Physica A 701 (2026) 132009

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

# Semi-supervised Hopfield model: Theoretical and Numerical results☆

Linda Albanese [a,d,e], Andrea Ladiana [b,d], Andrea Lepre [c,d,*]

[a] Dipartimento di Matematica e Fisica "Ennio De Giorgi", Università del Salento, Italy
[b] Dipartimento di Scienze di Base e Applicazioni all'Ingegneria, Sapienza Università di Roma, Italy
[c] Dipartimento di Matematica, Sapienza Università di Roma, Italy
[d] Istituto Nazionale di Alta Matematica Francesco Severi (INdAM), Roma, Italy
[e] Istituto Nazionale di Fisica Nucleare, Sezione di Lecce, Italy

ARTICLE INFO

*Keywords:*
Semi-supervised learning
Hopfield neural networks
Replica-symmetric statistical mechanics

ABSTRACT

In the daily practice of Machine Learning, fully labeled datasets are a luxury: labels demand expensive and time-consuming human annotation, whereas raw, unlabeled data can be harvested automatically and in bulk. Semi-supervised learning, where the network jointly exploits the few labeled and the many unlabeled examples at its disposal, is the standard answer to this asymmetry, yet a statistical mechanical theory of semi-supervised Hebbian learning is still lacking. In this paper we fill this gap for the Hopfield network: we prescribe a synaptic coupling given by the convex combination, weighted by a mixing parameter $\lambda \in [0,1]$, of the supervised and unsupervised Hebbian kernels built from the same archetypes, and we solve for the emergent computational capabilities of the resulting network.

A signal-to-noise analysis yields the one-step Mattis magnetization and the learning threshold, i.e., the minimum dataset size for stable retrieval, the latter being obtained in closed-form for a balanced dataset. Using Guerra's interpolation, we then derive the replica-symmetric quenched pressure in the high-storage regime, treating the correlated disorder generated by the supervised and unsupervised channels through a particular eigen-channel decomposition. The resulting phase diagram shows that a mixed strategy outperforms both pure protocols. Finally, we prove that the quenched pressure is convex in $\lambda$, so thermodynamics cannot select an interior mixture: $\lambda$ is therefore a learning hyperparameter.

All the analytical findings are successfully checked against extensive Monte Carlo simulations.

## 1. Introduction

In modern Machine Learning, labels are the expensive ingredient: annotating a dataset requires human expertise, time and money, whereas unlabeled data can be collected automatically and in massive quantities. As a consequence, the datasets a learning machine actually experiences are almost never fully labeled, nor completely unlabeled, and the standard answer of the Machine Learning community to this structural asymmetry is *semi-supervised learning*, where the (few) labeled examples and the (many) unlabeled ones are jointly exploited during training [1,2], from entropy-based regularization [3] to deep generative modeling [4]. Despite its ubiquity in applications, this paradigm still lacks a transparent, analytically solvable model within the Hebbian framework of

☆ This article is part of a Special issue entitled: 'Statistical mechanics for Artificial Learning' published in Physica A.

* Corresponding author.
*E-mail address:* andrea.lepre@uniroma1.it (A. Lepre).

https://doi.org/10.1016/j.physa.2026.132009

Available online 9 September 2026
0378-4371/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

associative memories, where the interplay between the two channels of information is under full mathematical control: providing such a model, and solving it, is the purpose of the present paper.

The natural framework for this program is the statistical mechanics of neural networks [5,6], which we now briefly outline.

Since Hopfield's seminal work on biologically inspired associative memories [7], the statistical mechanics of spin glasses [6,8] has proved to be a privileged route towards a quantitative understanding of information processing in neural networks [5,9–11]. The theory developed in the eighties by Amit, Gutfreund and Sompolinsky (AGS) [12] is a milestone in this field, which turned the Hopfield model, a fully-connected mean-field network whose pairwise couplings implement the celebrated Hebbian prescription, into a solvable statistical mechanical system, whose phase diagram sharply splits the space of the control parameters (noise level and storage load) into regions with different emergent computational capabilities.

Since then, the Hopfield model has become the fundamental *associative memory* network, spawning countless variations on a theme, including its dense generalizations with many-body couplings [13–20].

However, in its original formulation, the Hopfield network does not really *learn*: the patterns are experienced just once and directly stored in the synaptic matrix, while in Machine Learning jargon *learning* refers to the ability of the machine to build its own representation of unknown archetypes out of a corpus of their noisy examples. In recent years this conceptual gap has been bridged by turning the Hebbian storing prescription into genuine Hebbian *learning* rules, split into a supervised setting, where a teacher groups the examples class by class, and an unsupervised one, where the network experiences the whole dataset with no label at all: see in particular [21–23] and [24] for a derivation of these prescriptions from first principles via maximum entropy. Within these frameworks one can compute, beyond the storage capacity, the *threshold for learning*, that is the minimal number of examples per archetype allowing the network to infer, and subsequently retrieve, the archetypes hidden in the dataset: conceptually, this is the closest statistical mechanical analog of the sample complexity bounds of statistical learning theory [25].

So far, however, the supervised and unsupervised protocols have essentially been investigated separately, one at a time, while, as stressed above, the operationally relevant scenario is the mixed one. Semi-supervision has of course not escaped the attention of the statistical mechanics community, but it has been addressed along different routes. In the framework of on-line learning, the generalization error of a semi-supervised protocol has been obtained through a set of deterministic differential equations for the order parameters, which connect continuously the supervised and the semi-supervised prescriptions [26], albeit for a teacher-student perceptron trained by gradient-based rules.

A second, by now classical line of research addresses semi-supervised *clustering* and community detection: there the graph itself is the datum and a fraction of the node labels is revealed to the observer, so that the spin glass machinery delivers the distribution of the classifications rather than the sole minimal cut [27], the free energy of the constrained problem together with the fate of its criticality [28], the detectability thresholds of the semi-supervised stochastic block model [29] and the large deviations associated with the choice of the revealed subset [30]. Hopfield networks themselves have been employed in this very spirit, the couplings encoding the similarity among the items of a given dataset and the dynamics, restricted to the unlabeled nodes, propagating the labels of the observed ones: a strategy which proved effective even under strongly unbalanced labelings [31,32] and which is still exploited in modern computer vision pipelines for automatic image annotation [33]. In all these settings the network is a device for *transductive inference* on a fixed dataset: the labels are propagated among the very items whose mutual similarity defines the couplings, and no archetype is reconstructed out of a corpus of its noisy examples.

In this work, instead, we consider a Hopfield network supplied with a dataset made of $M_L$ labeled and $M_U$ unlabeled noisy examples of $K$ random archetypes, and we prescribe a synaptic coupling given by the convex combination of the supervised and unsupervised Hebbian kernels, ruled by a mixing parameter $\lambda \in [0, 1]$ whose extreme values recover the two pure protocols. As we shall prove, $\lambda$ must not be confused with the empirical fraction of labeled examples $M_L/(M_L + M_U)$, but it can be considered as a synaptic weight entering the learning rule, which the experimenter is free to tune. Its optimal value results from a non-trivial balance between the relative abundances of the two channels and the intrinsic quality of the examples, and it collapses onto the naive labeled fraction only in the limit of noiseless datasets.

Our results are both analytical and numerical. First, by a signal-to-noise analysis of the one-step dynamics at zero temperature, we compute in closed-form the Mattis magnetization achieved after a single network update, whence we extract the threshold for learning $M^\star(\alpha, r, \lambda)$ as a function of the storage load $\alpha$, the dataset quality $r$ and the mixing $\lambda$; these predictions are in complete agreement with Monte Carlo simulations. We would like to emphasize that the closed-form expression of $M^\star$ is derived for a balanced dataset, namely $M_L = M_U = M/2$: a single stability condition cannot fix the two channel sizes separately, hence this restriction is assumed throughout whenever the learning threshold is discussed. Next, we solve the model at the replica-symmetric level of description by means of Guerra's interpolation [34–36]. This step requires some care: since the two channels are built out of the same archetypes, the quenched noises they generate are mutually correlated. The supervised channel can be gaussianized by the Central Limit Theorem and the unsupervised one by the universality of the quenched noise established by Carmona and Hu [37,38], while the resulting structured disorder is treated exactly through a spectral decomposition into independent eigen-channels, each carrying its own eigenvalue and multiplicity. We thus obtain the quenched statistical pressure, the self-consistency equations for the order parameters and, in the zero-temperature limit, the phase diagram of the network: remarkably, at intermediate values of $\lambda$ the retrieval region is strictly larger than in both the pure protocols, namely semi-supervision genuinely pays off.

Finally, we address the question of the optimal mixing. We prove that the quenched pressure is convex in $\lambda$, so that an interior stationary point, characterized by an exact equipartition of the energies extracted from the two channels, is a *maximum* of the free energy: equilibrium thermodynamics, by itself, would push the synaptic rule towards the pure channels and never stop midway. This no-go result formally legitimizes the operational reading of $\lambda$ as an externally tuned hyperparameter and shifts the optimality criterion from the free energy to the retrieval capacity, whose maximization yields a closed-form expression for $\lambda^\star$.

The paper is structured as follows: in Section 2 we introduce the model, together with its control and order parameters; in Section 3 we develop the signal-to-noise analysis, deriving the one-step Mattis magnetization and the threshold for learning; in Section 4 we solve the model at the replica-symmetric level via Guerra's interpolation, obtaining the quenched statistical pressure, the self-consistency equations and the phase diagrams; Section 5 is devoted to the optimal mixing parameter, addressed both from the thermodynamic and from the retrieval perspective; finally, conclusions and outlooks are drawn in Section 6. Technical details and lengthy computations are collected in the Appendices.

## 2. Generalities

Let us consider $N$ neurons $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_N) \in \{-1, +1\}^N$ and a database $\mathcal{S} = \{\boldsymbol{\eta}^{\mu a}\}_{\mu=1,\dots,K}^{a=1,\dots,M}$, where $\eta_i^{\mu,a} \in \{-1, +1\}$. The elements of the dataset are assumed to be some noisy examples generated from $K$ archetypes $\boldsymbol{\xi}^\mu \in \{-1, +1\}^N$, $\mu = 1, \dots, K$, whose entries are i.i.d. Rademacher random variables, namely

$$\mathbb{P}(\xi_i^\mu = \pm 1) = \frac{1}{2}, \qquad i = 1, \dots, N, \quad \mu = 1, \dots, K. \tag{2.1}$$

Therefore, the dataset is constructed so as to emulate a *semi-supervised setting*: it is split into two subsets, the first one, made of $M_L$ elements and indexed by $a = 1, \dots, M_L$, being *labeled* and playing the role of the supervised channel, whereas the remaining $M_U$ elements, indexed by $c = 1, \dots, M_U$, are *unlabeled* and constitute the unsupervised channel, with $M = M_L + M_U$.

More precisely, each element of the database can be written as

$$\eta_i^{\mu a} = \xi_i^\mu \, \chi_i^{\mu a}, \qquad \mathbb{P}(\chi_i^{\mu a} = \pm 1) = \frac{1 \pm r}{2}, \quad r \in (0, 1], \tag{2.2}$$

where the binary random variables $\chi_i^{\mu a}$ fulfill

$$\mathbb{E}[\chi_i^{\mu a}] = r, \qquad \mathbb{E}[(\chi_i^{\mu a})^2] = 1, \qquad \mathbb{E}[\chi_i^{\mu a} \chi_i^{\mu b}] = \delta_{ab} + (1 - \delta_{ab}) r^2, \tag{2.3}$$

which yields, for the example matrix elements,

$$\mathbb{E}[\eta_i^{\mu a}] = 0, \qquad \mathbb{E}[\eta_i^{\mu a} \eta_j^{\nu b}] = \delta_{ij} \, \delta_{\mu\nu} \left[\delta_{ab} + (1 - \delta_{ab}) r^2\right]. \tag{2.4}$$

We stress that for $r = 1$, the dataset becomes a trivial set of copies of the archetypes.[1]

Following the seminal work by Hopfield [7] and the Hebbian prescription, the coupling matrix $J_{ij}$ is given by a convex combination of the couplings induced by the labeled and unlabeled subsets of the dataset

$$J_{ij}^\lambda = \frac{1}{N} \sum_{\mu=1}^{K} \left[ \frac{\lambda}{\Gamma_L} \sum_{a,b=1}^{M_L} \eta_i^{\mu a} \eta_j^{\mu b} + \frac{1 - \lambda}{\Gamma_U} \sum_{c=1}^{M_U} \eta_i^{\mu c} \eta_j^{\mu c} \right] = \lambda J_{ij}^{(L)} + (1 - \lambda) J_{ij}^{(U)}, \tag{2.6}$$

with $\lambda \in [0, 1]$ a mixing parameter tuning the relative weight of the two channels. The extreme cases $\lambda \in \{0, 1\}$ correspond to the absence of one of the two groups inside the database.

The Hamiltonian is the standard quadratic form

$$\begin{aligned} \mathcal{H}_N(\boldsymbol{\sigma} | \mathcal{S}, \lambda) &= -\frac{1}{2} \sum_{i,j=1}^{N} J_{ij}^\lambda \, \sigma_i \sigma_j \\ &= -\frac{1}{2N} \sum_{\mu=1}^{K} \sum_{i,j=1}^{N} \left[ \frac{\lambda}{\Gamma_L} \sum_{a,b=1}^{M_L} \eta_i^{\mu a} \eta_j^{\mu b} + \frac{1 - \lambda}{\Gamma_U} \sum_{c=1}^{M_U} \eta_i^{\mu c} \eta_j^{\mu c} \right] \sigma_i \sigma_j, \end{aligned} \tag{2.7}$$

with $\Gamma_L$ and $\Gamma_U$ two normalization factors fixed as follows[2]:

$$\Gamma_L := M_L^2 \left[ r^2 + \tfrac{1 - r^2}{M_L} \right], \qquad \Gamma_U := M_U \, r^2. \tag{2.8}$$

A schematic representation of the network is depicted in Fig. 1.

[1] Another way to rewrite the elements of the dataset with respect to the archetypes is to consider an additive noise rather than a multiplicative one. This can be achieved by decomposing each variable $\chi_i^{\mu,a}$ into its expectation $r$ and a zero-mean fluctuation with variance $1 - r^2$: $\chi_i^{\mu,a} = r + \tilde{\chi}_i^{\mu,a}$, $\mathrm{Var}(\tilde{\chi}_i^{\mu,a}) = \mathbb{E}(\tilde{\chi}_i^{\mu,a})^2 = 1 - r^2$, $i = 1, \dots, N$, $a = 1, \dots, M$, $\mu = 1, \dots, K$. Therefore,

$$\eta_i^{\mu a} = \xi_i^\mu (r + \tilde{\chi}_i^{\mu,a}) = r \, \xi_i^\mu + \zeta_i^{\mu a}, \qquad \zeta_i^{\mu a} := \xi_i^\mu \tilde{\chi}_i^{\mu,a}. \tag{2.5}$$

This will be useful in the signal analysis of Appendix B.

[2] The two factors differ because the two channels offer different objects to be normalized. The labels allow the supervised channel to merge all the $M_L$ examples of a given archetype into a single collective variable $\sum_a \eta_i^{\mu a}$, whose second moment follows from (2.3), $\mathbb{E}\left[\left(\sum_a \eta_i^{\mu a}\right)^2\right] = M_L^2 r^2 + M_L(1 - r^2) = \Gamma_L$; dividing by it gives the supervised block unit second moment. In the unsupervised channel no such collective variable can be formed, each example entering the coupling through its own outer product, of second moment $\mathbb{E}[(\eta_i^{\mu c})^2] = 1$: the same prescription would give $\Gamma_U = M_U$ and would leave the retrieval signal proportional to $r^2$, hence vanishing on poor datasets, so the unsupervised block is normalized on its signal, see Appendix A for more details.

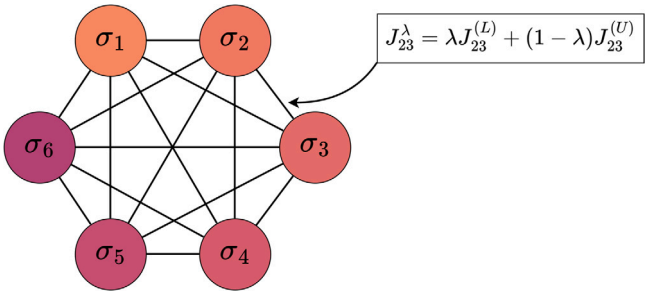



**Fig. 1.** Toy representation of the network whose Hamiltonian is defined in (2.7) with $N = 6$. Each neuron interacts with all the others through the effective synaptic coupling $J_{ij}^{\lambda}$ in (2.6), given by the convex combination of the supervised and unsupervised Hebbian contributions.

**Remark 1.** A clarification is now in order on the quality parameter $r$, which enters the normalizations (2.8) of the learning rule and might therefore look like a piece of privileged information that the network has no right to possess. It is not, and two questions have to be addressed in order to see it: whether $r$ can be recovered from the dataset alone, and how accurately it has to be recovered.

In a semi-supervised setting the labeled channel offers the most direct route to the first question. Denoting by $\bar{\eta}_{i,(L)}^{\mu} := \frac{1}{M_L}\sum_{a=1}^{M_L} \eta_i^{\mu a}$ the empirical mean over the labeled examples of the archetype $\mu$, introduced in (2.20), relations (2.3) give $\mathbb{E}[(\bar{\eta}_{i,(L)}^{\mu})^2] = r^2 + (1-r^2)/M_L = \mathcal{R}_L$, so that

$$\hat{r}^2 := \frac{M_L \widehat{\mathcal{R}}_L - 1}{M_L - 1}, \qquad \widehat{\mathcal{R}}_L := \frac{1}{KN}\sum_{\mu=1}^{K}\sum_{i=1}^{N}\left(\bar{\eta}_{i,(L)}^{\mu}\right)^2, \tag{2.9}$$

is an unbiased estimator of $r^2$ built exclusively on the observed examples. Since $\bar{\eta}_{i,(L)}^{\mu} = \xi_i^{\mu}\,\bar{\chi}_i^{\mu}$ and the square removes the archetype, the $KN$ terms of $\widehat{\mathcal{R}}_L$ are independent and bounded, so that $\hat{r}^2$ concentrates around its mean at the rate $(KN)^{-1/2}$. Only the labels are needed in order to form $\bar{\eta}_{i,(L)}^{\mu}$: however scarce the supervised channel may be, it is exactly what makes the quality of the dataset accessible to the learner, and the resulting estimate can then be fed into both the normalizations of (2.8).

Even with no label at all the parameter remains identifiable. By (2.4), the overlap $\frac{1}{N}\sum_i \eta_i^a \eta_i^b$ between two distinct unlabeled examples has expectation $r^2$ when they stem from the same archetype and 0 otherwise, with fluctuations of order $N^{-1/2}$: the histogram of the pairwise overlaps is thus a two-component mixture, one component centered at $r^2$ with the known weight $(M_U - 1)/(KM_U - 1)$ and one centered at zero, from which $r^2$ can be read off without knowing which example belongs to which archetype.

We turn to the second question, namely how much accuracy is actually required. Here the structure of the model gives a sharp answer, because $r$ enters the machine only through $\Gamma_L$ and $\Gamma_U$: whatever value is used, the resulting synaptic matrix is still a convex combination of the two Hebbian kernels, so that a wrong $r$ cannot alter the nature of the network, it can only move it to a different point of the $(\lambda, \beta)$ plane. Explicitly, writing the mis-specified normalizations as $\hat{\Gamma}_X = \Gamma_X(1+\delta_X)$, $X \in \{L, U\}$, the two weights of (2.6) become $\lambda/(1+\delta_L)$ and $(1-\lambda)/(1+\delta_U)$; collecting their sum $\mathcal{W}$ restores a convex combination and leaves

$$J^{\lambda}[\hat{\Gamma}] = \mathcal{W}\, J^{\lambda_{\text{eff}}}, \qquad \mathcal{W} := \frac{\lambda}{1+\delta_L} + \frac{1-\lambda}{1+\delta_U}, \qquad \lambda_{\text{eff}} := \frac{1}{\mathcal{W}}\,\frac{\lambda}{1+\delta_L}, \tag{2.10}$$

that is, exactly the correctly normalized coupling with the mixing $\lambda$ replaced by $\lambda_{\text{eff}}$ and, since only the product $\beta J$ enters the Boltzmann factor, the inverse temperature $\beta$ replaced by $\beta\mathcal{W}$. A misestimate of $r$ does not spoil the model, it reparametrizes it. Both displacements are small. For $\hat{r} = r(1+\varepsilon)$ one finds at first order in $\varepsilon$

$$\delta_U = 2\varepsilon, \qquad \delta_L = 2\varepsilon\,\frac{1 - 1/M_L}{1+\rho_L}, \tag{2.11}$$

so that the two shifts coincide up to corrections of order $1/M_L$ and $\rho_L$. Were they equal, the common factor would cancel between numerator and denominator of $\lambda_{\text{eff}}$ and the mixing would be left untouched, the whole effect of the error reducing to the rescaling $\beta \to \beta(1-2\varepsilon)$, immaterial in the zero-temperature regime in which the retrieval region is measured. What survives is a residual displacement of the mixing, $\lambda_{\text{eff}} - \lambda = O\big(\varepsilon(\rho_L + 1/M_L)\big)$, suppressed both by the accuracy of the estimate and by the size of the dataset.

The two answers meet here. The estimator (2.9) returns $r$ with a relative accuracy of order $(KN)^{-1/2}$ and the displacement of the mixing it induces is smaller still. The learning rule is thus safe by a wide margin, and the dependence of (2.8) on $r$ carries no hidden privileged information.

It is convenient to introduce now the *load* of the network, namely the ratio between the number of patterns and the size of the network in the thermodynamic limit

$$\alpha := \lim_{N\to+\infty} \frac{K}{N}. \tag{2.12}$$

In this work, we are mainly interested in the *high-storage regime*, namely when the load of the network $\alpha$ is a real finite number [12].

**Remark 2.** In the physical model under investigation self-interaction terms are excluded, that is, a neuron $\sigma_i$ interacts with any other neuron $\sigma_k$ with $k \neq i$. This should be accounted for in (2.7) by inserting a corrective contribution that neutralizes the diagonal terms. However, since this contribution is constant, it is neglected here in order to keep the notation simple.

In fact, for the unsupervised block:

$$J_{ii}^{(U)} = \frac{1-\lambda}{N\Gamma_U}\sum_{\mu,c} 1 = \frac{(1-\lambda)KM_U}{N\Gamma_U} = \frac{(1-\lambda)KM_U}{NM_U r^2} \to \frac{(1-\lambda)\alpha}{r^2}, \qquad N \to +\infty. \tag{2.13}$$

For the supervised block it is convenient to take the expectation value:

$$\mathbb{E}[J_{ii}^{(L)}] = \frac{\lambda}{N\Gamma_L}\sum_{\mu}\mathbb{E}\left[\left(\sum_a \eta_i^{\mu a}\right)^2\right] = \frac{\lambda K}{N\Gamma_L}\Gamma_L = \frac{\lambda K}{N} \to \lambda\alpha, \qquad N \to +\infty, \tag{2.14}$$

which are both constant and are therefore neglected.

We stress that this simplification concerns the analytical treatment only. In all the finite-size Monte Carlo simulations reported in this work the self-couplings are instead set exactly to zero, $J_{ii}^{\lambda} = 0$ for every $i = 1, \dots, N$, so that no neuron feeds back on itself and the numerical dynamics is free from any residual self-interaction.

**Remark 3.** After the introduction of the model, a remark regarding the role of $\lambda$ is in order.

It is important to distinguish the upstream stage of data collection from the downstream stage of synaptic learning. If one were allowed to actively allocate a fixed budget of $M$ examples, the trivial strategy for maximizing the retrieval capacity would be to collect only supervised examples, namely $M_L = M$ and $M_U = 0$. Indeed, the supervised channel benefits from the noise reduction induced by the Central Limit Theorem (CLT), while avoiding the divergent interference term of order $O(r^{-2})$ associated with the unsupervised channel [39,40].

In the present setting, however, $M_L$ and $M_U$ are external constraints fixed by the available dataset [1,2]. This standard paradigm in semi-supervised learning reflects the practical reality that acquiring labels is typically an expensive and time-consuming process requiring human expertise, whereas unlabeled data can be collected automatically and in massive quantities. Therefore, the parameter $\lambda \in [0,1]$ should *not* be interpreted as the empirical fraction of labeled data, for instance by imposing $\lambda = M_L/M$ (see for instance [24]). Rather, $\lambda$ is a genuine synaptic weight entering the Hebbian prescription, which dynamically tunes the relative relevance assigned by the network to the supervised and unsupervised channels during learning.[3]

Accordingly, in this manuscript we show that there exists an optimal value $\lambda^\star$ (see Section 5) which does not merely reflect the relative proportion of labeled and unlabeled examples, but rather results from a balance between such proportions and the intrinsic quality of the dataset. This confirms the role of $\lambda$ as a crucial and tunable hyperparameter of the synaptic learning rule.

Therefore, to sum up, the model is parametrized by the following *control parameters*, which are held fixed in the thermodynamic limit:

- the *load* of the network $\alpha$, defined in Eq. (2.12);
- the *quality* $r \in (0,1]$ of the examples, introduced for the creation of the database in Eq. (2.2);
- the per-channel *dataset sizes* $M_L, M_U$. Equivalently, one can also consider the per-channel *rescaled sizes*

$$\rho_X := \frac{1-r^2}{M_X r^2}, \qquad \mathcal{R}_X := r^2 + \frac{1-r^2}{M_X} = r^2(1+\rho_X), \tag{2.15}$$

  for $X \in \{L,U\}$. The scaling $M_X \propto r^{-2}$, $X \in \{L,U\}$, keeps $\rho_X$ of order one[4];
- the *mixing parameter* $\lambda \in [0,1]$ introduced in the Hamiltonian in (2.7) and the focus of Remark 3.
- the *inverse temperature* $\beta > 0$ tuning the stochasticity of the spins. Large $\beta$ values, corresponding to low temperatures, strongly favor low-energy states, while small $\beta$ values, corresponding to high temperatures, make higher-energy states more accessible. We will exploit it in the definition of the thermodynamic variables of the model (i.e., Boltzmann Gibbs probability distribution and partition function);

Then, in order to describe the macroscopic behavior of the system, one can introduce several *order parameters*. In particular, the retrieval order parameter is the standard *Mattis magnetization* which can be defined as the normalized dot product between $\boldsymbol{\xi}^\mu$ and $\boldsymbol{\sigma}$, namely for each $\mu = 1, \dots, K$ as

$$m^\mu(\boldsymbol{\sigma}|\boldsymbol{\xi}^\mu) := \frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu\sigma_i. \tag{2.16}$$

[3] We stress that the presence of a free coefficient weighting the unsupervised contribution against the supervised one is a recurrent feature of semi-supervised algorithms, rather than an artifact of the present construction: in entropy-regularized semi-supervised learning, for instance, the unlabeled examples enter the cost function through an entropy term whose relative weight is a hyperparameter tuned by the experimenter and not dictated by the labeled fraction [3], exactly as $\lambda$ here.

[4] $\rho_X$ has been also denoted as *entropy* of the dataset for its role in the description of the information contained in it. For more details, please see [21,39,40].

We can also define, for each $\mu = 1, \dots K$ and for each $a = 1, \dots M_X$, with $X \in \{L, U\}$, the *single example overlap*:

$$n_X^{\mu,a}(\boldsymbol{\sigma}|\boldsymbol{\eta}^\mu) := \frac{1}{N}\sum_{i=1}^{N} \eta_i^{\mu a}\sigma_i, \tag{2.17}$$

which can be used for writing the Hamiltonian in the following form

$$-\beta\mathcal{H}(\boldsymbol{\sigma}|S,\lambda) = \lambda\,\beta N\,\mathcal{E}_L(\boldsymbol{\sigma}|S,\lambda) + (1-\lambda)\,\beta N\,\mathcal{E}_U(\boldsymbol{\sigma}|S,\lambda) + O(1), \tag{2.18}$$

where

$$\mathcal{E}_L(\boldsymbol{\sigma}) := \frac{1}{2\Gamma_L N}\sum_{\mu=1}^{K}\Big(\sum_{a=1}^{M_L} n_L^{\mu,a}\Big)^2, \qquad \mathcal{E}_U(\boldsymbol{\sigma}) := \frac{1}{2\Gamma_U N}\sum_{\mu=1}^{K}\sum_{c=1}^{M_U}(n_U^{\mu,c})^2. \tag{2.19}$$

It is also useful to define the mean of the labeled and unlabeled examples:

$$\bar{\eta}_{i,(X)}^{\mu} := \frac{1}{M_X}\sum_{a=1}^{M_X}\eta_i^{\mu a} \tag{2.20}$$

and the corresponding *mean example magnetizations* are defined respectively as

$$n_L^{\mu}(\boldsymbol{\sigma}|\bar{\boldsymbol{\eta}}^\mu) := \frac{r}{\mathcal{R}_L}\frac{1}{N}\sum_{i=1}^{N}\bar{\eta}_{i,(L)}^{\mu}\sigma_i, \qquad n_U^{\mu}(\boldsymbol{\sigma}|\bar{\boldsymbol{\eta}}^\mu) := \frac{r}{\mathcal{R}_U}\frac{1}{N}\sum_{i=1}^{N}\bar{\eta}_{i,(U)}^{\mu}\sigma_i. \tag{2.21}$$

The prefactor $r/\mathcal{R}_X$ normalizes each $n_X^\mu$ so that, at the retrieval saddle $\boldsymbol{\sigma} = \boldsymbol{\xi}^\mu$, one has $\mathbb{E}[n_X^\mu] = 1 = \mathbb{E}[m^\mu]$.

**Remark 4.** Using again (2.17), it is possible to rewrite (2.7) in a matrix form. Indeed, if we consider the vector

$$\mathbf{n}_\mu := \big(n_L^{\mu,1}, \dots, n_L^{\mu,M_L}; n_U^{\mu,1}, \dots, n_U^{\mu,M_U}\big)^\top \in \mathbb{R}^M, \tag{2.22}$$

and we introduce the block matrix

$$W := \begin{pmatrix} \frac{\beta\lambda}{\Gamma_L}\mathbf{1}_{M_L}\mathbf{1}_{M_L}^\top & \mathbf{0} \\ \mathbf{0} & \frac{\beta(1-\lambda)}{\Gamma_U}\mathbb{I}_{M_U} \end{pmatrix} \in \mathbb{R}^{M\times M}, \tag{2.23}$$

where $\mathbf{1}_{M_L} \in \mathbb{R}^{M_L}$ is the all-ones vector and $\mathbb{I}_{M_U}$ is the identity matrix $M_U \times M_U$, we can rewrite the Hamiltonian (2.7) as

$$-\beta\,\mathcal{H}(\boldsymbol{\sigma}|S) = \frac{N}{2}\sum_{\mu=1}^{K}\mathbf{n}_\mu^\top W\,\mathbf{n}_\mu \;+\; O(1). \tag{2.24}$$

Because of the presence of quenched disorder, we also introduce the *two-replica overlap* (also known as *Edwards-Anderson overlap*) of the Ising spins, which reads

$$q_{ab}(\boldsymbol{\sigma}) := \frac{1}{N}\sum_{i=1}^{N}\sigma_i^{(a)}\sigma_i^{(b)}, \tag{2.25}$$

which measures the spin-glass character of the equilibrium measure (notice that, for $a = b$, $q_{aa}(\boldsymbol{\sigma})$ is identically equal to 1).

In order to develop the statistical mechanics techniques, we need to introduce also the *partition function* and the *quenched statistical pressure* at finite size $N$, defined as

$$\mathcal{Z}_{N,\beta}(S) := \sum_{\{\boldsymbol{\sigma}\}}\exp\big[-\beta\,\mathcal{H}_N(\boldsymbol{\sigma}|S)\big], \tag{2.26}$$

$$\mathcal{A}_N(\beta, r, \rho_L, \rho_U, \lambda) := \frac{1}{N}\,\mathbb{E}\big[\ln\mathcal{Z}_{N,\beta}(S)\big], \tag{2.27}$$

where $\sum_{\{\boldsymbol{\sigma}\}}$ runs over all possible configurations of $\boldsymbol{\sigma}$, $\mathbb{E} := \mathbb{E}_S$ is the quenched expectation over the elements of the database and $\exp\big[-\beta\,\mathcal{H}_N(\boldsymbol{\sigma}|S)\big]$ is denoted as *Boltzmann factor*.

It is possible to define the thermodynamic quenched statistical pressure

$$\mathcal{A}(\alpha,\beta,r,\rho_L,\rho_U,\lambda) := \lim_{N\to+\infty}\mathcal{A}_N(\alpha,\beta,r,\rho_L,\rho_U,\lambda) = \lim_{N\to+\infty}\frac{1}{N}\,\mathbb{E}\big[\ln\mathcal{Z}_{N,\beta}(S)\big], \tag{2.28}$$

assuming its existence and uniqueness.[5]

The Boltzmann–Gibbs measure on the spin configurations is

$$\mathbb{P}_{\beta,\lambda}(\boldsymbol{\sigma}|S) = \frac{1}{\mathcal{Z}_{N,\beta}(S)}\,e^{-\beta\mathcal{H}(\boldsymbol{\sigma}|S)}, \tag{2.29}$$

[5] The existence and uniqueness of the thermodynamic quenched statistical pressure have already been proven by Guerra and Toninelli and by Talagrand in [36,41] for the Sherrington-Kirkpatrick model [8]. Moreover, recent work extends this result to the Hopfield model [42].

and the expectation of any observable $\mathcal{O}(\boldsymbol{\sigma})$ is denoted by $\omega(\mathcal{O}) = \sum_{\boldsymbol{\sigma}} \mathbb{P}_{\beta,\lambda}(\boldsymbol{\sigma}|S)\mathcal{O}(\boldsymbol{\sigma})$; the disorder-averaged Boltzmann expectation is $\langle \mathcal{O} \rangle := \mathbb{E}[\omega(\mathcal{O})]$.

## 3. Signal-to-noise and stability analysis

The signal-to-noise ratio (SNR) [12] is a useful tool to understand some features of the model. In this section we will present the main results using this technique; for more details we refer to Appendix A.

Briefly, the SNR is based on the comparison between the retrieval part, namely the one linked to the patterns we have chosen to be retrieved, from now on denoted as the *signal*, and all the remaining contributions, which act as an intrinsic *noise* in the network. Without loss of generality, we assume that the signal is carried by the terms with $\mu = 1$, whereas those with $\mu \geq 2$ constitute the noise:

$$\begin{aligned}\mathcal{H}_N(\boldsymbol{\sigma}|S,\lambda) = &-\frac{1}{2N}\sum_{i,j=1}^{N}\left[\frac{\lambda}{\Gamma_L}\sum_{a,b=1}^{M_L}\eta_i^{1,a}\eta_j^{1,b} + \frac{1-\lambda}{\Gamma_U}\sum_{c=1}^{M_U}\eta_i^{1,c}\eta_j^{1,c}\right]\sigma_i\sigma_j \\ &-\frac{1}{2N}\sum_{\mu\geq 2}^{K}\sum_{i,j=1}^{N}\left[\frac{\lambda}{\Gamma_L}\sum_{a,b=1}^{M_L}\eta_i^{\mu a}\eta_j^{\mu b} + \frac{1-\lambda}{\Gamma_U}\sum_{c=1}^{M_U}\eta_i^{\mu c}\eta_j^{\mu c}\right]\sigma_i\sigma_j,\end{aligned} \tag{3.1}$$

therefore in terms of the Boltzmann factor we get

$$e^{-\beta\mathcal{H}(\boldsymbol{\sigma})} = \underbrace{\exp\left[\mathcal{B}_{\text{sig}}\right]}_{\mu=1}\cdot\underbrace{\exp\left[\mathcal{B}_{\text{noise}}\right]}_{\mu\geq 2}, \tag{3.2}$$

with, using (2.17):

$$\mathcal{B}_{\text{sig}} := \frac{\beta\lambda N}{2\Gamma_L}\left(\sum_{a=1}^{M_L} n_L^{1,a}\right)^2 + \frac{\beta(1-\lambda)N}{2\Gamma_U}\sum_{c=1}^{M_U}(n_U^{1,c})^2, \tag{3.3}$$

$$\mathcal{B}_{\text{noise}} := \frac{\beta\lambda N}{2\Gamma_L}\sum_{\mu\geq 2}^{K}\left(\sum_{a=1}^{M_L} n_L^{\mu,a}\right)^2 + \frac{\beta(1-\lambda)N}{2\Gamma_U}\sum_{\mu\geq 2}^{K}\sum_{c=1}^{M_U}(n_U^{\mu,c})^2. \tag{3.4}$$

We now define the model's dynamics in the $\beta \to \infty$ limit through the following update rule, our aim being to assess the state of the network at time $t = 1$, starting from the initial configuration $\boldsymbol{\sigma}^{(t=0)} = \boldsymbol{\xi}^1$:

$$\sigma_i^{(t+1)} = \text{sign}\left(h_i\sigma_i^{(t)}\right) \tag{3.5}$$

where

$$\begin{aligned}h_i &:= h_i^{(L)} + h_i^{(U)} \\ &:= \frac{\lambda}{N\Gamma_L}\sum_{\mu=1}^{K}\sum_{a,b=1}^{M_L}\eta_i^{\mu a}\sum_{j\neq i}\eta_j^{\mu b}\xi_j^1 + \frac{1-\lambda}{N\Gamma_U}\sum_{\mu=1}^{K}\sum_{c=1}^{M_U}\eta_i^{\mu c}\sum_{j\neq i}\eta_j^{\mu c}\xi_j^1.\end{aligned} \tag{3.6}$$

After deriving the state of the network at $t = 1$, we are interested in the behavior of the one-step Mattis magnetization on the target archetype $\boldsymbol{\xi}^1$ in the zero-temperature and thermodynamic limits, namely

$$m^{(1)}(\boldsymbol{\xi}^1) := \frac{1}{N}\sum_{i=1}^{N}\text{sgn}(h_i)\xi_i^1. \tag{3.7}$$

Since, by the CLT, the quantity $h_i\xi_i^1$ is asymptotically Gaussian as $N \to +\infty$, we can compute the thermodynamic one-step Mattis magnetization on the target archetype $\boldsymbol{\xi}^1$ as

$$m^{(1)}(\boldsymbol{\xi}^1) \xrightarrow{N\to\infty} \text{erf}\left(\frac{S}{\sqrt{2V}}\right), \tag{3.8}$$

where $S := \mathbb{E}\left[h_i\,\xi_i^1\right] = \mathbb{E}\left[(h_i^{(L)} + h_i^{(U)})\xi_i^1\right]$ and $V := \text{Var}\left(h_i\,\xi_i^1\right)$. Let us start with the computation of the signal; for all the details we refer to Appendix A. Focusing only on the results, the contributions to the signal from the supervised and unsupervised parts are

$$\mathbb{E}\left[h_i^{(L),\,\mu=1}\xi_i^1\right] = \frac{\lambda}{N\Gamma_L}M_L r(N-1)M_L r \xrightarrow{N\to\infty} \frac{\lambda M_L^2 r^2}{\Gamma_L} = \frac{\lambda}{1+\rho_L} \tag{3.9}$$

$$\mathbb{E}\left[h_i^{(U),\,\mu=1}\xi_i^1\right] = \frac{1-\lambda}{N\Gamma_U}M_U r(N-1)r \xrightarrow{N\to\infty} \frac{(1-\lambda)M_U r^2}{\Gamma_U} = 1-\lambda, \tag{3.10}$$

therefore

$$S = \mathbb{E}\left[(h_i^{(L)} + h_i^{(U)})\xi_i^1\right] \xrightarrow{N\to\infty} \frac{\lambda}{1+\rho_L} + (1-\lambda). \tag{3.11}$$

As far as the noise terms are concerned, the calculations are more cumbersome: the variance of $\xi_i^1 h_i$ receives several contributions from every cross-term of $\left(h_i^{(L)} + h_i^{(U)}\right)^2$. As shown again in Appendix A, the variance can be written as

$$V = \mathrm{Var}(h_i \xi_i^1) \xrightarrow{N\to\infty} C(\lambda, r, \rho_L, \rho_U)\,\alpha + R(\lambda, r, \rho_L, \rho_U), \tag{3.12}$$

where we have put together the slow *noise interference terms* in $C(\lambda, r, \rho_L, \rho_U)$ and the *within-class noise* in $R(\lambda, r, \rho_L, \rho_U)$:

$$C(\lambda, r, \rho_L, \rho_U) = \lambda^2 + (1-\lambda)^2\left[1 + \frac{1-r^4}{M_U r^4}\right] + \frac{2\lambda(1-\lambda)}{1+\rho_L}, \tag{3.13}$$

$$R(\lambda, r, \rho_L, \rho_U) = \frac{\lambda^2 \rho_L}{(1+\rho_L)^2} + (1-\lambda)^2 \rho_U. \tag{3.14}$$

Now, putting (3.13) and (3.14) in (3.8) we get

$$m^{(1)}(\boldsymbol{\xi}^1) = \mathrm{erf}\left(\frac{\frac{\lambda}{1+\rho_L} + (1-\lambda)}{\sqrt{2(C(\lambda, r, \rho_L, \rho_U)\alpha + R(\lambda, r, \rho_L, \rho_U))}}\right). \tag{3.15}$$

Some limiting behaviors are immediate:

- As $\alpha \to 0$ (*low storage*), the variance $V$ reduces to $R(\lambda, r, \rho_L, \rho_U)$ and thus $m^{(1)} \to \mathrm{erf}(S/\sqrt{2R})$. If, in addition, $r = 1$ (noiseless data), then $\rho_L = \rho_U = 0$ and therefore $R = 0$, whence $m^{(1)}(\boldsymbol{\xi}^1) \to 1$, namely perfect retrieval.
- As $r \to 0$ (*low-quality data*), at fixed $\rho_U$ and for every $\lambda < 1$, the interference coefficient *diverges*, $C(\lambda, r \to 0, \rho_L, \rho_U) = O(r^{-2})$. Indeed, at fixed $\rho_U$ the quartic contribution in (3.13) can be recast as $\frac{1-r^4}{M_U r^4} = \frac{1+r^2}{r^2}\,\rho_U$, which blows up as $r^{-2}$. Consequently the cross-talk noise overwhelms the signal and $m^{(1)}(\boldsymbol{\xi}^1)$ goes to 0.
- As $\lambda \to 1$ (*supervised limit*), $S = 1/(1+\rho_L)$, $V = \alpha + \rho_L/(1+\rho_L)^2$, hence

$$m^{(1)}_{\lambda=1}(\boldsymbol{\xi}^1) = \mathrm{erf}\left(\frac{1}{\sqrt{2\alpha(1+\rho_L)^2 + 2\rho_L}}\right), \tag{3.16}$$

  reproducing the results obtained in [21].
- As $\lambda \to 0$ (*unsupervised limit*), $S = 1$, $V = \alpha[1 + (1-r^4)/(M_U r^4)] + \rho_U$, hence

$$m^{(1)}_{\lambda=0} = \mathrm{erf}\left(\frac{1}{\sqrt{2\alpha[1 + (1-r^4)/(M_U r^4)] + 2\rho_U}}\right), \tag{3.17}$$

  reproducing [21,23].

In order to estimate the minimal number of examples per archetype required for a stable retrieval phase, we impose the stability condition $m^{(1)}(\boldsymbol{\xi}) > \mathrm{erf}(\theta)$, where $\theta$ is a threshold of order unity; we choose $\theta = \frac{1}{\sqrt{2}}$, since it is the natural stability boundary at which the signal is at least as large as the standard deviation of the noise. We thus get the condition

$$C(\lambda, r, \rho_L, \rho_U)\alpha + R(\lambda, r, \rho_L, \rho_U) < \frac{\frac{\lambda}{1+\rho_L} + (1-\lambda)}{2\theta^2}. \tag{3.18}$$

Our goal is now to determine the minimal number of examples in the supervised and unsupervised datasets from the condition (3.18). As it stands, however, the problem is not analytically tractable, since a single inequality cannot fix the two sizes $M_L$ and $M_U$ separately: a further condition would be needed. We therefore restrict ourselves to a *balanced* dataset, with the same number of examples in each channel, namely $M_L = M_U = M/2$.

In this case it is possible to find a closed form of (3.18), namely

$$M^\star(\alpha, r, \lambda; \theta) \simeq \frac{\alpha(1-\lambda)^2\,\frac{1-r^4}{r^4} + \left[\lambda^2 + (1-\lambda)^2\right]\frac{1-r^2}{r^2}}{\frac{1}{2\theta^2} - \alpha}, \qquad \alpha < \frac{1}{2\theta^2}. \tag{3.19}$$

The analytical behavior of $M^\star$ is in complete agreement with Monte Carlo (MC) simulations, as shown in Fig. 2 for different values of the size of the network $N$ and number of patterns $K$.

## 4. Analytical results in RS assumption

The SNR analysis and the numerical simulations presented in Section 3 provide a solid heuristic foundation for understanding the learning thresholds of the semi-supervised Hopfield model. To corroborate these findings analytically, we now show how the same

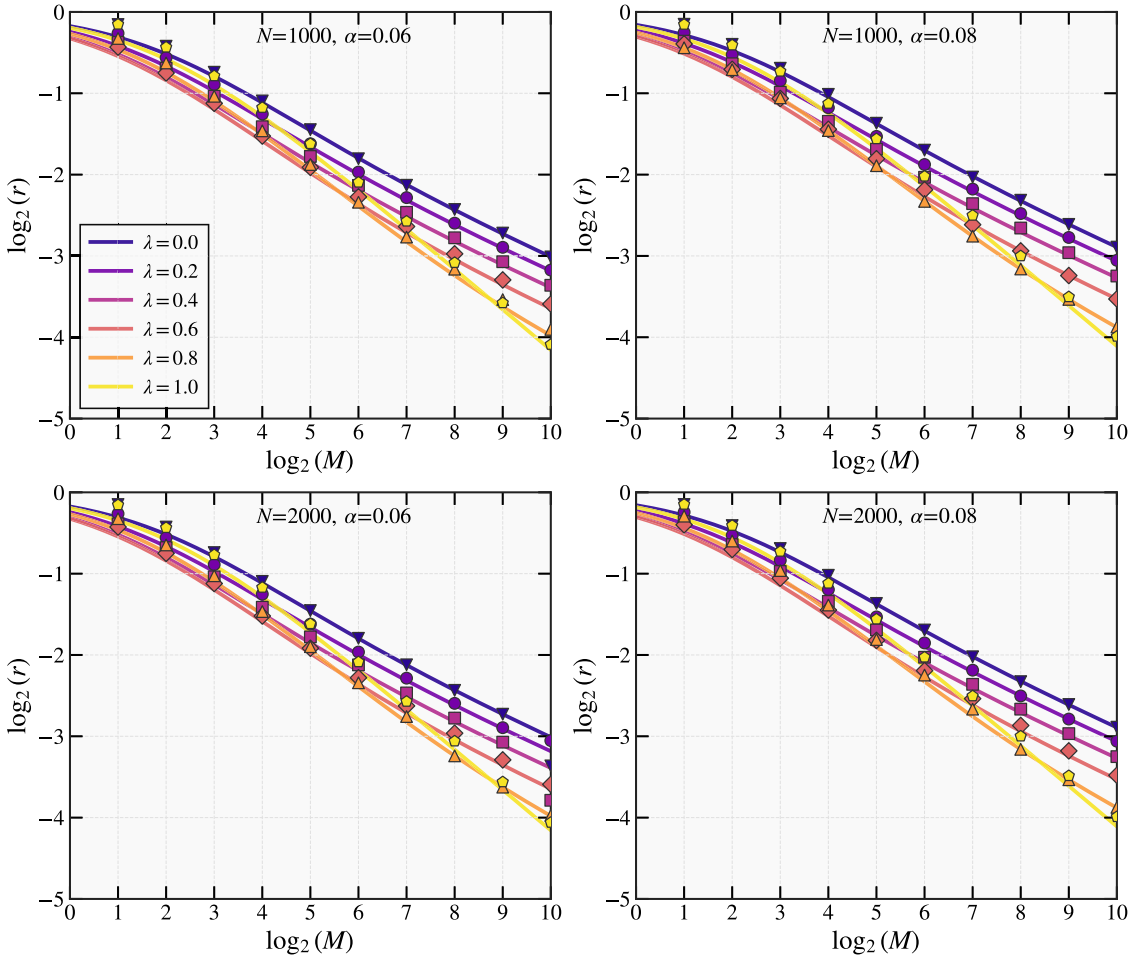



**Fig. 2.** Comparison between SNR theoretical predictions (solid lines) and noiseless ($T = 0$) MC simulations (symbols; see Appendix E.2) for the semi-supervised Hopfield network, showing the phase boundaries in the $(\log_2 M, \log_2 r)$ plane for different values of the mixing parameter $\lambda \in \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$. The four panels display different system sizes $N$ and network loads $\alpha_N = K/N$: (a) $N = 1000$, $\alpha = 0.06$, (b) $N = 1000$, $\alpha = 0.08$, (c) $N = 2000$, $\alpha = 0.06$, and (d) $N = 2000$, $\alpha = 0.08$. The theoretical curves represent the learning threshold boundary derived from the stability condition $m^{(1)}(\xi^1) > 0.9$, using the analytical one-step Mattis magnetization (3.15). The numerical MC data points are obtained by fixing the number of examples $M$ (ranging exponentially from 2 to $2^{12}$) and determining via bisection the critical dataset quality $r$ such that the average one-step Mattis magnetization $m^{(1)}$ reaches the retrieval threshold 0.9. Note the excellent agreement between the theoretical predictions and numerical outcomes across all parameters.

macroscopic behavior can be recovered within the Replica-Symmetric (RS) framework by means of *Guerra's interpolation method* [34], a route which is controlled at every step but which, as we stress below, still rests on an ansatz on the order parameters.[6]

Our main objective in this section is to derive the thermodynamic quenched statistical pressure $\mathcal{A}(\alpha, \beta, r, \rho_L, \rho_U, \lambda)$ in the high-storage regime. Since we are interested in the retrieval phase, we isolate, as in , the first archetype $\xi^1$ as the target pattern to be retrieved, separating its coherent contribution (the signal) from that of the remaining $K - 1$ non-condensed archetypes. The latter act as a source of cross-talk noise, whose handling requires accounting for the structural differences between the supervised

[6] A comment is in order here, since it would be easy to overstate the status of this derivation. Guerra's interpolation differs from the replica trick in that no analytic continuation in the number of replicas is involved: the interpolating measure (4.31) is well defined for every $t \in [0, 1]$, the sum rule (4.29) is an exact identity at finite $N$, and the streaming (4.37) is computed exactly, with no approximation. What the interpolation does *not* establish is the validity of the RS picture itself: the closure of the streaming rests on the assumption (4.30) that the order parameters concentrate on deterministic values in the thermodynamic limit, which is precisely the hypothesis underlying the replica-symmetric solution obtained through the replica trick. In this respect the two routes share the same assumptions on the order parameters, and all the results of this section, phase diagrams included, are to be understood as replica-symmetric. The RS solution of Hopfield-like models is moreover known to become unstable at low temperature and large load, below the de Almeida-Thouless (AT) line [43,44]: the low-temperature portions of the boundaries reported below are therefore to be read as approximate, and we return to this point in the Conclusions.

and unsupervised channels. While the detailed treatment of the signal component is relegated to Appendix B, the analysis of this structured noise is developed in the following subsections.

### 4.1. Treating the noise with Gaussian approximation

When we analyze the Hamiltonian of the model (see Eq. (2.7)), we see a structural difference between the supervised and unsupervised channels. This makes the noise treatment subtle. In short, the supervised channel produces a random field $\bar{\eta}^{\mu}_{i,(L)}$ that is itself a sum of $M_L$ i.i.d. Rademacher random variables, hence amenable to a CLT argument. The unsupervised channel, conversely, exposes the individual random variable $\eta^{\mu,c}_i$, a single Rademacher entry, inside the quadratic form, and the CLT does not apply to it. In this case one resorts to the universality of the quenched noise for spin-glass models established by Carmona and Hu [37] and extended to bipartite systems in [38]. Let us see in detail how this is handled.

We will begin by rewriting the supervised coupling $J^{(L)}_{ij}$ as

$$J^{(L)}_{ij} = \frac{1}{N\Gamma_L}\sum_{\mu=2}^{K}\sum_{a,b=1}^{M_L}\eta^{\mu,a}_i\eta^{\mu,b}_j = \frac{1}{N\mathcal{R}_L}\sum_{\mu=2}^{K}\left(\frac{1}{M_L}\sum_{a=1}^{M_L}\eta^{\mu,a}_i\right)\left(\frac{1}{M_L}\sum_{b=1}^{M_L}\eta^{\mu,b}_j\right)$$
$$= \frac{1}{N\mathcal{R}_L}\sum_{\mu=2}^{K}\bar{\eta}^{\mu}_{i,(L)}\bar{\eta}^{\mu}_{j,(L)}, \tag{4.1}$$

where in the last step we have used that $\Gamma_L = \mathcal{R}_L M_L^2$ and (2.20).

Now, focusing on $\bar{\eta}^{\mu}_{i,(L)}$, we can rewrite it as

$$\bar{\eta}^{\mu}_{i,(L)} = \frac{1}{M_L}\sum_{a=1}^{M_L}\eta^{\mu,a}_i = \xi^{\mu}_i\frac{1}{M_L}\sum_{a=1}^{M_L}\chi^{\mu,a}_i, \tag{4.2}$$

where we have exploited the fact that our examples are now labeled, hence we know the corresponding archetype. For $\mu \geq 2$ the prefactor $\xi^{\mu}_i$ is a Rademacher independent of everything else (and independent across $i$), while the inner empirical mean of i.i.d. Bernoulli's converges, by CLT, to a Gaussian as $M_L \to \infty$. Specifically, by (2.3),

$$\mathbb{E}\Big[\frac{1}{M_L}\sum_{a=1}^{M_L}\chi^{\mu,a}_i\Big] = r, \qquad \mathrm{Var}\Big[\frac{1}{M_L}\sum_{a=1}^{M_L}\chi^{\mu,a}_i\Big] = \frac{1-r^2}{M_L}, \tag{4.3}$$

hence, multiplied by the independent sign $\xi^{\mu}_i$ (whose effect is to recentre the mean to zero for $\mu \geq 2$ because $\mathbb{E}_{\xi}[\xi^{\mu}_i] = 0$), one obtains

$$\bar{\eta}^{\mu}_{i,(L)} \xrightarrow{M_L\to\infty,\,\mu\geq 2} \sqrt{\mathcal{R}_L}\,\tilde{\lambda}^{\mu,(L)}_i, \qquad \tilde{\lambda}^{\mu,(L)}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \tag{4.4}$$

where $\mathcal{R}_L = r^2 + (1-r^2)/M_L$ is the second moment of $\bar{\eta}^{\mu}_{i,(L)}$ defined in (2.15). Under the replacement, the supervised noise term (4.1) becomes

$$J^{(L)}_{ij} \to \frac{1}{N}\sum_{\mu=2}^{K}\tilde{\lambda}^{\mu,(L)}_i\tilde{\lambda}^{\mu,(L)}_j \tag{4.5}$$

which can be replaced in the Hamiltonian of the model (3.1). Now let us focus on the unsupervised term, namely

$$J^{(U)}_{ij} = \frac{1}{N\Gamma_U}\sum_{\mu=2}^{K}\sum_{c=1}^{M_U}\eta^{\mu,c}_i\eta^{\mu,c}_j. \tag{4.6}$$

In the unsupervised channel the random field $\eta^{\mu,c}_i$ is a single Rademacher variable, not a sum, so CLT does not apply directly. However, what we ultimately need is not a distributional replacement of *each* random variable, but a replacement that leaves the *quenched pressure* of the model unchanged in the thermodynamic limit. This is exactly the content of the Carmona-Hu universality theorem for spin glass models [37].

The theorem, extended to bipartite mean-field spin glasses in [38], covers the setting closest to the two-species structure that emerges below. Its application to our unsupervised channel is direct: each $\eta^{\mu,c}_i$ can be considered as a random variable of zero mean (for $\mu \geq 2$) and unit variance, so the quenched pressure is invariant under the replacement

$$\eta^{\mu,c}_i \xrightarrow{\text{Carmona–Hu}} \tilde{\lambda}^{\mu,c}_i, \qquad \tilde{\lambda}^{\mu,c}_i \sim \mathcal{N}(0,1), \quad \mu \geq 2, \tag{4.7}$$

with covariance structure $\mathbb{E}[\tilde{\lambda}^{\mu,c}_i\tilde{\lambda}^{\mu,c'}_i] = \delta_{cc'} + (1-\delta_{cc'})r^2$.

After the replacement, the unsupervised noise becomes

$$J^{(U)}_{ij} \longrightarrow \frac{1}{N\Gamma_U}\sum_{\mu\geq 2}^{K}\sum_{c=1}^{M_U}\tilde{\lambda}^{\mu,c}_i\tilde{\lambda}^{\mu,c}_j \tag{4.8}$$

which can be also reinserted in (3.1).

### 4.2. Linearization of the partition function

After the Gaussian replacements (4.4) and (4.7) the noise contribution in (3.1) is a sum of squares of linear combinations of $\sigma_i$. To convert each square into a linear form, we apply the standard Hubbard Stratonovich (HS) Gaussian identity

$$\exp\Big[\tfrac{1}{2}A\,x^2\Big] = \sqrt{\tfrac{1}{2\pi}}\int_{\mathbb{R}} dz\,\exp\Big[-\tfrac{1}{2}z^2+\sqrt{A}\,z\,x\Big] = \int \mathcal{D}z\,\exp\big[\sqrt{A}\,z\,x\big], \tag{4.9}$$

where $\mathcal{D}z := dz\,e^{-z^2/2}/\sqrt{2\pi}$ is the standard Gaussian measure on $\mathbb{R}$. Then, we apply (4.9) to each $\mu \geq 2$ in the supervised term of (4.5) with $A = \beta\lambda/N$ and $x = \sum_i \tilde{\lambda}_i^{\mu,(L)}\sigma_i$, introducing one Gaussian variable $z_\mu^{(L)}$ per non-condensed archetype, and we get

$$\exp\Big[\frac{\beta\lambda}{2N}\Big(\sum_i \tilde{\lambda}_i^{\mu,(L)}\sigma_i\Big)^2\Big] = \int \mathcal{D}z_\mu^{(L)}\,\exp\Big[\sqrt{\frac{\beta\lambda}{N}}\,z_\mu^{(L)}\sum_i \tilde{\lambda}_i^{\mu,(L)}\sigma_i\Big]. \tag{4.10}$$

Instead, for the unsupervised part, we apply (4.9) to each $(\mu, c)$ with $\mu \geq 2$ and $c = 1, \dots, M_U$ in (3.1), with $A = \beta(1-\lambda)/(N\,r^2)$ (using $\Gamma_U = M_U r^2$, the prefactor reads $\beta(1-\lambda)/(N M_U r^2)$, which is shared by the $M_U$ squares; the $M_U$ factor is absorbed in the per-example HS variable). Introducing one Gaussian variable $z_\mu^c$ per pair $(\mu, c)$,

$$\exp\Big[\frac{\beta(1-\lambda)}{2N\Gamma_U}\Big(\sum_i \tilde{\lambda}_i^{\mu,c}\sigma_i\Big)^2\Big] = \int \mathcal{D}z_\mu^c\,\exp\Big[\sqrt{\frac{\beta(1-\lambda)}{N\Gamma_U}}\,z_\mu^c\sum_i \tilde{\lambda}_i^{\mu,c}\sigma_i\Big]. \tag{4.11}$$

By substituting (4.10) and (4.11) into (3.1), we can rewrite the partition function as

$$\begin{aligned}\mathcal{Z}_{N,\beta}(S) = \sum_{\{\boldsymbol{\sigma}\}}\int\prod_{\mu\geq 2}\mathcal{D}z_\mu^{(L)}\prod_{c=1}^{M_U}\mathcal{D}z_\mu^c\;\exp\Bigg[&\frac{\beta N}{2}\big[\lambda(1+\rho_L)\,n_L^2+(1-\lambda)(1+\rho_U)\,n_U^2\big]\\ &+\sqrt{\frac{\beta\lambda}{N}}\sum_{\mu\geq 2}z_\mu^{(L)}\sum_i\tilde{\lambda}_i^{\mu,(L)}\sigma_i\\ &+\sqrt{\frac{\beta(1-\lambda)}{N\Gamma_U}}\sum_{\mu\geq 2}\sum_c z_\mu^c\sum_i\tilde{\lambda}_i^{\mu,c}\sigma_i\Bigg].\end{aligned} \tag{4.12}$$

where the first term in the exponent is an algebraic rewriting of the signal contribution in terms of the order parameters, and it is computed in Appendix B. This is our starting point for the recovery of the expression of the quenched statistical pressure (2.27) via Guerra's interpolation of the next section.

**Remark 5** (*The Relation Between Databases*)**.**

We highlight that the two databases are made from the same archetypes: the two channels are therefore strongly correlated, and their mutual relationship has to be analyzed as well.

Collecting the cross-covariances we have established, the Gaussian random fields acting on each spin $\sigma_i$ from a non-condensed archetype $\mu \geq 2$ form an $M_U + 1$-dimensional centered Gaussian vector

$$\tilde{\lambda}_i^\mu := \big(\tilde{\lambda}_i^{\mu,(L)},\ \tilde{\lambda}_i^{\mu,1},\ \dots,\ \tilde{\lambda}_i^{\mu,M_U}\big) \in \mathbb{R}^{1+M_U},\ i = 1,\dots,N,\ \mu = 1,\dots,K \tag{4.13}$$

with covariance matrix structured by the labeled vs unlabeled split. The slow noise in (3.1) is governed by the $M \times M$ matrix $SW$, where $W$ is the block-diagonal weight matrix, already defined in (2.23), but rewritten here for simplicity[7]

$$W = \begin{pmatrix} W_L & 0 \\ 0 & W_U \end{pmatrix}, \qquad W_L := \frac{\beta\lambda}{\Gamma_L}\mathbf{1}_{M_L}\mathbf{1}_{M_L}^\top, \quad W_U := \frac{\beta(1-\lambda)}{\Gamma_U}\mathbb{I}_{M_U}, \tag{4.14}$$

and $S$ is the Gram matrix in the example space $\mathbb{R}^M = \mathbb{R}^{M_L} \oplus \mathbb{R}^{M_U}$ governed by[8]

$$S_{ab} = r^2 + (1-r^2)\delta_{ab}, \qquad a, b = 1, \dots, M. \tag{4.15}$$

Although $W$ is block-diagonal, $S$ is not: the rank-one piece $r^2\mathbf{1}_M\mathbf{1}_M^\top$ of (4.15) couples every labeled example to every unlabeled one, so $SW$ acquires non-vanishing off-diagonal blocks and the two channels do *not* decouple. Using $\Gamma_L = M_L^2 r^2(1+\rho_L)$ and $\Gamma_U = M_U r^2$ the four blocks read

$$\begin{aligned}(SW)_{LL} &= \frac{\beta\lambda}{M_L}\,\mathbf{1}_{M_L}\mathbf{1}_{M_L}^\top, & (SW)_{LU} &= \frac{\beta(1-\lambda)}{M_U}\,\mathbf{1}_{M_L}\mathbf{1}_{M_U}^\top,\\ (SW)_{UL} &= \frac{\beta\lambda}{M_L(1+\rho_L)}\,\mathbf{1}_{M_U}\mathbf{1}_{M_L}^\top, & (SW)_{UU} &= \frac{\beta(1-\lambda)}{\Gamma_U}\big[r^2\mathbf{1}_{M_U}\mathbf{1}_{M_U}^\top + (1-r^2)\mathbb{I}_{M_U}\big],\end{aligned} \tag{4.16}$$

The cross-block entries are $O(1/M)$, but since $\|\mathbf{1}_{M_X}\| = \sqrt{M_X}$ their spectral weight is $O(1)$ and they cannot be dropped. The (non-symmetric) matrix $SW$ leaves three mutually orthogonal subspaces of $\mathbb{R}^M$ invariant.

[7] The coefficients appearing in $W$ originate from the Hubbard–Stratonovich transformation.

[8] The appearance of the prefactors $\sqrt{\beta\lambda/\Gamma_L}$ and $\sqrt{\beta(1-\lambda)/\Gamma_U}$ in front of the supervised and unsupervised contributions is justified by the Hubbard–Stratonovich transformation.

- *Unsupervised bulk* (multiplicity $M_U - 1$). For $\mathbf{v} \in \mathbb{R}^{M_U}$ with $\mathbf{1}_{M_U}^\top \mathbf{v} = 0$, the vector $(\mathbf{0}_{M_L}, \mathbf{v})$ is annihilated by $(SW)_{LL}$ and $(SW)_{LU}$, while $(SW)_{UU}(\mathbf{0}_{M_L}, \mathbf{v}) = \frac{\beta(1-\lambda)(1-r^2)}{\Gamma_U}\mathbf{v}$, giving the eigenvalue, with multiplicity $d = M_U - 1$

$$\mu_U^{\text{bulk}} = \frac{\beta(1-\lambda)(1-r^2)}{M_U r^2} = \beta(1-\lambda)\,\rho_U. \tag{4.17}$$

- *Supervised null space* (multiplicity $M_L - 1$). For $\mathbf{u} \in \mathbb{R}^{M_L}$ with $\mathbf{1}_{M_L}^\top \mathbf{u} = 0$ one has $(SW)(\mathbf{u}, \mathbf{0}_{M_U})^\top = \mathbf{0}$, therefore the value 0 is the eigenvalue with multiplicity $d = M_L - 1$.
- *Condensed sector* (the remaining 2 dimensions). The plane spanned by $\boldsymbol{e}_L = (\mathbf{1}_{M_L}, \mathbf{0}_{M_U})$ and $\boldsymbol{e}_U = (\mathbf{0}_{M_L}, \mathbf{1}_{M_U})$ is invariant, $SW\,\boldsymbol{e}_L = \beta\lambda\,\boldsymbol{e}_L + \frac{\beta\lambda}{1+\rho_L}\,\boldsymbol{e}_U$ and $SW\,\boldsymbol{e}_U = \beta(1-\lambda)\,\boldsymbol{e}_L + \beta(1-\lambda)\kappa\,\boldsymbol{e}_U$, where $\kappa := 1 + \rho_U$, so in the basis $\{\boldsymbol{e}_L, \boldsymbol{e}_U\}$ the action is

$$\mathbb{M}_2 = \beta \begin{pmatrix} \lambda & (1-\lambda) \\ \dfrac{\lambda}{1+\rho_L} & (1-\lambda)\kappa \end{pmatrix}, \tag{4.18}$$

with $\det \mathbb{M}_2 = \beta^2\lambda(1-\lambda)\big(\kappa - \frac{1}{1+\rho_L}\big) = \beta^2\lambda(1-\lambda)\big(\rho_U + \frac{\rho_L}{1+\rho_L}\big)$ and eigenvalues

$$\mu_\pm = \frac{\beta}{2}\Big[\lambda + (1-\lambda)\kappa \pm \sqrt{\big(\lambda - (1-\lambda)\kappa\big)^2 + \frac{4\lambda(1-\lambda)}{1+\rho_L}}\,\Big]. \tag{4.19}$$

The dimensions add up, $2 + (M_U - 1) + (M_L - 1) = M$. Writing $\tilde{\mu}_k := \mu_k/\beta$, the spectrum of $SW$ is

$$\{(\tilde{\mu}_k, d_k)\}_{k=1,\dots,M} = \big\{(\tilde{\mu}_+, 1),\ (\tilde{\mu}_-, 1),\ (\tilde{\mu}_U^{\text{bulk}},\ M_U - 1),\ (0,\ M_L - 1)\big\}. \tag{4.20}$$

Being similar to the symmetric $S^{1/2} W S^{1/2}$, $SW$ has a real non-negative spectrum, which is what permits the channel diagonalization used in the interpolation.

Coming back to the definition of the partition function, the problem is that (4.12) contains $1 + M_U$ analog variables $(z_\mu^{(L)};\, z_\mu^1, \dots, z_\mu^{M_U})$ per non-condensed archetype $\mu \geq 2$, coupled to the spins through the quenched random fields $(\tilde{\lambda}_i^{\mu,(L)}, \tilde{\lambda}_i^{\mu,1}, \dots, \tilde{\lambda}_i^{\mu,M_U})$, with the two channel-specific prefactors $\sqrt{\beta\lambda/N}$ and $\sqrt{\beta(1-\lambda)/(N\Gamma_U)}$. It is possible to show now that, exploiting an orthogonal rotation into the eigenbasis of $SW$ as in (4.14) and (4.15), this noise contribution becomes a sum of independent analog variables, with the $\lambda$-dependence absorbed entirely into the spectrum.

Therefore, thanks to this analysis, one can write the noise terms of the semi-supervised model (3.1) in a compact form similar to the construction of the supervised classic case [21].

Indeed, extending the $\boldsymbol{z}_\mu$ vector to $\mathbb{R}^M$ by appending $M_L - 1$ phantom components in the null space of $W_L$ (which carry no coupling to the spins and will play no role) we can write the noise part of (4.12) as

$$\mathcal{B}_\mu^{\text{noise}} = \sqrt{\frac{1}{N}} \sum_{i=1}^{N} \sigma_i \left(\tilde{\mathbf{z}}_\mu\right)^\top W^{1/2}\, \tilde{\boldsymbol{\Lambda}}_i^\mu, \tag{4.21}$$

where $W^{1/2}$ is the positive square root of the weight matrix $W$ defined in (4.14), and $\tilde{\boldsymbol{\Lambda}}_i^\mu \in \mathbb{R}^M$ is the vector of (Gaussian) random fields at site $i$ for archetype $\mu$, with covariance

$$\mathbb{E}\big[\tilde{\Lambda}_{\mu i}^a\, \tilde{\Lambda}_{\mu j}^b\big] = \delta_{ij}\, S_{ab}, \tag{4.22}$$

with $S$ the Gram matrix of (4.15). Eq. (4.21) reduces to (4.12) block-wise: the rank-one piece of $W^{1/2}$ in the $L$-block projects the supervised coupling onto the symmetric direction $\mathbf{1}_L/\sqrt{M_L}$, while the diagonal $U$-block of $W^{1/2}$ is proportional to $\sqrt{\beta(1-\lambda)/\Gamma_U}\,\mathbb{I}_{M_U}$.

Now, let $O$ be the orthogonal $M \times M$ matrix that diagonalizes $W^{1/2} S W^{1/2}$,

$$O^\top (W^{1/2} S W^{1/2})\, O = \operatorname{diag}(\beta\tilde{\mu}_1, \dots, \beta\tilde{\mu}_M), \tag{4.23}$$

with $\{\tilde{\mu}_k\}$ the spectrum in (4.20) (the non-zero spectrum of $W^{1/2} S W^{1/2}$ coincides with that of $SW$). Introducing rotated variables

$$\mathbf{z}_\mu^{\text{rot}} := O^\top \tilde{\mathbf{z}}_\mu, \qquad \boldsymbol{\Lambda}_i^{\text{rot},\mu} := O^\top W^{1/2}\, \tilde{\boldsymbol{\Lambda}}_i^\mu, \tag{4.24}$$

the orthogonal invariance of the Gaussian measure ensures $\mathcal{D}\tilde{\mathbf{z}}_\mu = \mathcal{D}\mathbf{z}_\mu^{\text{rot}}$, while the rotated random fields decouple across $k$ with variance $\beta\tilde{\mu}_k$,

$$\mathbb{E}\big[\Lambda_i^{\text{rot},\mu k}\, \Lambda_j^{\text{rot},\mu k'}\big] = \delta_{ij}\, \delta_{kk'}\, \beta\tilde{\mu}_k. \tag{4.25}$$

Writing $\Lambda_i^{\text{rot},\mu k} = \sqrt{\beta\tilde{\mu}_k}\, \lambda_i^{\mu,k}$ with $\lambda_i^{\mu,k} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ standard, inserting $\mathbb{I} = OO^\top$ and renaming $z_\mu^{\text{rot},k}$ as $z_\mu^k$, the noise (4.21) becomes

$$\mathcal{B}_\mu^{\text{noise}} = \sqrt{\frac{\beta}{N}} \sum_{k:\,\tilde{\mu}_k \neq 0} \sqrt{\tilde{\mu}_k}\, z_\mu^k \sum_{i=1}^{N} \lambda_i^{\mu k} \sigma_i, \qquad \lambda_i^{\mu k} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1). \tag{4.26}$$

The $k$-sum runs only over the $M_U + 1$ eigen-channels with non-zero $\tilde{\mu}_k$: the $M_L - 1$ phantom components corresponding to the supervised null space carry $\sqrt{\tilde{\mu}_k} = 0$ and decouple identically.

Thus, after the rotation, we have independent eigen-modes $z_\mu^k$, one per eigenvalue $\tilde{\mu}_k$ of $SW$ with multiplicity $d_k$, and the natural slow-noise order parameters are the two-replica overlaps

$$p_{ab}^k := \frac{1}{K-1}\sum_{\mu\geq 2}^{K} z_\mu^{k,(a)} z_\mu^{k,(b)} \quad k = 1,\dots,M_U+1 \text{ such that } \tilde{\mu}_k \neq 0. \tag{4.27}$$

### 4.3. Guerra's interpolation

After the discussion in Sections 4.1 and 4.2, the final Hamiltonian under consideration is the following

$$\begin{aligned} -\beta\mathcal{H}_N(\boldsymbol{\sigma}|S,\lambda,\boldsymbol{z}) =& \frac{\beta N}{2}\left[\lambda(1+\rho_L)\,n_L^2 + (1-\lambda)(1+\rho_U)\,n_U^2\right] \\ &+ \sqrt{\frac{\beta}{N}}\sum_{k:\,\tilde{\mu}_k\neq 0}\sqrt{\tilde{\mu}_k}\, z_\mu^k \sum_{i=1}^{N}\lambda_i^{\mu k}\sigma_i. \end{aligned} \tag{4.28}$$

Our goal now is to find the quenched statistical pressure (2.27) of the model in (4.28) using Guerra's interpolation. The strategy is standard [19–21,34,35]: we introduce a one-parameter family of partition functions $\mathcal{Z}_N(t)$, $t\in[0,1]$, interpolating between the original model at $t=1$ and a one-body solvable model at $t=0$. Then, we solve the simpler $t=0$ model exactly, compute the streaming $\partial_t\mathcal{A}(t)$, and recover the original quenched pressure via the fundamental theorem of calculus,

$$\mathcal{A}(t=1) = \mathcal{A}(t=0) + \int_0^1 \partial_t\mathcal{A}(t)\,dt. \tag{4.29}$$

We work within the RS approximation, namely we assume that the relevant order parameters self-average around deterministic values in the thermodynamic limit. More precisely, for any order parameter $X$ with limiting value $\langle X\rangle$, we assume

$$\lim_{N\to\infty}\mathbb{P}_N(X) = \delta(X-\langle X\rangle). \tag{4.30}$$

We will see that under the RS ansatz the integrand in (4.29) becomes $t$-independent so the integral becomes trivial.

For the detailed computation of each of the terms presented above, we refer to Appendix C; in this section we report the main tools and results.

Let $\boldsymbol{\theta} = \{\theta_i\}_{i=1,\dots,N}$ and $\boldsymbol{\psi} = \{\psi_\mu^k\}_{\mu=1,\dots,K}^{k=1,\dots,M}$, where the index $k$ runs over all the eigenvalues of $SW$ and carries the multiplicities $d_k$ of (4.20), be auxiliary standard Gaussian fields, independent of the quenched disorder and of each other, and let $J_m\in\mathbb{R}$ be an external field conjugate to the magnetization $m$. The Guerra interpolating partition function is

$$\mathcal{Z}_N(t,J_m) = \lim_{J_m\to 0}\sum_{\{\boldsymbol{\sigma}\}}\int\prod_{\mu\geq 2}\prod_{k:\,\tilde{\mu}_k\neq 0}\mathcal{D}z_\mu^k\,\mathcal{B}(t,J_m), \tag{4.31}$$

where the interpolating Boltzmann factor $\mathcal{B}(t,J_m)$ reads, after the diagonalization in the eigen-channel basis (4.26),

$$\begin{aligned} \mathcal{B}(t,J_m) = \exp\Bigg\{ & -\frac{1}{2}\sum_{\mu\geq 2}\sum_{k:\,\tilde{\mu}_k\neq 0}\left[1-(1-t)\beta\tilde{\mu}_k(1-\langle q_{12}\rangle)\right](z_\mu^k)^2 \\ & + \sqrt{t}\sqrt{\frac{\beta}{N}}\sum_{\mu\geq 2}\sum_{k:\,\tilde{\mu}_k\neq 0}\sqrt{\tilde{\mu}_k}\, z_\mu^k\sum_i\lambda_i^{\mu,k}\sigma_i \\ & + t\,\frac{\beta N}{2}\left[\lambda(1+\rho_L)\,n_L^2 + (1-\lambda)(1+\rho_U)\,n_U^2\right] \\ & + (1-t)\,\beta\left[\lambda(1+\rho_L)\langle n_L\rangle\, N n_L + (1-\lambda)(1+\rho_U)\langle n_U\rangle\, N n_U\right] \\ & + \sqrt{(1-t)\,\alpha\beta\sum_{k:\,\tilde{\mu}_k\neq 0} d_k\,\tilde{\mu}_k\,\langle p_{12}^k\rangle}\;\sum_i\theta_i\sigma_i \\ & + \sqrt{(1-t)\,\beta\langle q_{12}\rangle}\sum_{\mu\geq 2}\sum_{k:\,\tilde{\mu}_k\neq 0}\sqrt{\tilde{\mu}_k}\,\psi_\mu^k z_\mu^k - J_m\,\beta N m\Bigg\}, \end{aligned} \tag{4.32}$$

where $J_m$ is a conjugate field which allows us to recover the magnetization through

$$\langle m\rangle = -\beta^{-1}\partial_{J_m}\mathcal{A}(t=1,J_m)\Big|_{J_m=0}. \tag{4.33}$$

The interpolating quenched statistical pressure is defined as

$$\mathcal{A}_N(t,J_m) := \frac{1}{N}\,\mathbb{E}\ln\mathcal{Z}_N(t,J_m), \qquad \mathcal{A}_N(t) := \mathcal{A}_N(t,0), \tag{4.34}$$

and the corresponding generalized Boltzmann expectation is

$$\omega_t(\cdot) := \frac{1}{\mathcal{Z}_N(t,J_m)}\sum_{\{\boldsymbol{\sigma}\}}\int\prod\mathcal{D}z\,(\cdot)\,\mathcal{B}(t,J_m), \qquad \langle\cdot\rangle_t := \mathbb{E}[\omega_t(\cdot)]. \tag{4.35}$$

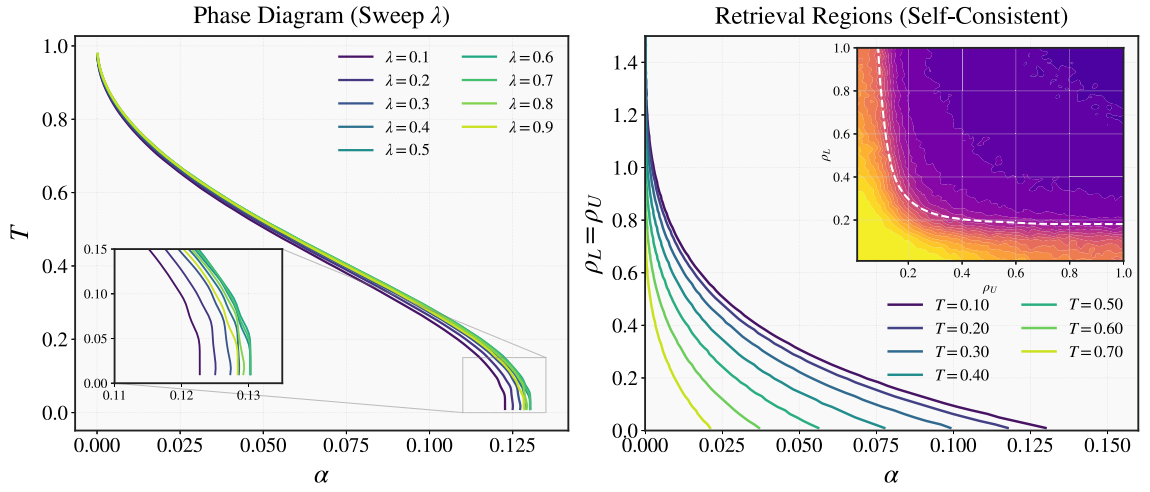



**Fig. 3.** Sensitivity of the retrieval phase to the mixing parameter $\lambda$ and to the dataset noise level $\rho = \rho_L = \rho_U$. Both panels are obtained from the numerical solution of the RS self-consistency equations (Appendix E.3); the inset of the right panel is the only Monte Carlo datum of this figure. *Left panel:* RS retrieval boundary $\alpha_c(T)$ in the $(\alpha, T)$ plane for $r = 0.8$, $M_L = M_U = 30$, and nine values of the mixing parameter $\lambda \in \{0.1, 0.2, \dots, 0.9\}$ (color-coded from dark blue to yellow). The maximum critical load $\alpha_c(T \to 0)$ is not attained at the extremal values $\lambda = 0$ or $\lambda = 1$, but at an intermediate value $\lambda \approx 0.5$–$0.6$, indicating that a mixed semi-supervised coupling genuinely outperforms any pure learning strategy. The inset magnifies the low-temperature region ($T \lesssim 0.15$, $\alpha \in [0.11, 0.135]$), resolving the non-monotone ordering of the curves with respect to $\lambda$. *Right panel:* RS retrieval boundary in the $(\alpha, \rho)$ plane for $r = 0.7$ and fixed mixing parameter $\lambda = 0.5$, at temperatures $T \in \{0.10, 0.20, \dots, 0.70\}$. Each curve separates the retrieval phase (below, high magnetization) from the paramagnetic phase (above): lower temperature and smaller $\rho$ (better dataset quality) expand the retrieval basin. The inset shows a MC heat map of the Mattis magnetization in the $(\rho_U, \rho_L)$ plane (deterministic one-step dynamics at $T = 0$, Appendix E.2), with the white dashed line marking the analytical Signal To Noise stability boundary (3.18); the color encodes the Mattis magnetization $m(\xi^1|\sigma) \in [0, 1]$.

We stress that for $t = 1$ all the one-body terms vanish and we recover the original model in (4.28). Instead, for $t = 0$ we get a new solvable model whose quenched statistical pressure is easier to find (see Appendix C for details):

$$\begin{aligned}\mathcal{A}_{N,\lambda}(t{=}0, J_m) = & -\frac{K}{2N}\sum_k d_k\left[\ln\big(1-\beta\tilde{\mu}_k(1-\langle q_{12}\rangle)\big) - \frac{\beta\tilde{\mu}_k\langle q_{12}\rangle}{1-\beta\tilde{\mu}_k(1-\langle q_{12}\rangle)}\right] + \ln 2 \\ & + \mathbb{E}_{\chi,\theta}\ln\cosh\left[\beta\langle m\rangle\left(\frac{\lambda}{M_L r}\sum_{a=1}^{M_L}\chi^{1,a} + \frac{1-\lambda}{M_U r}\sum_{c=1}^{M_U}\chi^{1,c}\right) - \beta J_m + \theta\sqrt{\frac{K}{N}\beta\sum_k d_k\,\tilde{\mu}_k\,\langle p_{12}^k\rangle}\right].\end{aligned} \tag{4.36}$$

All we need to apply the Fundamental Theorem of Calculus is the computation of the integral in (4.29), which reads

$$\begin{aligned}\partial_t\mathcal{A}_N(t, J_m) = & \frac{\beta}{2}\sum_{X\in\{L,U\}} c_X(1+\rho_X)\big(\Delta[n_X^2] - \langle n_X\rangle^2\big) \\ & - \frac{K\beta}{2N}\sum_k d_k\,\tilde{\mu}_k\big(\Delta[p_{12}^k q_{12}] + \langle p_{12}^k\rangle(1-\langle q_{12}\rangle)\big),\end{aligned} \tag{4.37}$$

where $c_L = \lambda$, $c_U = 1-\lambda$, and we have introduced the fluctuations of $n_X$ and of $p_{12}^k q_{12}$ as

$$\Delta[n_X^2] := \mathbb{E}\langle(n_X - \langle n_X\rangle)^2\rangle_t, \qquad \Delta[p_{12}^k q_{12}] := \mathbb{E}\langle(p_{12}^k - \langle p_{12}^k\rangle)(q_{12} - \langle q_{12}\rangle)\rangle_t. \tag{4.38}$$

Under the RS ansatz (4.30) in the thermodynamic limit every centered fluctuation in (4.38) vanishes, so $\Delta[n_X^2] \to 0$ and $\Delta[p_{12}^k q_{12}] \to 0$. The streaming becomes $t$-independent:

$$\partial_t\mathcal{A}(t, J_m) = -\frac{\beta}{2}\big[\lambda(1+\rho_L)\langle n_L\rangle^2 + (1-\lambda)(1+\rho_U)\langle n_U\rangle^2\big] - \frac{\alpha\beta}{2}(1-\langle q_{12}\rangle)\sum_k d_k\,\tilde{\mu}_k\,\langle p_{12}^k\rangle, \tag{4.39}$$

where we have used also the definition of the load of the network in (2.12). Since $\partial_t\mathcal{A}$ is $t$-independent, the integral $\int_0^1 \partial_t\mathcal{A}\,dt$ in (4.29) equals (4.39) directly. Also in this case, for all the computations we refer to Appendix C.

Applying the definition of the load (2.12) also to the one-body terms, we finally obtain the following expression for the quenched statistical pressure of the model defined by the Hamiltonian (2.7):

$$\begin{aligned}\mathcal{A}_{\mathrm{RS}} = & \lim_{J_m\to 0}\mathcal{A}_{\mathrm{RS}}(J_m) \\ = & \lim_{J_m\to 0}\left\{-\frac{\alpha}{2}\sum_k d_k\left[\ln\big(1-\beta\tilde{\mu}_k(1-\langle q\rangle)\big) - \frac{\beta\tilde{\mu}_k\langle q\rangle}{1-\beta\tilde{\mu}_k(1-\langle q\rangle)}\right]\right.\end{aligned}$$

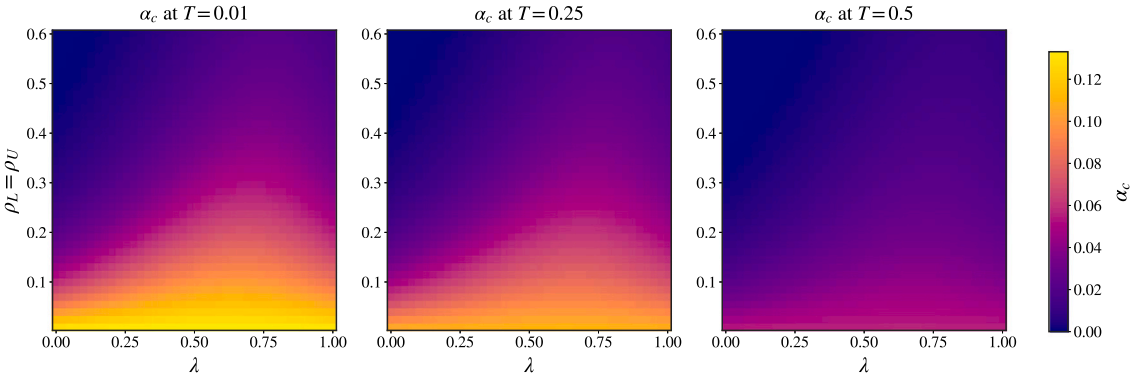



**Fig. 4.** Heat map of the critical retrieval load $\alpha_c$ as a function of the mixing parameter $\lambda \in [0,1]$ and the rescaled noise level $\rho = \rho_L = \rho_U \in [0.01, 0.60]$, computed from the RS self-consistency equations (Appendix E.3) for $M_L = M_U = 30$ at three temperatures $T \in \{0.01, 0.25, 0.50\}$. In all panels the color encodes the maximal load $\alpha_c$: high-capacity configurations appear bright (yellow), low-capacity ones dark (blue). As expected, increasing $\rho$ (deteriorating dataset quality) reduces $\alpha_c$ uniformly, while increasing temperature contracts the retrieval region. The key observation, visible across all temperatures, is that for any fixed $\rho > 0$ the maximum of $\alpha_c$ over $\lambda$ is attained at an *interior* value $\lambda \in (0,1)$, strictly away from both the purely unsupervised ($\lambda = 0$) and purely supervised ($\lambda = 1$) extremes. Quantitatively, at $T = 0.01$ and $\rho = 0.01$, one finds $\alpha_c(\lambda = 0) \approx 0.127$, $\alpha_c(\lambda = 1) \approx 0.131$, and $\alpha_c(\lambda \approx 0.59) \approx 0.133$: a balanced mixed strategy outperforms both pure channels.

$$- \frac{\beta}{2}\left[\lambda(1+\rho_L)\langle n_L\rangle^2 + (1-\lambda)(1+\rho_U)\langle n_U\rangle^2\right] - \frac{\alpha\beta}{2}(1-\langle q\rangle)\sum_k d_k\tilde{\mu}_k\langle p_{12}^k\rangle + \ln 2$$
$$+ \mathbb{E}_{\chi,\theta}\ln\cosh\left[\beta\left(\lambda\langle n_L\rangle\frac{1}{M_L r}\sum_a \chi^{1a} + (1-\lambda)\langle n_U\rangle\frac{1}{M_U r}\sum_c \chi^{1c}\right) - \beta J_m + \theta\sqrt{\alpha\beta\sum_k d_k\tilde{\mu}_k\langle p_{12}^k\rangle}\right]\Bigg\}. \tag{4.40}$$

The order parameters $\langle m\rangle, \langle n_L\rangle, \langle n_U\rangle, \langle q_{12}\rangle, \langle p_{12}^k\rangle$ are determined by extremizing the RS statistical pressure (4.40) with respect to each of them. We compute the derivatives one by one in Appendix C. Here, we report the self-consistency equations in the large dataset limit $M_X \to \infty$, for $X \in \{L, U\}$ (see Appendix C.1):

$$\langle m\rangle = \mathbb{E}_z\left[\tanh\left(\beta\langle m\rangle + \beta z\sqrt{\alpha\sum_k d_k\frac{\tilde{\mu}_k^2\langle q_{12}\rangle}{\left[1-\beta\tilde{\mu}_k(1-\langle q_{12}\rangle)\right]^2} + \langle m\rangle^2\left(\lambda^2\rho_L + (1-\lambda)^2\rho_U\right)}\right)\right], \tag{4.41}$$

$$\langle q_{12}\rangle = \mathbb{E}_z\left[\tanh^2\left(\beta\langle m\rangle + \beta z\sqrt{\alpha\sum_k d_k\frac{\tilde{\mu}_k^2\langle q_{12}\rangle}{\left[1-\beta\tilde{\mu}_k(1-\langle q_{12}\rangle)\right]^2} + \langle m\rangle^2\left(\lambda^2\rho_L + (1-\lambda)^2\rho_U\right)}\right)\right], \tag{4.42}$$

$$\langle p_{12}^k\rangle = \frac{\beta\tilde{\mu}_k\langle q_{12}\rangle}{\left[1-\beta\tilde{\mu}_k(1-\langle q_{12}\rangle)\right]^2}, \tag{4.43}$$

$$\langle n_L\rangle = \langle n_U\rangle = \langle m\rangle. \tag{4.44}$$

To extract the retrieval boundary and the structure of the phase diagram, we specialize the large-dataset RS self-consistency system (4.41)–(4.43) (where $\langle n_L\rangle = \langle n_U\rangle = \langle m\rangle$) to the zero-temperature regime $\beta \to \infty$, following Amit's approach [5]. In this limit the spin-glass overlap saturates, $\langle q\rangle \to 1$, while the combination $\mathcal{Q} := \beta(1-\langle q\rangle)$ stays finite. The two surviving order parameters $\langle m\rangle$ and $\mathcal{Q}$ are then fixed by the closed zero-temperature self-consistency system (derived in detail in Appendix C.2)

$$\langle m\rangle = \mathrm{erf}\left(\frac{\langle m\rangle}{\sqrt{2\left(\alpha\sum_k \frac{d_k\tilde{\mu}_k^2}{(1-\tilde{\mu}_k\mathcal{Q})^2} + \langle m\rangle^2(\lambda^2\rho_L + (1-\lambda)^2\rho_U)\right)}}\right), \tag{4.45}$$

$$\mathcal{Q} = \sqrt{\frac{2}{\pi\left(\alpha\sum_k \frac{d_k\tilde{\mu}_k^2}{(1-\tilde{\mu}_k\mathcal{Q})^2} + \langle m\rangle^2(\lambda^2\rho_L + (1-\lambda)^2\rho_U)\right)}} \cdot \exp\left[-\frac{\langle m\rangle^2}{2\left(\alpha\sum_k \frac{d_k\tilde{\mu}_k^2}{(1-\tilde{\mu}_k\mathcal{Q})^2} + \langle m\rangle^2(\lambda^2\rho_L + (1-\lambda)^2\rho_U)\right)}\right]. \tag{4.46}$$

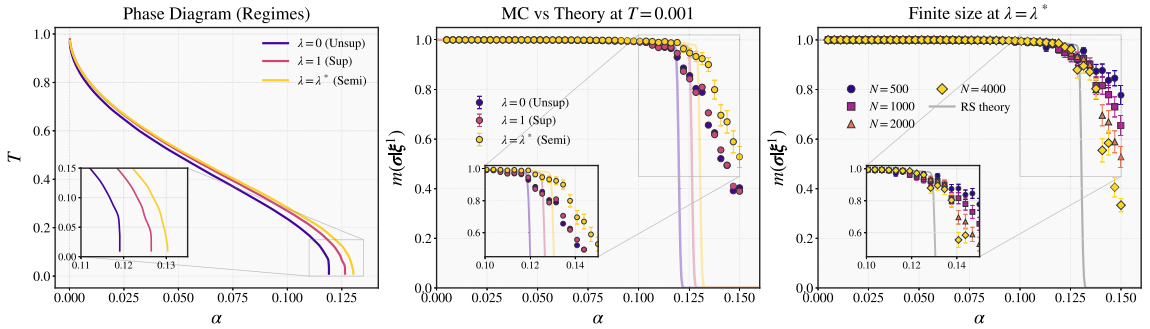



**Fig. 5.** Phase diagrams and retrieval performance of the semi-supervised Hopfield model for $r = 0.8$, $M_L = M_U = 30$. *Left panel:* retrieval boundary $\alpha_c(T)$ in the $(\alpha, T)$ plane for the three representative learning regimes: fully unsupervised ($\lambda = 0$), fully supervised ($\lambda = 1$), and semi-supervised with the pointwise-optimal mixing $\lambda = \lambda^*$ from (5.17), obtained from the numerical solution of the RS self-consistency equations (Appendix E.3). The optimal semi-supervised boundary lies strictly to the right of both limiting cases, demonstrating that the enlarged retrieval region is a genuine effect of the mixed Hebbian learning rule. The inset magnifies the low-temperature region ($T \lesssim 0.15$, $\alpha \in [0.11, 0.135]$), resolving the separation between the three curves. *Central panel:* Mattis magnetization $m = m(\sigma|\xi^1)$ as a function of the storage load $\alpha$ at near-zero temperature ($T = 0.001$), for the same three regimes. Solid lines are replica-symmetric (RS) theoretical predictions; symbols with error bars are equilibrium Glauber Monte Carlo estimates (Appendix E.2) ($N = 2000$, 50 disorder realizations). Both the analytical curves and the numerical data display a sharp, collective drop of $m$ within a narrow window of $\alpha$, signaling a thermodynamic phase transition. The inset zooms into the transition region ($\alpha \in [0.10, 0.15]$, $m \in [0.45, 1.0]$), highlighting the quantitative agreement between theory and simulation. *Right panel:* finite-size analysis of the same transition at the optimal mixing $\lambda = \lambda^*$. The equilibrium Monte Carlo protocol of the central panel is repeated for the four network sizes $N = 500, 1000, 2000, 4000$, with identical dataset parameters, identical thermal parameters, identical statistics ($n_{\text{real}} = 50$) and the same grid of loads; the curve at $N = 2000$ is the semi-supervised curve of the central panel itself, and the gray line is the corresponding replica-symmetric prediction. Away from the transition the four curves are indistinguishable within the error bars, whereas inside the transition region the drop steepens monotonically with the size: at the largest load explored, $\alpha = 0.150$, the disorder-averaged magnetization decreases from 0.78 at $N = 500$ to 0.66, 0.53 and 0.33 as the size is doubled, the replica-symmetric boundary lying at $\alpha_c^{\text{RS}}(\lambda^*) \simeq 0.129$. The inset magnifies the same window as the central panel.

Equation (4.45) is the zero-temperature retrieval equation, whose non-trivial solution $\langle m \rangle > 0$ delimits the retrieval phase, while (4.46) closes the system by fixing $\mathcal{Q} = \beta(1 - \langle q \rangle)$. We solve this system numerically to obtain the phase diagram of Fig. 3; evaluating the resulting critical retrieval load over the $(\lambda, \rho)$ plane, with $\rho := \rho_L = \rho_U$, then yields the capacity map of Fig. 4. The numerical procedure, namely the damped iteration of the self-consistency system and the continuation in the load used to locate $\alpha_c$, is detailed in Appendix E.3.

We stress once more that the boundaries obtained in this way are replica-symmetric: they are exact within the ansatz (4.30), and they are expected to be only approximate in the low-temperature, high-load corner of the phase diagram, where the RS solution of Hopfield-like models is known to be unstable [43,44]. Locating the AT line of the present model, and thereby delimiting the region in which the diagrams below are quantitatively reliable, is left for future work.

## 5. Optimal mixing parameter

Asking which mixing $\lambda \in [0, 1]$ is *optimal* is ambiguous until one declares by which principle the selection is made, and the two natural principles pull in opposite directions. On the one hand one may treat $\lambda$ on the same footing as the other macroscopic variables and let equilibrium thermodynamics select it, i.e. look for the stationary point of the quenched pressure along the mixing family: this is a *global* criterion, blind to which minimum the system is sitting in. On the other hand one may keep $\lambda$ for what it operationally is, a hyperparameter of the synaptic rule (2.6), fixed by the experimenter before the dynamics starts, and tune it so as to maximize the *local* stability of the target pattern, hence the critical load: this is the criterion implicit in the signal-to-noise analysis of Section 3. The two routes do not select the same value, and the discrepancy is not a technical accident: as we show below, thermodynamics alone *refuses* the mixture altogether. We denote by $\lambda_{\text{eq}}$ the thermodynamic stationarity point and by $\lambda^\star$ the retrieval-optimal mixing, and we address them in turn. Throughout this section we work in the large dataset limit $M_L, M_U \to \infty$, in which the self-consistency system reduces to (4.41)–(4.43).

*Thermodynamic stationarity*

Let $\Omega := \{\langle m \rangle, \langle q_{12} \rangle, \langle p_{12}^1 \rangle, \langle p_{12}^2 \rangle, \dots, \langle p_{12}^{M_U+1} \rangle\}$ denote the set of RS order parameters. The pressure $\mathcal{A} = \mathcal{A}(\lambda, \Omega)$ depends on $\lambda$ both explicitly and implicitly through the elements of $\Omega$, which are fixed by the self-consistency Eqs. (4.41)–(4.43). Its total derivative with respect to $\lambda$ is

$$\frac{d\mathcal{A}}{d\lambda} = \frac{\partial \mathcal{A}}{\partial \lambda} + \sum_{O \in \Omega} \frac{\partial \mathcal{A}}{\partial O} \frac{\partial O}{\partial \lambda}. \tag{5.1}$$

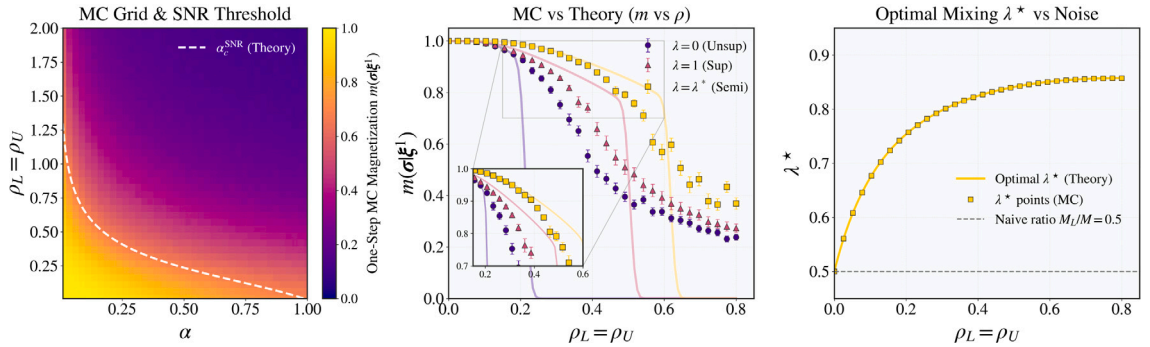



**Fig. 6.** MC validation of the retrieval phase boundaries in the $(\alpha, \rho)$ plane. *Left panel:* heat map of the one-step Mattis magnetization $m(\boldsymbol{\sigma}|\boldsymbol{\xi}^1)$ on a $40 \times 40$ grid of $(\alpha, \rho)$ values ($N = 2000$, $M_L = M_U = 30$, $\lambda = 0.5$, zero-temperature deterministic one-step dynamics, Appendix E.2, 32 disorder realizations). The color encodes $m \in [0, 1]$ (plasma scale); the white dashed curve is the analytical SNR threshold $\alpha_c^{\mathrm{SNR}}(\rho)$ from (5.9), which faithfully tracks the numerically observed boundary between high-$m$ and low-$m$ regions across the entire parameter space. *Center panel:* magnetization $m(\boldsymbol{\sigma}|\boldsymbol{\xi}^1)$ as a function of the rescaled noise parameter $\rho$ at fixed load $\alpha = 0.02$ and near-zero temperature ($T = 0.001$), for the three regimes: fully unsupervised ($\lambda = 0$, circles), fully supervised ($\lambda = 1$, triangles), and semi-supervised with the pointwise-optimal mixing $\lambda = \lambda^*(\alpha, r)$. Solid lines are RS theoretical predictions (Appendix E.3); symbols with error bars are equilibrium Glauber Monte Carlo data (Appendix E.2) ($N = 2000$, $M_L = M_U = 30$, 50 disorder realizations). As $\rho$ increases (dataset quality deteriorates), all curves decay, but the optimal semi-supervised strategy maintains a higher magnetization over the entire range. The inset magnifies the high-magnetization region ($\rho \in [0.15, 0.6]$, $m \in [0.7, 1.0]$). *Right panel:* closed-form optimal mixing parameter $\lambda^\star$ as a function of $\rho_L = \rho_U$, evaluated directly from (5.17). The solid curve denotes the exact stationary root $\lambda^\star = 1 - x^\star$ from (5.17); square markers indicate the discrete values evaluated in the MC simulations. Notably, at perfect dataset quality $\rho_L = \rho_U = 0$ (corresponding to $r = 1$), within-class noise vanishes and the system enters the degenerate regime where the optimal mixing reduces to the empirical fraction of labeled examples, $\lambda^\star\big|_{r\to 1} = M_L/(M_L + M_U) = 0.5$ (see Remark 6). For any $r < 1$ ($\rho_L = \rho_U > 0$), noise corrections detach $\lambda^\star$ from the naive ratio 0.5, smoothly increasing it towards $\sim 0.86$ to adaptively suppress unlabeled interference.

The self-consistency equations are precisely the stationarity conditions $\partial \mathcal{A}/\partial O = 0$, so every term in the implicit sum vanishes and only the explicit $\lambda$-dependence survives[9]:

$$\frac{d\mathcal{A}}{d\lambda} = \frac{\partial \mathcal{A}}{\partial \lambda}\bigg|_{\Omega \text{ fixed}}. \tag{5.2}$$

The $\lambda$−gradient of the pressure is therefore controlled by the explicit $\lambda$−dependence of the quenched statistical pressure $\mathcal{A}$ alone, with the order parameters frozen at their self-consistent values.

We remind that we can write the Boltzmann weight as in Eq. (2.18):

$$-\beta \mathcal{H}(\boldsymbol{\sigma}) = \lambda\, \beta N\, \mathcal{E}_L(\boldsymbol{\sigma}) + (1-\lambda)\, \beta N\, \mathcal{E}_U(\boldsymbol{\sigma}), \tag{5.3}$$

and $\mathcal{E}_X$, for $X \in \{L, U\}$, were already defined in Eq. (2.19).

Differentiating (5.3) at fixed $\boldsymbol{\sigma}$ gives $\partial_\lambda(-\beta\mathcal{H}) = \beta N[\mathcal{E}_L - \mathcal{E}_U]$, so by the identity (5.2),

$$\frac{\partial \mathcal{A}}{\partial \lambda} = \beta \lim_{N\to\infty} \mathbb{E}\, \omega\big(\mathcal{E}_L - \mathcal{E}_U\big). \tag{5.4}$$

The stationarity condition $\partial \mathcal{A}/\partial \lambda = 0$ therefore acquires a transparent physical content: the thermodynamically stationary mixing $\lambda_{\mathrm{eq}}$ equipartitions the macroscopic energy extracted from the labeled and unlabeled datasets,

$$\mathbb{E}\, \omega(\mathcal{E}_L) = \mathbb{E}\, \omega(\mathcal{E}_U). \tag{5.5}$$

The variational status of this condition, whether $\lambda_{\mathrm{eq}}$ is a maximum or a minimum of the pressure, and whether it is unique, is completely fixed by an elementary but consequential structural fact: the Hamiltonian (2.18) is convex in $\lambda$, so that $\lambda \mapsto N^{-1} \ln \mathcal{Z}_{N,\beta}$ has second derivative

$$\frac{\partial^2}{\partial \lambda^2} \frac{\ln \mathcal{Z}_{N,\beta}}{N} = \beta^2 N\, \mathrm{Var}_\omega\big(\mathcal{E}_L - \mathcal{E}_U\big) \;\geq\; 0, \tag{5.6}$$

a Gibbs variance, hence non-negative for every $N$, every $\beta > 0$ and every realization of the dataset $S$. Convexity is preserved by the quenched average and by pointwise limits, so both $\mathcal{A}_N$ and $\mathcal{A}$ are convex in the mixing parameter.[10] The consequences we need are collected, with their proofs, in Appendix D; here we only recall them and comment on their physical content.

[9] This stationarity property of the RS variational free energy is standard in the statistical mechanics of disordered systems [6,10,45]: the total derivative of the quenched pressure with respect to a control parameter reduces to its explicit part because the order parameters sit at a saddle point. It is the finite-temperature, classical counterpart of the Hellmann–Feynman theorem of quantum mechanics [46–48], where the derivative of an eigenvalue with respect to a parameter of the Hamiltonian $\mathcal{H}$ equals the expectation of $\partial_\lambda \mathcal{H}$ on the corresponding eigenstate, the first-order variation of the state itself dropping out by normalization and stationarity.

First, the derivative $d\mathcal{A}/d\lambda = \beta\,\mathbb{E}\,\omega(\mathcal{E}_L - \mathcal{E}_U)$ is non-decreasing, so the equipartition condition (5.5) has a solution set which is a closed interval, reduced to the single point $\lambda_{\text{eq}}$ whenever the convexity is strict: the energy balance identifies a well-defined mixing, and no spurious multiplicity of thermodynamic optima can arise. Second, and decisively, an interior solution of (5.5) is a global *minimum* of $\lambda \mapsto \mathcal{A}(\beta, \lambda)$, i.e. a global *maximum* of the quenched free energy; equivalently, by the inequality

$$\mathcal{A}(\beta, \lambda) \;\le\; \lambda\,\mathcal{A}(\beta, 1) + (1-\lambda)\,\mathcal{A}(\beta, 0), \tag{5.7}$$

the pressure of the mixture never exceeds the interpolation of the two pure channels. The mixed protocol is thus, thermodynamically speaking, the *least* favorable one: a selection principle based on relaxation to the lowest free energy at fixed temperature would push the synaptic rule towards the pure channels $\lambda \in \{0, 1\}$ and would never stop at an interior point. Equilibrium thermodynamics, by itself, does not select semi-supervised learning, it rejects it. This is the precise sense in which the convexity of $\mathcal{A}$ plays a "no-go" role: it forbids reading $\lambda$ as an order parameter that the system chooses spontaneously, and it thereby provides the formal justification for treating it as what it operationally is, an externally imposed hyperparameter of the learning rule.

Once $\lambda$ is recognized as a hyperparameter, the objective changes accordingly. One is no longer interested in the global free-energy landscape, but in the stability of one specific minimum, the one aligned with the target pattern. The natural figure of merit is then the local signal-to-noise ratio of Section 3: the signal $S(\lambda)$ against the total noise variance $V(\lambda) = C(\lambda)\alpha + R(\lambda)$, where $C$ collects the cross-pattern interference and $R$ the within-class fluctuations of the examples. These two contributions compete. The interference term alone is minimized by pure supervision, $C'(\lambda) < 0$ on $(0, 1)$, and would drive $\lambda \to 1$; but $R(\lambda)$ is convex and penalizes both extremes of the mixing range, since discarding either channel throws away the averaging that suppresses the within-class noise. It is exactly this trade-off, absent from the pressure, which only sees the total energy balance, that places the noise minimum, hence the capacity maximum, at an interior $\lambda^\star \in (0, 1)$. Far from diminishing the role of $\lambda$, the dichotomy confirms it as a genuine hyperparameter: its optimal tuning encodes information on the dataset (quality $r$ and per-channel sizes $M_L, M_U$) that the free energy alone cannot see. Note also that, although exact, the equipartition characterization (5.5) is not usable in practice to locate any mixing in closed-form, which makes the passage to the SNR observables necessary as well as conceptually mandatory.

*Retrieval capacity maximization*

We recall that the one-step magnetization takes the form

$$m^{(1)}(\xi^1) = \operatorname{erf}\left( \frac{\dfrac{\lambda}{1+\rho_L} + (1-\lambda)}{\sqrt{2(C(\lambda, r, \rho_L, \rho_U)\alpha + R(\lambda, r, \rho_L, \rho_U))}} \right). \tag{5.8}$$

Imposing the usual condition $m^{(1)}(\xi^1) > \operatorname{erf}(\theta)$ with the canonical choice $\theta = 1/\sqrt{2}$, we get an estimate for the critical load:

$$\alpha_c^{\text{SNR}} = \frac{(S(\lambda, r, \rho_L, \rho_U))^2 - R(\lambda, r, \rho_L, \rho_U)}{C(\lambda, r, \rho_L, \rho_U)}. \tag{5.9}$$

where $S(\lambda, r, \rho_L, \rho_U)$, $R(\lambda, r, \rho_L, \rho_U)$ and $C(\lambda, r, \rho_L, \rho_U)$ are given in Eqs. (3.11), (3.13) and (3.14). Since $\alpha_c^{\text{SNR}}$ is a physical capacity only when non-negative, we determine the optimal mixing as the constrained maximizer

$$\lambda^\star = \arg\max_{\lambda \in [0,1]} \alpha_c^{\text{SNR}}(\lambda) \qquad \text{subject to} \qquad \alpha_c^{\text{SNR}}(\lambda) \ge 0. \tag{5.10}$$

The stationarity condition $\partial_\lambda \alpha_c^{\text{SNR}} = 0$ contains cubic terms which cancel identically, so that it reduces to the quadratic equation (having set $x := 1 - \lambda$)

$$\mathfrak{A}\,x^2 - \mathfrak{B}\,x + \mathfrak{C} = 0, \tag{5.11}$$

with coefficients:

$$\mathfrak{A} = \eta\big[\rho_U + \eta(1-2\eta) - 2(1-\eta)(\bar{\rho}_U + 2\eta)\big], \tag{5.12}$$

$$\mathfrak{B} = \rho_U + \eta(1-2\eta) + (1-\eta)(1-2\eta)(\bar{\rho}_U + 2\eta), \tag{5.13}$$

$$\mathfrak{C} = \eta(1-\eta)(3-2\eta), \tag{5.14}$$

$$\eta = \frac{\rho_L}{1+\rho_L}, \tag{5.15}$$

$$\bar{\rho}_U = \frac{1-r^4}{M_U r^4} = \frac{1+r^2}{r^2}\,\rho_U. \tag{5.16}$$

[10] Here and in Appendix D the energy densities of (2.19) are normalized so that $-\beta\mathcal{H} = \beta N[\lambda\mathcal{E}_L + (1-\lambda)\mathcal{E}_U]$ is extensive, i.e. $\mathcal{E}_L = \frac{1}{2r_L}\sum_\mu \big(\sum_a n_L^{\mu a}\big)^2$ and $\mathcal{E}_U = \frac{1}{2r_U}\sum_{\mu,c}(n_U^{\mu c})^2$.

Here $\bar{\rho}_U$ is the quartic counterpart of the rescaled size $\rho_U$ introduced in (2.15). The physical root, i.e. the one lying in $[0,1]$, is

$$\lambda^\star = 1 - x^\star, \qquad x^\star = \frac{2\,\mathfrak{C}}{\mathfrak{B} + \sqrt{\mathfrak{B}^2 - 4\mathfrak{A}\mathfrak{C}}}. \tag{5.17}$$

Since $C(\lambda) > 0$, the constraint in (5.10) reduces to $\alpha_c^{\mathrm{SNR}} \geq 0 \Leftrightarrow S^2 \geq R$ and never binds at an interior optimum: either $\alpha_c^{\mathrm{SNR}}(\lambda^\star) > 0$, and (5.17) is the maximizer, or $\alpha_c^{\mathrm{SNR}}(\lambda^\star) \leq 0$, in which case $\alpha_c^{\mathrm{SNR}}(\lambda) \leq 0$ for all $\lambda$ and no retrieval solution exists. The equality $S^2(\lambda^\star) = R(\lambda^\star)$ therefore fixes a critical quality $r_{\min}(M_L, M_U)$.

An explicit expression for the threshold follows from the fact that, at the boundary of the retrieval region, the constrained maximum vanishes, $\alpha_c^{\mathrm{SNR}}(\lambda^\star) = 0$ together with $\partial_\lambda \alpha_c^{\mathrm{SNR}} = 0$. Since $\alpha_c^{\mathrm{SNR}} = (S^2 - R)/C$ with $C > 0$, this forces the numerator $\mathcal{G}(\lambda) = S^2 - R$ to develop a double root. Being quadratic in $\lambda$,

$$\mathcal{G}(\lambda) = n_0 + n_1\lambda + n_2\lambda^2, \qquad n_0 = 1 - \rho_U, \quad n_1 = 2(\rho_U - \eta), \quad n_2 = \eta(2\eta - 1) - \rho_U, \tag{5.18}$$

its discriminant reads $\frac{1}{4}\Delta_\lambda = (1-\eta)\big[\eta - \rho_U(2\eta - 1)\big]$ (note that $\mathcal{G}$ is independent of $\bar{\rho}_U$, so the threshold does not depend on the load or on the interference structure). Setting $\Delta_\lambda = 0$ and using $\eta = \rho_L/(1+\rho_L)$, the boundary reduces to the symmetric condition

$$\frac{1}{\rho_L} + \frac{1}{\rho_U} = 1 \qquad \Longleftrightarrow \qquad \rho_L\rho_U = \rho_L + \rho_U. \tag{5.19}$$

Inserting $\rho_{L,U} = (1 - r^2)/(M_{L,U}\, r^2)$ gives $(M_L + M_U)\, r^2 = 1 - r^2$, i.e. the closed-form retrieval threshold

$$r_{\min} = \frac{1}{\sqrt{M_L + M_U + 1}}. \tag{5.20}$$

For $r < r_{\min}$ the within-class fluctuations $\rho_{L,U} \sim 1/r^2$ overcome the signal and no retrieval solution exists; the optimal mixing $\lambda^\star$ is therefore defined only for $r > r_{\min}$.

**Remark 6.** At perfect dataset quality $r \to 1$ the two channels become noiseless, $\rho_L, \rho_U \to 0$ (hence $\eta \to 0$) and $\bar{\rho}_U \to 0$, so that all the coefficients of the stationarity quadratic (5.11) vanish simultaneously, $\mathfrak{A}, \mathfrak{B}, \mathfrak{C} \to 0$, and the closed-form (5.17) becomes indeterminate. This is not a divergence but a degeneracy: at $r = 1$ one has $S = 1$, $C = 1$ and $R = 0$ for every $\lambda$, so the capacity is flat,

$$\alpha_c^{\mathrm{SNR}}(\lambda) = \frac{S^2 - R}{C} = 1 \qquad \text{for all } \lambda \in [0,1], \tag{5.21}$$

and the maximizer is not unique: every mixing is optimal.

The degeneracy is lifted at first order in $\varepsilon := 1 - r^2$. Using $\eta \simeq \varepsilon/M_L$, $\rho_U \simeq \varepsilon/M_U$ and $\bar{\rho}_U \simeq 2\varepsilon/M_U$, one finds $\mathfrak{A} = O(\varepsilon^2)$ while $\mathfrak{B}, \mathfrak{C} = O(\varepsilon)$, so that

$$x^\star = \frac{2\mathfrak{C}}{\mathfrak{B} + \sqrt{\mathfrak{B}^2 - 4\mathfrak{A}\mathfrak{C}}} \xrightarrow{r\to 1} \frac{\mathfrak{C}}{\mathfrak{B}} = \frac{3\eta}{\rho_U + \bar{\rho}_U + 3\eta} \longrightarrow \frac{M_U}{M_L + M_U}, \tag{5.22}$$

and therefore

$$\lambda^\star\Big|_{r\to 1} = \frac{M_L}{M}. \tag{5.23}$$

Thus $\lambda^\star(r)$ is continuous and finite up to $r = 1$, where it reduces to the empirical fraction of labeled examples. Remarkably, this is the only regime in which the optimal synaptic weight coincides with the naive proportion $M_L/M$ that, as stressed in Remark 3, does not in general control the learning rule: when the dataset is ideal the within-class noise balance vanishes and only the counting of examples survives, whereas for any $r < 1$ the $O(1 - r^2)$ corrections detach $\lambda^\star$ from $M_L/M$.

The retrieval consequences of the optimal mixing $\lambda^\star$ are illustrated in Figs. 5 and 6. In the $(\alpha, T)$ plane (Fig. 5, left) the retrieval boundary evaluated at $\lambda = \lambda^\star$ lies strictly to the right of both pure channels at every temperature, so the mixed rule genuinely enlarges the retrieval region; the near-zero-temperature magnetization curves (center) exhibit a sharp retrieval-to-spin-glass transition, with the RS prediction and the MC data collapsing onto the same critical load. That this drop is the finite-size signature of a thermodynamic transition, rather than a crossover, is supported by the finite-size analysis of the right panel, where the same Monte Carlo protocol is repeated at $\lambda = \lambda^\star$ for four network sizes and the drop is seen to steepen monotonically with $N$ around the RS critical load. The robustness of this picture against dataset quality is summarized in Fig. 6: the one-step magnetization heat map (left) shows that the closed-form SNR threshold $\alpha_c^{\mathrm{SNR}}(\rho)$ of (5.9) faithfully tracks the numerically observed boundary across the whole $(\alpha, \rho)$ plane, while at fixed load (center) the optimal mixture sustains a higher magnetization than either pure channel as the noise $\rho$ grows. Finally (right), $\lambda^\star(\rho)$ detaches from the naive labeled fraction $M_L/M = 1/2$ as soon as $\rho > 0$ and increases monotonically towards $\simeq 0.86$: the network adaptively shifts weight onto the supervised channel to suppress the diverging unlabeled interference, confirming that $\lambda^\star$ is a genuine, dataset-dependent hyperparameter rather than a proxy for the proportion of labels.

## 6. Conclusions and outlooks

In this paper we introduced and solved a semi-supervised extension of the Hopfield model, conceived to mirror the operationally relevant scenario of Machine Learning practice, where labeled examples are scarce and unlabeled ones abundant. The network is

supplied with a dataset made of $M_L$ labeled and $M_U$ unlabeled noisy examples of $K$ unknown archetypes, and its synaptic coupling is the convex combination $J^\lambda = \lambda J^{(L)} + (1-\lambda)J^{(U)}$ of the supervised and unsupervised Hebbian kernels, with the mixing parameter $\lambda \in [0,1]$ tuning the relative weight assigned by the network to the two channels of information.

Below we summarize the main outcomes of our work.

First, by a signal-to-noise analysis of the zero-temperature one-step dynamics we obtained the Mattis magnetization on the target archetype in closed-form, whence we derived the threshold for learning $M^\star(\alpha, r, \lambda)$, namely the minimal dataset size guaranteeing a stable retrieval state at a given storage load $\alpha$ and dataset quality $r$, in closed-form for a balanced dataset $M_L = M_U = M/2$: the resulting learning boundaries in the $(M, r)$ plane are in complete agreement with Monte Carlo simulations, at all the inspected values of $\lambda$ and of the network size.

Second, by adapting Guerra's interpolation to the present setting, we derived the RS quenched statistical pressure in the high-storage regime, together with the self-consistency equations for the order parameters and their zero-temperature limit. The main technical hurdle lies in the quenched noise: since labeled and unlabeled examples stem from the same archetypes, the two channels do not decouple and the slow noise is governed by a non-trivial matrix whose exact spectral decomposition, two hybridized condensed modes, an unsupervised bulk and a supervised null space, allows one to rewrite the disorder as a collection of independent eigen-channels. The resulting phase diagrams show that the retrieval region achieved at intermediate mixing is strictly larger than those of both the pure protocols: semi-supervised Hebbian learning genuinely outperforms its supervised and unsupervised limits.

Finally, we proved that the quenched pressure is convex in the mixing parameter, so that its interior stationary point, characterized by the exact equipartition of the macroscopic energies extracted from the two subsets of the dataset, is a maximum of the free energy: equilibrium thermodynamics, by itself, rejects the mixture and would drive the synaptic rule towards the pure channels. Far from being a drawback, this no-go result provides the formal justification for treating $\lambda$ as what it operationally is, an externally imposed hyperparameter of the learning rule, rather than an order parameter spontaneously selected by the system.

Once the optimality criterion is shifted from the free energy to the retrieval capacity, the trade-off between cross-archetype interference (minimized by pure supervision) and within-class fluctuations (which penalize both extremes of the mixing range) places the capacity maximum at an interior value $\lambda^\star \in (0,1)$, for which we provided a closed-form expression, together with the minimal dataset quality $r_{\min} = 1/\sqrt{M_L + M_U + 1}$ below which no retrieval is possible, whatever the mixing. Remarkably, $\lambda^\star$ coincides with the empirical fraction of labeled examples $M_L/(M_L + M_U)$ only for noiseless datasets. As the dataset quality deteriorates, $\lambda^\star$ progressively detaches from this value and shifts weight onto the supervised channel to suppress the diverging unlabeled interference: the optimal mixing is a genuine, dataset-dependent hyperparameter and not a proxy for the proportion of labels.

As a technical remark, we stress that the strategy pursued here, the gaussianization of the two channels via CLT and universality of the quenched noise [37,38], followed by the exact diagonalization of the correlated disorder into independent eigen-channels, is of broad applicability and can be exported to other multi-channel Hebbian settings beyond the present one.

As for future perspectives, several directions naturally stem from this work. First, the whole construction can be extended to dense architectures, where neurons interact in groups of $P > 2$ units [39,40]: there, the interplay between density and semi-supervision is expected to be particularly rich, as the supervised and unsupervised channels are known to respond very differently to the interaction order. Second, our analysis is confined to the RS level of description, which, below the AT line, is expected to be only approximate: relaxing this assumption, e.g. by inspecting the instability line of the RS solution [43,44] and by moving to broken replica-symmetry schemes [34], would sharpen the low-temperature portion of the phase diagram and delimit the region in which the present boundaries are quantitatively reliable. Third, the present framework is naturally suited to be exported to hetero-associative architectures [49,50], where multiple layers of neurons learn, store and retrieve structured tuples of patterns and whose supervised and unsupervised Hebbian protocols have been recently worked out [51–53]: there, semi-supervision would intertwine with the cooperative interplay among layers, possibly enlarging the retrieval region along both directions at once. Finally, having recognized $\lambda$ as a hyperparameter, it is natural to ask whether the network itself could adaptively adjust it during training, namely to promote the mixing parameter to a slow dynamical variable: we plan to report soon on this topic.

## CRediT authorship contribution statement

**Linda Albanese:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Andrea Ladiana:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Andrea Lepre:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

L.A. acknowledges funding from the project "Patto Territoriale del Sistema Universitario Pugliese" (CUP F61B23000370006). L.A., A.Ladiana and A.Lepre are members of the GNFM group within INdAM which is acknowledged.

The authors acknowledge the use of the Lagrange Multi-GPU Server at the Department of Mathematics, Sapienza University of Rome, for computational resources supporting this work.

## Appendix A. Evaluation of the momenta of the effective post synaptic potential

In this section we compute in detail the two main quantities of Section 3: $S := \mathbb{E}\big[h_i\,\xi_i^1\big] = \mathbb{E}\big[(h_i^{(L)} + h_i^{(U)})\xi_i^1\big]$ and $V := \mathrm{Var}\big(h_i\,\xi_i^1\big)$, where we recall that:

$$\begin{aligned} h_i &:= h_i^{(L)} + h_i^{(U)} \\ &= \frac{\lambda}{N\Gamma_L}\sum_{\mu=1}^{K}\sum_{a,b=1}^{M_L}\eta_i^{\mu a}\sum_{j\neq i}\eta_j^{\mu b}\,\xi_j^1 + \frac{1-\lambda}{N\Gamma_U}\sum_{\mu=1}^{K}\sum_{c=1}^{M_U}\eta_i^{\mu c}\sum_{j\neq i}\eta_j^{\mu c}\,\xi_j^1. \end{aligned} \tag{A.1}$$

We split each contribution into a *signal* part (terms with $\mu = 1$, the archetype to be retrieved) and a *noise* part (terms with $\mu \geq 2$, the other archetypes that act as quenched interference):

$$h_i^{(L)} = h_i^{(L),\,\mu=1} + h_i^{(L),\,\mu\geq 2}, \tag{A.2}$$

$$h_i^{(U)} = h_i^{(U),\,\mu=1} + h_i^{(U),\,\mu\geq 2}, \tag{A.3}$$

where, using $\eta_j^{1b}\xi_j^1 = \chi_j^{1b}$,

$$h_i^{(L),\,\mu=1} = \frac{\lambda}{N\Gamma_L}\Big(\sum_{a=1}^{M_L}\chi_i^{1a}\Big)\Big(\sum_{b=1}^{M_L}\sum_{j\neq i}\chi_j^{1b}\Big)\,\xi_i^1, \tag{A.4}$$

$$h_i^{(U),\,\mu=1} = \frac{1-\lambda}{N\Gamma_U}\sum_{c=1}^{M_U}\chi_i^{1c}\sum_{j\neq i}\chi_j^{1c}\,\xi_i^1, \tag{A.5}$$

$$h_i^{(L),\,\mu\geq 2} = \frac{\lambda}{N\Gamma_L}\sum_{\mu\geq 2}\Big(\sum_{a}\eta_i^{\mu a}\Big)\Big(\sum_{b}\sum_{j\neq i}\eta_j^{\mu b}\xi_j^1\Big), \tag{A.6}$$

$$h_i^{(U),\,\mu\geq 2} = \frac{1-\lambda}{N\Gamma_U}\sum_{\mu\geq 2}\sum_{c}\eta_i^{\mu c}\sum_{j\neq i}\eta_j^{\mu c}\xi_j^1. \tag{A.7}$$

The factor $\xi_i^1$ pulled out of the signal terms makes manifest that, in expectation, these terms produce a contribution *aligned* with $\xi_i^1$,precisely the signal we want to control.

We begin by computing the signal, noting that for $\mu \geq 2$, the inner sum $\sum_{j\neq i}\eta_j^{\mu b}\xi_j^1$ involves the product $\xi_j^\mu\xi_j^1$ of two independent Rademacher variables, hence its expectation vanishes: $\mathbb{E}\big[\sum_{j\neq i}\eta_j^{\mu b}\xi_j^1\big] = 0$. The only non-zero contribution to $S$ comes from the $\mu = 1$ pieces (A.4) and (A.5).

Starting from (A.4) and multiplying by $\xi_i^1$ (which cancels the $\xi_i^1$ already in the formula since $(\xi_i^1)^2 = 1$):

$$h_i^{(L),\,\mu=1}\xi_i^1 = \frac{\lambda}{N\Gamma_L}\Big(\sum_{a=1}^{M_L}\chi_i^{1a}\Big)\Big(\sum_{b=1}^{M_L}\sum_{j\neq i}\chi_j^{1b}\Big). \tag{A.8}$$

The two factors involve disjoint sets of site indices, hence by independence of the $\chi$'s across sites,

$$\mathbb{E}\big[h_i^{(L),\,\mu=1}\xi_i^1\big] = \frac{\lambda}{N\Gamma_L}\Big(\sum_{a}\mathbb{E}[\chi_i^{1a}]\Big)\Big(\sum_{b}\sum_{j\neq i}\mathbb{E}[\chi_j^{1b}]\Big) = \frac{\lambda}{N\Gamma_L}M_L r(N-1)M_L r, \tag{A.9}$$

having used $\mathbb{E}[\chi_i^{1a}] = r$. Taking $N\to\infty$ and substituting $\Gamma_L = M_L^2\mathcal{R}_L = M_L^2 r^2(1+\rho_L)$,

$$\mathbb{E}\big[h_i^{(L),\,\mu=1}\xi_i^1\big] \xrightarrow{N\to\infty} \frac{\lambda M_L^2 r^2}{\Gamma_L} = \frac{\lambda}{1+\rho_L}. \tag{A.10}$$

The $1/(1+\rho_L)$ comes from the within-class fluctuation correction $1-r^2$ in $\mathcal{R}_L$. In the large dataset limit $M_L\to\infty$ one has $\rho_L\to 0$ and the signal reduces to $\lambda$ exactly.

Starting from (A.5) and multiplying by $\xi_i^1$:

$$h_i^{(U),\,\mu=1}\xi_i^1 = \frac{1-\lambda}{N\Gamma_U}\sum_{c=1}^{M_U}\chi_i^{1c}\sum_{j\neq i}\chi_j^{1c}. \tag{A.11}$$

By the same disjoint-site argument as before, but now with each example $c$ contributing independently,

$$\mathbb{E}\big[h_i^{(U),\,\mu=1}\xi_i^1\big] = \frac{1-\lambda}{N\Gamma_U}\sum_{c=1}^{M_U}\mathbb{E}[\chi_i^{1c}]\sum_{j\neq i}\mathbb{E}[\chi_j^{1c}] = \frac{1-\lambda}{N\Gamma_U}M_U r(N-1)r. \tag{A.12}$$

Taking $N\to\infty$ and using $\Gamma_U = M_U r^2$,

$$\mathbb{E}\big[h_i^{(U),\,\mu=1}\xi_i^1\big] \xrightarrow{N\to\infty} \frac{(1-\lambda)M_U r^2}{\Gamma_U} = 1-\lambda. \tag{A.13}$$

Notice that no $\rho_U$ correction appears here, because $\Gamma_U$ has been chosen to be exactly $M_U r^2$.

The variance of $h_i$ receives contributions from *every* cross-product in $\Big(h_i^{(L)} + h_i^{(U)}\Big)^2$. There are five qualitatively distinct contributions, that we label from (a) to (e):

(a) supervised *interference* (terms with $\mu \neq 1$ in $(h_i^{(L)})^2$);
(b) supervised *within-class fluctuations* (terms with $\mu = 1$ in $(h_i^{(L)})^2$);
(c) unsupervised *interference* (terms with $\mu \neq 1$ in $(h_i^{(U)})^2$);
(d) unsupervised *within-class fluctuations* (terms with $\mu = 1$ in $(h_i^{(U)})^2$);
(e) supervised-unsupervised *cross-channel* covariance ($\mu \neq 1$ in $2h_i^{(L)}h_i^{(U)}$).

The cross-channel covariance at $\mu = 1$ vanishes once we recombine the two channels with the signal $\xi_i^1$ already subtracted, and is therefore not listed.

Starting from (a), we define for simplicity:

$$A_\mu^{(L)} := \sum_{a=1}^{M_L} \eta_i^{\mu a}, \qquad B_\mu^{(L)} := \sum_{b=1}^{M_L} \sum_{j\neq i} \eta_j^{\mu b}\, \xi_j^1, \tag{A.14}$$

so that $h_i^{(L),\mu\geq 2} = (\lambda/(N\Gamma_L)) \sum_{\mu\geq 2} A_\mu^{(L)} B_\mu^{(L)}$. The two factors involve disjoint site indices, hence they are independent. Using $\mathbb{E}[\chi^a \chi^{a'}] = \delta_{aa'} + (1-\delta_{aa'})r^2$,

$$\mathbb{E}\big[(A_\mu^{(L)})^2\big] = \sum_{a,a'} \mathbb{E}[\chi^a \chi^{a'}] = M_L + M_L(M_L-1)r^2 = M_L^2 r^2 + M_L(1-r^2) = M_L^2 \mathcal{R}_L = \Gamma_L, \tag{A.15}$$

$$\mathbb{E}\big[(B_\mu^{(L)})^2\big] = \sum_{b,b'} \mathbb{E}[\chi^b \chi^{b'}] \sum_{j\neq i} 1 \;=\; (N-1)\, M_L^2 \mathcal{R}_L \;\simeq\; N\,\Gamma_L. \tag{A.16}$$

The means $\mathbb{E}[A_\mu^{(L)}]$ and $\mathbb{E}[B_\mu^{(L)}]$ vanish for $\mu \neq 1$ (by $\mathbb{E}_\xi[\xi^\mu] = 0$), so the variance of $A_\mu^{(L)} B_\mu^{(L)}$ is just $\mathbb{E}[A^2]\mathbb{E}[B^2] = N\Gamma_L^2$. Summing over the $K-1 \simeq K$ independent contributions,

$$\mathrm{Var}\big(h_i^{(L),\,\mu\geq 2}\big) = \frac{\lambda^2}{N^2\Gamma_L^2}(K-1)N\Gamma_L^2 \xrightarrow{N\to\infty} \lambda^2\,\alpha. \tag{A.17}$$

Instead, for the case (b), and so $\mu = 1$, $A_1^{(L)} = \sum_a \chi_i^{1a}$ and $B_1^{(L)} = \sum_b \sum_{j\neq i} \chi_j^{1b}$ have non-zero means $\mathbb{E}[A_1] = M_L r$, $\mathbb{E}[B_1] = (N-1)M_L r$, and variances $\mathrm{Var}(A_1) = M_L(1-r^2)$, $\mathrm{Var}(B_1) = (N-1)M_L(1-r^2)$. The full variance of the product $A_1 B_1$ (with $A_1, B_1$ independent) is given by the standard formula

$$\mathrm{Var}(A_1 B_1) = \mathrm{Var}(A_1)\,\mathrm{Var}(B_1) + \mathrm{Var}(A_1)\mathbb{E}[B_1]^2 + \mathbb{E}[A_1]^2\,\mathrm{Var}(B_1). \tag{A.18}$$

The three contributions scale, respectively, as $NM_L^2(1-r^2)^2$, $N^2M_L^3r^2(1-r^2)$, and $NM_L^3r^2(1-r^2)$. The dominant one in $N$ is the middle term $N^2M_L^3r^2(1-r^2)$. Using $\Gamma_L^2 = M_L^4\mathcal{R}_L^2 = M_L^4 r^4(1+\rho_L)^2$,

$$\mathrm{Var}\big(h_i^{(L),\,\mu=1}\big) = \frac{\lambda^2 M_L^3 r^2(1-r^2)}{\Gamma_L^2} = \frac{\lambda^2(1-r^2)}{M_L r^2(1+\rho_L)^2} \xrightarrow{N\to\infty} \frac{\lambda^2\rho_L}{(1+\rho_L)^2}. \tag{A.19}$$

The factor $1/(1+\rho_L)^2$ is the same factor that appears (squared in the signal) in (A.10): in the SNR ratio $S/\sqrt{V}$ the two cancel cleanly.

For cases (c) and (d), we define $Y_\mu^c \equiv \eta_i^{\mu c} \sum_{j\neq i} \eta_j^{\mu c} \xi_j^1$. Then $\mathbb{E}[Y_\mu^c] = 0$ (because $\mathbb{E}_\xi[\xi^\mu] = 0$) and

$$\mathbb{E}[(Y_\mu^c)^2] = \mathbb{E}[(\eta_i^{\mu c})^2] \sum_{j\neq i} \mathbb{E}\big[(\xi_j^\mu)^2(\xi_j^1)^2(\chi_j^{\mu c})^2\big] = 1\cdot(N-1) \simeq N. \tag{A.20}$$

For $c \neq c'$, both $\eta_i^{\mu c}, \eta_i^{\mu c'}$ share the same archetype $\xi_i^\mu$ but the two $\chi$'s are independent, so $\mathbb{E}[\eta_i^{\mu c}\eta_i^{\mu c'}] = r^2$ and (by the same site-by-site analysis) $\mathbb{E}[Y_\mu^c Y_\mu^{c'}] = (N-1)r^4 \simeq Nr^4$. Summing:

$$\sum_{c,c'=1}^{M_U} \mathbb{E}[Y_\mu^c Y_\mu^{c'}] = M_U N + M_U(M_U-1)Nr^4 \simeq NM_U(1+(M_U-1)r^4) = NM_U^2 r^4\left[1 + \frac{1-r^4}{M_U r^4}\right]. \tag{A.21}$$

Multiplying by $(K-1) \simeq \alpha N$ and $(1-\lambda)^2/(N^2\Gamma_U^2)$ with $\Gamma_U = M_U r^2$,

$$\mathrm{Var}\big(h_i^{(U),\,\mu\geq 2}\big) = \frac{(1-\lambda)^2}{N^2\Gamma_U^2}(\alpha N)(NM_U^2 r^4)\Big[1 + \frac{1-r^4}{M_U r^4}\Big] \xrightarrow{N\to\infty} (1-\lambda)^2\alpha\Big[1 + \frac{1-r^4}{M_U r^4}\Big]. \tag{A.22}$$

The unsupervised within-class fluctuations have the structure $\sum_{c=1}^{M_U} X_c$ (sum over examples), not $(\sum_{c=1}^{M_U} X_c)^2$, so each example contributes independently to the variance. With $X_c = \chi_i^{1,c} \sum_{j\neq i} \chi_j^{1,c}$ (after using $\eta^{1,c} = \xi^1 \chi^{1,c}$ and pulling out $\xi_i^1\xi_j^1$),

$$\mathrm{Var}(X_c) = \mathrm{Var}(\chi_i^{1c})\,\mathrm{Var}\Big(\sum_{j\neq i}\chi_j^{1c}\Big) + \mathbb{E}[\chi_i^{1c}]^2\,\mathrm{Var}\Big(\sum_{j\neq i}\chi_j^{1c}\Big) + \mathrm{Var}(\chi_i^{1c})\mathbb{E}\Big[\sum_{j\neq i}\chi_j^{1c}\Big]^2. \tag{A.23}$$

The dominant term is $\mathrm{Var}(\chi_i^{1c})\,\mathbb{E}\big[\sum_{j\neq i}\chi_j^{1c}\big]^2 = (1-r^2)(N-1)^2r^2 \simeq N^2r^2(1-r^2)$ (it dominates the other two cross-products, which are only $O(N)$ each). Summing over $c$ (independent examples), $\sum_c \mathrm{Var}(X_c) \simeq M_U N^2 r^2(1-r^2)$. Multiplying by $(1-\lambda)^2/(N^2\Gamma_U^2)$ with $\Gamma_U^2 = M_U^2 r^4$,

$$\mathrm{Var}\big(h_i^{(U),\mu=1}\big) = \frac{(1-\lambda)^2}{N^2 M_U^2 r^4} M_U N^2 r^2(1-r^2) = \frac{(1-\lambda)^2(1-r^2)}{M_U r^2} \xrightarrow{N\to\infty} (1-\lambda)^2\rho_U. \tag{A.24}$$

The last term (e) is unique to the semi-supervised model. Contributions (a) and (c) both involve sums over $\mu \neq 1$, sharing the *same* non-condensed archetype $\xi^\mu$ (only the example index $a$ vs. $c$ differs). They are therefore *not* independent, and their covariance must be added to the variance with a factor 2. Writing $h_i^{(L),\,\mu\geq 2} = (\lambda/(N\Gamma_L))\sum_{\mu\geq 2} A_\mu^{(L)} B_\mu^{(L)}$ and $h_i^{(U),\,\mu\geq 2} = ((1-\lambda)/(N\Gamma_U))\sum_{\mu\geq 2}\sum_{c=1}^{M_U} Y_\mu^c$,

$$\mathrm{Cov}\big(h_i^{(L),\,\mu\geq 2}, h_i^{(U),\,\mu\geq 2}\big) = \frac{\lambda(1-\lambda)}{N^2\Gamma_L\Gamma_U}\,(K-1)\sum_{c=1}^{M_U}\mathbb{E}\big[A_\mu^{(L)}\,Y_\mu^c\big]. \tag{A.25}$$

By disjoint-site factorization,

$$\mathbb{E}[A_\mu^{(L)} Y_\mu^c] = \mathbb{E}\big[A_\mu^{(L)}\eta_i^{\mu c}\big]\cdot\mathbb{E}\Big[B_\mu^{(L)\prime}\sum_{j\neq i}\eta_j^{\mu c}\xi_j^1\Big], \tag{A.26}$$

where $B_\mu^{(L)\prime}$ denotes the $j$-sum part of $B_\mu^{(L)}$ (the $b$-sum factors out cleanly). Now

$$\mathbb{E}\big[A_\mu^{(L)}\eta_i^{\mu c}\big] = \sum_{a=1}^{M_L}\mathbb{E}\big[\eta_i^{\mu a}\eta_i^{\mu c}\big] = \sum_{a=1}^{M_L} r^2 = M_L r^2, \tag{A.27}$$

$$\mathbb{E}\Big[B_\mu^{(L)\prime}\sum_{j\neq i}\eta_j^{\mu c}\xi_j^1\Big] = \sum_{b=1}^{M_L}\sum_{j\neq i}\mathbb{E}\big[\eta_j^{\mu b}\eta_j^{\mu c}\big] = (N-1)M_L r^2, \tag{A.28}$$

where we used $\mathbb{E}[\eta_i^{\mu a}\eta_i^{\mu c}] = \mathbb{E}[(\xi_i^\mu)^2\chi_i^{\mu a}\chi_i^{\mu c}] = \mathbb{E}[\chi]^2 = r^2$ (different examples $a \neq c$ have independent $\chi$'s). Inserting:

$$\begin{aligned}\mathrm{Cov}\big(h_i^{(L),\,\mu\geq 2}, h_i^{(U),\,\mu\geq 2}\big) &= \frac{\lambda(1-\lambda)}{N^2\Gamma_L\Gamma_U}\cdot(K-1)\cdot M_U\cdot M_L r^2\cdot(N-1)M_L r^2\\ &\xrightarrow{N\to\infty} \frac{\lambda(1-\lambda)\,\alpha\,M_U M_L^2 r^4}{\Gamma_L\Gamma_U} = \frac{\lambda(1-\lambda)\,\alpha\,M_U M_L^2 r^4}{M_L^2\mathcal{R}_L\cdot M_U r^2}\\ &= \frac{\lambda(1-\lambda)\,\alpha\,r^2}{\mathcal{R}_L} = \frac{\lambda(1-\lambda)\,\alpha}{1+\rho_L}.\end{aligned} \tag{A.29}$$

This contribution enters the variance with the factor 2 from $\mathrm{Var}(h^L+h^U) = \mathrm{Var}(h^L)+\mathrm{Var}(h^U)+2\,\mathrm{Cov}(h^L,h^U)$, giving the cross-channel piece $2\lambda(1-\lambda)\alpha/(1+\rho_L)$. In the large-$M_L$ limit $\rho_L \to 0$, the cross piece reduces to $2\lambda(1-\lambda)\alpha$.

Collecting (A.10) and (A.13), the total signal is

$$S = \mathbb{E}\big[h_i\,\xi_i^1\big] = \frac{\lambda}{1+\rho_L} + (1-\lambda). \tag{A.30}$$

Collecting (a)+(b)+(c)+(d)+2×(e) yields the total noise variance

$$V = \mathrm{Var}(h_i) = C(\lambda, r, \rho_L, \rho_U)\,\alpha + R(\lambda, r, \rho_L, \rho_U), \tag{A.31}$$

which splits into a slow-noise interference $C\alpha$ and a within-class noise $R$ with

$$C(\lambda, r, \rho_L, \rho_U) = \lambda^2 + (1-\lambda)^2\Big[1+\frac{1-r^4}{M_U r^4}\Big] + \frac{2\lambda(1-\lambda)}{1+\rho_L}, \tag{A.32}$$

$$R(\lambda, r, \rho_L, \rho_U) = \frac{\lambda^2\rho_L}{(1+\rho_L)^2} + (1-\lambda)^2\rho_U, \tag{A.33}$$

which is exactly what we have in Section 3.

## Appendix B. Computation of the signal

In this appendix we will derive a compact form for the signal term (3.3), which we rewrite here for simplicity:

$$\mathcal{B}_{\mathrm{sig}} = \frac{\beta\lambda N}{2\Gamma_L}\Big(\sum_{a=1}^{M_L} n_L^{1a}\Big)^2 + \frac{\beta(1-\lambda)N}{2\Gamma_U}\sum_{c=1}^{M_U}(n_U^{1c})^2 \tag{B.1}$$

Indeed, using the decomposition (2.5) and the definitions (2.21), the inner sum in the supervised signal is

$$\sum_{a=1}^{M_L} n^{1a} = \frac{1}{N}\sum_{i=1}^{N}\Big(\sum_{a=1}^{M_L}\eta_i^{1a}\Big)\sigma_i = \frac{M_L\mathcal{R}_L}{r}\,n_L^1, \tag{B.2}$$

where we used $\sum_a \eta_i^{1a} = \xi_i^1\sum_a \chi_i^{1a}$ together with the normalization prefactor $r/\mathcal{R}_L$ entering (2.21) (the identification becomes exact under $\mathbb{E}_\chi$, with corrections suppressed by powers of $N^{-1/2}$ under the RS ansatz). Inserting (B.2) into (B.1),

$$\frac{\beta\lambda N}{2\Gamma_L}\Big(\sum_a n^{1a}\Big)^2 = \frac{\beta\lambda N}{2\Gamma_L}\,\frac{M_L^2\mathcal{R}_L^2}{r^2}\,n_L^2 = \frac{\beta\lambda N}{2}\mathcal{R}_L\,\frac{n_L^2}{r^2} = \frac{\beta\lambda N}{2}\,(1+\rho_L)\,(n_L^1)^2, \tag{B.3}$$

having used $\Gamma_L = M_L^2\mathcal{R}_L$ and $\mathcal{R}_L/r^2 = 1+\rho_L$.

The unsupervised signal is, instead, a sum of squares of the example overlaps $n^{1c}$. Under the RS ansatz the empirical second moment $\frac{1}{M_U}\sum_c (n^{1c})^2$ self-averages and, expressed through the example magnetization $n_U$ of (2.21), collapses onto

$$\frac{\beta(1-\lambda)N}{2\Gamma_U}\sum_{c=1}^{M_U}(n^{1c})^2 = \frac{\beta(1-\lambda)N}{2}\,(1+\rho_U)\,(n_U^1)^2 + O(N^{-1}), \tag{B.4}$$

the factor $(1+\rho_U) = \mathcal{R}_U/r^2$ arising, exactly as in the supervised case, from the prefactor $r/\mathcal{R}_U$ in the definition of $n_U$ together with $\Gamma_U = M_U r^2$. Putting (B.3) and (B.4) together,

$$\mathcal{B}_{\text{sig}} = \frac{\beta N}{2}\left[\lambda(1+\rho_L)\,n_L^2 + (1-\lambda)(1+\rho_U)\,n_U^2\right]. \tag{B.5}$$

**Remark 7** (*Sum of Squares vs. Square of a Sum*)**.** The two channels are structurally different already at the level of the signal. The supervised contribution is the square of a sum over labeled examples, $\left(\sum_a n^{1a}\right)^2 = (M_L\mathcal{R}_L/r)^2 n_L^2$, so (B.3) is an exact algebraic identity. The unsupervised contribution is a sum of squares, $\sum_c (n^{1c})^2$, which is not reducible to $n_U^2$ by algebra alone: writing $n^{1c} = r\,m + \nu_c$ with $\nu_c = N^{-1}\sum_i \zeta_i^{1c}\sigma_i$ ($\mathbb{E}[\nu_c] = 0$, $\mathbb{E}[\nu_c^2] = (1-r^2)/N$), one has $\sum_c (n^{1c})^2 = M_U r^2 m^2 + \sum_c \nu_c^2 + \dots$, whose leading-$N$ part reproduces (B.4) once $n_U$ is identified at the retrieval saddle through $\langle n_U\rangle = \langle m\rangle$ (Appendix C.1)[11]; the example-to-example fluctuations $\nu_c$ are the microscopic origin of the within-class noise $\rho_U$. We keep the compact form (B.4) throughout.

## Appendix C. Computations of the terms in Guerra's interpolation

In this section we report the detailed calculations for the quenched statistical pressure and self-consistency equations via Guerra's interpolation in Section 4.3.

Let us start with the computation of the $t$-derivative of the interpolating pressure $\mathcal{A}_N(t, J_m)$ term by term. The general scheme is

$$\partial_t \mathcal{A}_N(t, J_m) = \frac{1}{N}\,\mathbb{E}\left[\omega_t(\partial_t \mathcal{B}(t, J_m))\right], \tag{C.1}$$

where $\partial_t \mathcal{B}(t, J_m)$ is the partial derivative of the interpolating Boltzmann factor (4.32) with respect to $t$ at fixed disorder and fixed interpolating variables. Differentiating each of the seven blocks in (4.32),

$$\begin{aligned}
\partial_t \mathcal{B}(t, J_m) = &\underbrace{+\frac{1}{2}\beta(1-\langle q_{12}\rangle)\sum_{\mu\geq 2}\sum_k \tilde{\mu}_k (z_\mu^k)^2}_{\text{(I)}} + \underbrace{\frac{1}{2\sqrt{t}}\sqrt{\frac{\beta}{N}}\sum_{\mu\geq 2}\sum_k \sqrt{\tilde{\mu}_k}\, z_\mu^k \sum_i \lambda_i^{\mu k}\sigma_i}_{\text{(II)}} \\
&+ \underbrace{\frac{\beta N}{2}\left[\lambda(1+\rho_L)n_L^2 + (1-\lambda)(1+\rho_U)n_U^2\right]}_{\text{(III)}} \\
&- \underbrace{\beta\left[\lambda(1+\rho_L)\langle n_L\rangle N n_L + (1-\lambda)(1+\rho_U)\langle n_U\rangle N n_U\right]}_{\text{(IV)}} \\
&- \underbrace{\frac{1}{2\sqrt{1-t}}\sqrt{\alpha\beta\sum_k d_k\tilde{\mu}_k\langle p^k\rangle}\sum_i \theta_i\sigma_i}_{\text{(V)}} - \underbrace{\frac{1}{2\sqrt{1-t}}\sqrt{\beta\langle q_{12}\rangle}\sum_{\mu\geq 2}\sum_k \sqrt{\tilde{\mu}_k}\,\psi_\mu^k z_\mu^k}_{\text{(VI)}}
\end{aligned} \tag{C.2}$$

where we stress that the sign on the blocks (V), (VI) comes from differentiating $\sqrt{1-t}$.

The terms involving the auxiliary Gaussian fields $\tilde{\lambda}$, $\theta$, $\psi$ are handled by repeated application of Stein's lemma, also known as Wick's theorem: for any smooth function $F(g)$ of a centered unit Gaussian variable $g$ for which the two expectations $\mathbb{E}\,(gF(g))$ and $\mathbb{E}\left(\partial_g F(g)\right)$ both exist, the relation

$$\mathbb{E}[gF(g)] = \mathbb{E}\left[\partial_g F(g)\right] \qquad \text{for } g\sim\mathcal{N}(0,1) \tag{C.3}$$

holds.

For the two-body block in (II), $g = \lambda_i^{\mu,k}$ and $F = \omega_t(z_\mu^k\sigma_i)$. The dependence of $F$ on $g$ enters only through the interpolating exponent, which contains $g$ in (II) (proportional to $\sqrt{t}$). Differentiating,

$$\mathbb{E}\left[\lambda_i^{\mu,k}\,\omega_t(z_\mu^k\sigma_i)\right] = \sqrt{\frac{t\beta\tilde{\mu}_k}{N}}\,\mathbb{E}\left[\omega_t((z_\mu^k)^2) - \omega_t(z_\mu^k\sigma_i)^2\right], \tag{C.4}$$

[11] We stress, that here we have used the (2.5) decomposition of the examples.

where the first term is a one-replica correlator and the second is a two-replica connected correlator (the second derivative of $\omega_t$ with respect to a parameter in the Boltzmann weight gives one replica diagonal and one replica off-diagonal piece). Summing (C.4) over $i, \mu, k$ and reintroducing the prefactor $1/(2\sqrt{t})$ from (II), then using the definitions (2.25), (4.27) with the multiplicities $d_k$,

$$\frac{1}{N}\mathbb{E}\big[\omega_t(\mathrm{II}')\big] = \frac{K\beta}{2N}\sum_k d_k\,\tilde{\mu}_k\big[\mathbb{E}\omega_t(p_{11}^k) - \mathbb{E}\omega_t(p_{12}^k\,q_{12})\big], \tag{C.5}$$

one term per eigen-channel, weighted by its eigenvalue $\tilde{\mu}_k$ and multiplicity $d_k$.

Applying Stein's lemma (C.3) to the block (V) with $g = \theta_i$:

$$\mathbb{E}[\theta_i\,\omega_t(\sigma_i)] = \sqrt{(1-t)\alpha\beta\sum_k d_k\tilde{\mu}_k\langle p^k\rangle}\,\mathbb{E}\big[\omega_t(1)-\omega_t(\sigma_i)^2\big] = \sqrt{(1-t)\alpha\beta\sum_k d_k\tilde{\mu}_k\langle p^k\rangle}\,\mathbb{E}\omega_t(1-q_{12}), \tag{C.6}$$

giving, after summing over $i$ and the $1/(2\sqrt{1-t})$ prefactor cancels,

$$-\frac{1}{N}\mathbb{E}\big[\omega_t(\mathrm{V})\big] = -\frac{\alpha\beta\sum_k d_k\tilde{\mu}_k\langle p^k\rangle}{2}\big[1-\mathbb{E}\langle q_{12}\rangle_t\big]. \tag{C.7}$$

Analogous Stein contractions on $\psi_\mu^k$ in (VI) yield

$$-\frac{1}{N}\mathbb{E}\big[\omega_t(\mathrm{VI})\big] = -\frac{\alpha\beta\,\langle q\rangle}{2}\sum_k d_k\,\tilde{\mu}_k\,\mathbb{E}\omega_t(p_{11}^k - p_{12}^k). \tag{C.8}$$

Putting (C.5)–(C.8) together with the elementary contributions (I), (III), (IV), and introducing the centered fluctuations

$$\Delta[n_X^2] = \mathbb{E}\omega_t(n_X - \langle n_X\rangle)^2, \qquad \Delta[p^k q] = \mathbb{E}\omega_t((p_{12}^k - \langle p^k\rangle)(q_{12} - \langle q\rangle)), \tag{C.9}$$

the streaming can be cast in the compact form

$$\begin{aligned}\partial_t\mathcal{A}_N(t,J_m) = {}& \frac{\beta}{2}\sum_{X\in\{L,U\}} c_X(1+\rho_X)\big(\Delta[n_X^2] - \langle n_X\rangle^2\big)\\ & - \frac{\alpha\beta}{2}\sum_k d_k\,\tilde{\mu}_k\big(\Delta[p^k q] + \langle p^k\rangle(1-\langle q\rangle)\big),\end{aligned} \tag{C.10}$$

with $c_L = \lambda$, $c_U = 1-\lambda$. Under the RS ansatz (4.30) every centered fluctuation in (4.38) vanishes in the thermodynamic limit, $\Delta[n_X^2]\to 0$ and $\Delta[p_{12}^k q_{12}]\to 0$. The streaming becomes $t$-independent:

$$\partial_t\mathcal{A}(t,J_m) = -\frac{\beta}{2}\big[\lambda(1+\rho_L)\langle n_L\rangle^2 + (1-\lambda)(1+\rho_U)\langle n_U\rangle^2\big] - \frac{\alpha\beta}{2}(1-\langle q_{12}\rangle)\sum_k d_k\,\tilde{\mu}_k\,\langle p_{12}^k\rangle. \tag{C.11}$$

Since $\partial_t\mathcal{A}$ is $t$-independent, the integral $\int_0^1 \partial_t\mathcal{A}\,dt$ in (4.29) equals (4.39) directly.

Now, let us proceed to compute the one-body terms, which in general are easier to compute than the original model. At $t=0$ the terms which survive are:

$$\begin{aligned}\mathcal{B}(t=0,J_m) = {}& -\tfrac{1}{2}\sum_{\mu\geq 2}\sum_k\big[1-\beta\tilde{\mu}_k(1-\langle q\rangle)\big](z_\mu^k)^2 + \sqrt{\beta\langle q\rangle}\sum_{\mu\geq 2}\sum_k\sqrt{\tilde{\mu}_k}\,\psi_\mu^k z_\mu^k\\ & + \beta\big[\lambda(1+\rho_L)\langle n_L\rangle N n_L + (1-\lambda)(1+\rho_U)\langle n_U\rangle N n_U\big]\\ & + \sqrt{\alpha\beta\sum_k d_k\tilde{\mu}_k\langle p^k\rangle}\sum_i\theta_i\sigma_i - J_m\beta N m,\end{aligned} \tag{C.12}$$

The Boltzmann factor at $t=0$ factorizes over the analog $z$-sector and the spin sector, and each reduces to a product over single degrees of freedom. We treat them separately.

The analog sector is Gaussian. Using

$$\int \mathcal{D}z\,e^{-\frac{A}{2}z^2+Bz} = A^{-1/2}\exp\left(\frac{B^2}{2A}\right), \tag{C.13}$$

each eigen-channel mode $z_\mu^k$ integrates with $A = 1-\beta\tilde{\mu}_k(1-\langle q\rangle)$ and $B = \sqrt{\beta\tilde{\mu}_k\langle q\rangle}\,\psi_\mu^k$:

$$\begin{aligned}&\int \mathcal{D}z_\mu^k\,\exp\Big(-\tfrac{1}{2}[1-\beta\tilde{\mu}_k(1-\langle q_{12}\rangle)](z_\mu^k)^2 + \sqrt{\beta\tilde{\mu}_k\langle q_{12}\rangle}\,\psi_\mu^k z_\mu^k\Big)\\ &\qquad = \big[1-\beta\tilde{\mu}_k(1-\langle q_{12}\rangle)\big]^{-1/2}\exp\Big[\frac{\beta\tilde{\mu}_k\langle q_{12}\rangle\,(\psi_\mu^k)^2}{2[1-\beta\tilde{\mu}_k(1-\langle q_{12}\rangle)]}\Big].\end{aligned} \tag{C.14}$$

Taking $\mathbb{E}_\psi[(\psi_\mu^k)^2] = 1$, summing over $\mu = 2,\dots,K$ (a factor $K-1\to\alpha N$) and over the eigen-channels $k$ with their multiplicities $d_k$ from the spectrum (4.20), we obtain the analog contribution to $\mathcal{A}(0)$:

$$\mathcal{A}^{(\mathrm{an})}(0) = -\frac{\alpha}{2}\sum_k d_k\Big[\ln\big(1-\beta\tilde{\mu}_k(1-\langle q\rangle)\big) - \frac{\beta\tilde{\mu}_k\langle q\rangle}{1-\beta\tilde{\mu}_k(1-\langle q\rangle)}\Big]. \tag{C.15}$$

The null modes ($\tilde{\mu}_k = 0$, multiplicity $M_L - 1$) contribute nothing; the bulk ($\tilde{\mu}_U^{\mathrm{bulk}} = (1-\lambda)\rho_U$, multiplicity $M_U - 1$) and the two condensed modes $\tilde{\mu}_\pm$ supply the entire analog free energy.

The remaining one-body terms in (C.12) act on $\sigma_i$ independently. Using the example decomposition (2.5) we can write:

$$\beta\lambda(1+\rho_L)\langle n_L\rangle N n_L = \beta\lambda(1+\rho_L)\langle n_L\rangle \cdot \frac{r}{\mathcal{R}_L}\cdot\frac{1}{M_L}\sum_{i,a}\eta_i^{1a}\sigma_i = \beta\lambda\,\langle n_L\rangle\,\frac{1}{M_L r}\sum_{i,a}\eta_i^{1a}\sigma_i, \tag{C.16}$$

where we used $(1+\rho_L)\cdot r/\mathcal{R}_L = 1/r$. Using $\eta_i^{1a} = \xi_i^1\,\chi_i^{1a}$, the on-site field acting on $\sigma_i$ from the labeled channel is

$$h_i^{(L)} := \beta\lambda\langle n_L\rangle\,\xi_i^1\cdot\frac{1}{M_L r}\sum_{a=1}^{M_L}\chi_i^{1a}. \tag{C.17}$$

Analogously for the unlabeled channel,

$$h_i^{(U)} := \beta(1-\lambda)\langle n_U\rangle\,\xi_i^1\cdot\frac{1}{M_U r}\sum_{c=1}^{M_U}\chi_i^{1c}. \tag{C.18}$$

Adding the cavity field from (V) and the source field $-J_m\beta\xi_i^1$ from (VII), the full one-site field is

$$h_i = h_i^L + h_i^U + \sqrt{\alpha\beta\Pi}\,\theta_i - J_m\beta\,\xi_i^1. \tag{C.19}$$

The sum over $\sigma_i \in \{-1,+1\}$ produces $2\cosh(h_i)$. Performing the gauge transformation $\sigma_i \to \xi_i^1\sigma_i$ (under which the law of $\chi_i^{1a}$ is unchanged because $\chi$ is independent of $\xi$ and the $\xi_i^1$ have the same distribution as their negatives), the explicit $\xi_i^1$ in the cavity fields cancels, the source becomes $-J_m\beta$ acting on $\sigma_i$ directly, and the cavity-field random variable $\theta_i$ remains a standard Gaussian. The resulting one-site Boltzmann sum at $J_m \to 0$ reads

$$\ln 2 + \mathbb{E}_{\chi,\theta}\,\ln\cosh\Big[\beta\Big(\lambda\langle n_L\rangle\,\frac{1}{M_L r}\sum_a\chi^{1a} + (1-\lambda)\langle n_U\rangle\,\frac{1}{M_U r}\sum_c\chi^{1c}\Big) + \theta\sqrt{\alpha\beta\sum_k d_k\tilde{\mu}_k\langle p^k\rangle}\Big], \tag{C.20}$$

Summing over the $N$ sites and dividing by $N$ gives the spin contribution to the Cauchy datum:

$$\mathcal{A}^{(\sigma)}(0) = \ln 2 + \mathbb{E}_{\chi,\theta}\,\ln\cosh\Big[\beta\Big(\lambda\langle n_L\rangle\,\tfrac{1}{M_L r}\sum_a\chi^{1a} + (1-\lambda)\langle n_U\rangle\,\tfrac{1}{M_U r}\sum_c\chi^{1c}\Big) - \beta J_m + \theta\sqrt{\alpha\beta\sum_k d_k\tilde{\mu}_k\langle p^k\rangle}\Big]. \tag{C.21}$$

Putting (C.15) and (C.21) together:

$$\mathcal{A}(t{=}0, J_m) = -\frac{\alpha}{2}\sum_k d_k\Big[\ln\big(1-\beta\tilde{\mu}_k(1-\langle q\rangle)\big) - \frac{\beta\tilde{\mu}_k\langle q\rangle}{1-\beta\tilde{\mu}_k(1-\langle q\rangle)}\Big] + \ln 2 \tag{C.22}$$

$$+\ \mathbb{E}_{\chi,\theta}\,\ln\cosh\Big[\beta\big(\lambda\langle n_L\rangle\,\tfrac{1}{M_L r}\sum_a\chi^{1a} + (1-\lambda)\langle n_U\rangle\,\tfrac{1}{M_U r}\sum_c\chi^{1c}\big) - \beta J_m + \theta\sqrt{\alpha\beta\sum_k d_k\tilde{\mu}_k\langle p^k\rangle}\Big]. \tag{C.23}$$

Inserting (4.36) and (4.39) into the sum rule (4.29) we obtain

$$\begin{aligned}\mathcal{A}_{\mathrm{RS}}(J_m) = &-\frac{\alpha}{2}\sum_k d_k\Big[\ln\big(1-\beta\tilde{\mu}_k(1-\langle q\rangle)\big) - \frac{\beta\tilde{\mu}_k\langle q\rangle}{1-\beta\tilde{\mu}_k(1-\langle q\rangle)}\Big]\\ &-\frac{\beta}{2}\big[\lambda(1+\rho_L)\langle n_L\rangle^2 + (1-\lambda)(1+\rho_U)\langle n_U\rangle^2\big] - \frac{\alpha\beta}{2}(1-\langle q\rangle)\sum_k d_k\tilde{\mu}_k\langle p^k\rangle\\ &+\ln 2 + \mathbb{E}_{\chi,\theta}\,\ln\cosh\Big[\beta\big(\lambda\langle n_L\rangle\,\tfrac{1}{M_L r}\sum_a\chi^{1a} + (1-\lambda)\langle n_U\rangle\,\tfrac{1}{M_U r}\sum_c\chi^{1c}\big) - \beta J_m + \theta\sqrt{\alpha\beta\sum_k d_k\tilde{\mu}_k\langle p^k\rangle}\Big].\end{aligned} \tag{C.24}$$

This is the RS quenched statistical pressure of the semi-supervised Hopfield model, expressed in terms of the five order parameters $\langle m\rangle, \langle n_L\rangle, \langle n_U\rangle, \langle q\rangle, \langle p^k\rangle$ and of the control parameters $\alpha, \beta, r, \rho_L, \rho_U, \lambda$.

The five order parameters $\langle m\rangle, \langle n_L\rangle, \langle n_U\rangle, \langle q\rangle, \langle p^k\rangle$ are determined by extremizing the RS statistical pressure (C.24). We compute the derivatives one by one, treating $\langle n_L\rangle$ and $\langle n_U\rangle$ as genuinely independent parameters, and derive the collapse $\langle n_L\rangle = \langle n_U\rangle = \langle m\rangle$ at the end as a consequence of the saddle equations.

For simplicity we introduce this notation:

$$\Xi(\chi,\theta) := \beta\,G_n(\chi) + \theta\sqrt{\alpha\beta\,\Pi}, \qquad D_k := 1-\beta\tilde{\mu}_k(1-\langle q\rangle), \qquad \Pi := \sum_k d_k\tilde{\mu}_k\langle p^k\rangle, \tag{C.25}$$

with

$$G_n(\chi) := \lambda\langle n_L\rangle\,\frac{1}{M_L r}\sum_{a=1}^{M_L}\chi^{1a} + (1-\lambda)\langle n_U\rangle\,\frac{1}{M_U r}\sum_{c=1}^{M_U}\chi^{1c}. \tag{C.26}$$

The analog overlap $\langle p^k\rangle$ is the two-replica overlap of the Gaussian eigen-modes $z_\mu^k$, and it is fixed by the analog (Gaussian) sector of the partition function. At the RS saddle each eigen-channel $z_\mu^k$ is governed by the effective single-mode weight already isolated in (C.14),

$$\propto \exp\Big[-\tfrac{1}{2}\,D_k\,(z_\mu^k)^2 + \sqrt{\beta\tilde{\mu}_k\langle q\rangle}\,\psi_\mu^k\,z_\mu^k\Big], \qquad \psi_\mu^k \sim \mathcal{N}(0,1), \tag{C.27}$$

so that its conditional mean is $\omega(z_\mu^k) = \sqrt{\beta\tilde{\mu}_k\langle q\rangle}\,\psi_\mu^k/D_k$. The disorder-averaged two-replica overlap then follows from the Gaussian second moment, using $\mathbb{E}_\psi[(\psi_\mu^k)^2] = 1$,

$$\langle p^k\rangle = \mathbb{E}_\psi\big[\omega(z_\mu^k)^2\big] = \frac{\beta\tilde{\mu}_k\langle q\rangle}{D_k^2}\,\mathbb{E}_\psi[(\psi_\mu^k)^2] = \frac{\beta\tilde{\mu}_k\langle q\rangle}{D_k^2}, \tag{C.28}$$

that is, using (C.25), the standard saddle-point equation

$$\langle p^k \rangle = \frac{\beta \tilde{\mu}_k \langle q \rangle}{\left[1 - \beta \tilde{\mu}_k (1 - \langle q \rangle)\right]^2}. \tag{C.29}$$

This is the analog-overlap self-consistency equation.

Then, differentiating (C.24) with respect to $\langle q \rangle$ and using (C.29) (which cancels several pieces), the terms involving the logarithm and the geometric series collapse and exactly compensate the Stein-lemma transformation of the cavity term. Explicitly, the analog logarithmic block contributes

$$\frac{\partial}{\partial \langle q \rangle} \left\{ -\frac{\alpha}{2} \sum_k d_k \left[ \ln D_k - \frac{\beta \tilde{\mu}_k \langle q \rangle}{D_k} \right] \right\} = -\frac{\alpha}{2} \sum_k d_k \frac{\beta^2 \tilde{\mu}_k^2 \langle q \rangle}{D_k^2}, \tag{C.30}$$

where the two pieces generated by $\partial_{\langle q \rangle} \ln D_k = \beta \tilde{\mu}_k / D_k$ cancel against the corresponding piece of $\partial_{\langle q \rangle} (\beta \tilde{\mu}_k \langle q \rangle / D_k)$, leaving only the squared denominator. The bilinear term gives $\partial_{\langle q \rangle} \left[ -\frac{\alpha\beta}{2} (1 - \langle q \rangle) \sum_k d_k \tilde{\mu}_k \langle p^k \rangle \right] = +\frac{\alpha\beta}{2} \sum_k d_k \tilde{\mu}_k \langle p^k \rangle$, and inserting (C.29) this equals $+\frac{\alpha}{2} \sum_k d_k \beta^2 \tilde{\mu}_k^2 \langle q \rangle / D_k^2$, which cancels (C.30) identically. Hence no analog contribution survives, and $\langle q \rangle$ is fixed by its definition as the two-replica spin overlap of the effective single-site measure, $\omega(\sigma_i) = \tanh \Xi$, so that

$$\langle q \rangle = \mathbb{E}_{\chi,\theta} \left[ \omega(\sigma_i)^2 \right] = \mathbb{E}_{\chi,\theta} \left[ \tanh^2 \Xi \right], \tag{C.31}$$

that is

$$\langle q \rangle = \mathbb{E}_{\chi,\theta} \tanh^2 \left[ \beta \left( \lambda \langle n_L \rangle \frac{1}{M_L r} \sum_a \chi^{1a} + (1 - \lambda) \langle n_U \rangle \frac{1}{M_U r} \sum_c \chi^{1c} \right) + \theta \sqrt{\alpha \beta \Pi} \right]. \tag{C.32}$$

This is the spin-glass overlap self-consistency, with the within-channel fluctuation noise embedded in the $\chi$-sums and the slow noise hidden in $\Pi$.

The two example magnetizations $\langle n_L \rangle, \langle n_U \rangle$ enter (C.24) both through the term $-\frac{\beta}{2} \sum_X c_X (1 + \rho_X) \langle n_X \rangle^2$ and through the cavity argument, where they multiply the $\chi$-sums of the corresponding channel. With $c_L = \lambda$, $c_U = 1 - \lambda$ and $\partial G_n / \partial \langle n_X \rangle = c_X \frac{1}{M_X r} \sum \chi^{1\cdot}$, the two derivatives read explicitly

$$\frac{\partial}{\partial \langle n_X \rangle} \left[ -\tfrac{\beta}{2} c_X (1 + \rho_X) \langle n_X \rangle^2 \right] = -\beta\, c_X (1 + \rho_X) \langle n_X \rangle, \tag{C.33}$$

$$\frac{\partial}{\partial \langle n_X \rangle} \mathbb{E}_{\chi,\theta} \left[ \ln \cosh \Xi \right] = \mathbb{E}_{\chi,\theta} \left[ \beta\, c_X \frac{1}{M_X r} \sum \chi^{1\cdot} \tanh \Xi \right]. \tag{C.34}$$

Setting $\partial \mathcal{A}_{\mathrm{RS}} / \partial \langle n_X \rangle = 0$ for $X \in \{L, U\}$, the common factor $\beta c_X$ cancels and one is left with $(1 + \rho_X) \langle n_X \rangle = \mathbb{E}_{\chi,\theta} [\frac{1}{M_X r} \sum \chi^{1\cdot} \tanh \Xi]$, i.e. the two coupled saddle equations

$$\langle n_X \rangle = \frac{1}{1 + \rho_X} \mathbb{E}_{\chi,\theta} \left[ \frac{1}{M_X r} \sum \chi^{1\cdot} \cdot \tanh\left( \beta\, G_n(\chi) + \theta \sqrt{\alpha \beta \Pi} \right) \right], \qquad X \in \{L, U\}, \tag{C.35}$$

where the $\chi$-sum is over labeled (resp. unlabeled) examples. The prefactor $1/(1 + \rho_X)$ comes from differentiating the quadratic $(1 + \rho_X) \langle n_X \rangle^2$ piece and combining with the cavity derivative.

The Mattis magnetization $\langle m \rangle = -\beta^{-1} \partial_{J_m} \mathcal{A}_{\mathrm{RS}} \big|_{J_m = 0}$ is obtained directly by differentiating (C.24) with respect to $J_m$. The field $J_m$ enters the pressure only through the term $-\beta J_m$ inside the cavity argument, so

$$\partial_{J_m} \mathcal{A}_{\mathrm{RS}} = \mathbb{E}_{\chi,\theta} \left[ (-\beta) \tanh(\Xi - \beta J_m) \right], \tag{C.36}$$

and setting $J_m = 0$,

$$\langle m \rangle = -\beta^{-1} \mathbb{E}_{\chi,\theta} \left[ (-\beta) \tanh \Xi \right] = \mathbb{E}_{\chi,\theta} \left[ \tanh \Xi \right], \tag{C.37}$$

that is

$$\langle m \rangle = \mathbb{E}_{\chi,\theta} \tanh \left[ \beta\, G_n(\chi) + \theta \sqrt{\alpha \beta \Pi} \right], \tag{C.38}$$

with $G_n(\chi)$ as in (C.26). Collecting (C.29), (C.32), (C.35) and (C.38) we have a closed system in the five order parameters $\langle m \rangle, \langle n_L \rangle, \langle n_U \rangle, \langle q \rangle, \langle p^k \rangle$.

### C.1. Large datasets limit

We now specialize the RS self-consistency Eqs. (C.29), (C.32), (C.35) and (C.38) to the large dataset regime $M_L, M_U \to \infty$ with the rescaled sizes $\rho_L, \rho_U = O(1)$ held fixed. In this regime the empirical means of the $\chi$-variables that enter $G_n(\chi)$ can be linearized by the CLT, the example magnetizations $\langle n_L \rangle, \langle n_U \rangle$ collapse onto the Mattis magnetization $\langle m \rangle$, and the slow noise acquires its closed-form eigenvalue-resolved expression.

For each channel $X \in \{L, U\}$ the empirical mean of $M_X$ i.i.d. Rademacher flips, after the prefactor $1/(M_X r)$, converges by the CLT to

$$\frac{1}{M_X r} \sum_{a=1}^{M_X} \chi^{1a} \xrightarrow{M_X \to \infty} 1 + \sqrt{\rho_X}\, Z_X, \qquad Z_X \sim \mathcal{N}(0, 1), \tag{C.39}$$

with $Z_L$ and $Z_U$ independent. Inserting (C.39) into the combined mean (C.26),

$$G_n(\chi) \xrightarrow{M\to\infty} \lambda\langle n_L\rangle + (1-\lambda)\langle n_U\rangle + \lambda\langle n_L\rangle\sqrt{\rho_L}\, Z_L + (1-\lambda)\langle n_U\rangle\sqrt{\rho_U}\, Z_U, \tag{C.40}$$

the deterministic part being the linear combination of the two example magnetizations and the random part a Gaussian of variance $\lambda^2\langle n_L\rangle^2\rho_L + (1-\lambda)^2\langle n_U\rangle^2\rho_U$.

Inserting (C.40) into the saddle Eqs. (C.35) for $\langle n_X\rangle$ and applying Stein's lemma (C.3) on $Z_X$ to evaluate the correlator $\mathbb{E}_{Z_X}[(1+\sqrt{\rho_X}Z_X)\tanh(\beta\, G_n(\chi) + \theta\sqrt{\alpha\beta\Pi})]$,

$$\langle n_X\rangle = \frac{1}{1+\rho_X}\Big[\lambda_X\langle n_X\rangle + \rho_X\,\lambda_X\beta(1-\langle q\rangle)\,\langle n_X\rangle\Big] + \frac{\lambda_{X'}\langle n_{X'}\rangle}{1+\rho_X}, \tag{C.41}$$

where we have introduced the shorthand $\lambda_L := \lambda$, $\lambda_U := 1-\lambda$, and the second term collects the contribution from the other channel. The first term is the leading $\mathbb{E}_\chi[(1+\sqrt{\rho_X}Z_X)]\cdot\mathbb{E}_\theta[\tanh]$ piece, the second is the Stein contraction of the $\sqrt{\rho_X}Z_X$ factor with the tanh derivative (which produces $\beta(1-\langle q\rangle)$ times the $\langle n_X\rangle$ inside $G_n$). Reorganizing,

$$\big[1+\rho_X - \lambda_X\rho_X\beta(1-\langle q\rangle)\big]\langle n_X\rangle = \lambda_L\langle n_L\rangle + \lambda_U\langle n_U\rangle = \langle m\rangle, \tag{C.42}$$

the last identity following from (C.38) once the CLT replacement is performed inside the tanh. Solving (C.42) for $\langle n_X\rangle$,

$$\langle n_X\rangle = \frac{\langle m\rangle}{1+\rho_X - \lambda_X\rho_X\beta(1-\langle q\rangle)}, \qquad X\in\{L,U\}, \tag{C.43}$$

which is the semi-supervised analog of the supervised result in [21]. In the strict large dataset limit $\rho_X \to 0$ the denominator reduces to 1 and we obtain the collapse

$$\langle n_L\rangle = \langle n_U\rangle = \langle m\rangle. \tag{C.44}$$

Under (C.44) the argument of the hyperbolic tangent in (C.38) and (C.32) reduces to $\beta\langle m\rangle\, G(\chi)$ with the simplified combined mean

$$G(\chi) := \lambda\,\frac{1}{M_L r}\sum_a \chi^{1a} + (1-\lambda)\,\frac{1}{M_U r}\sum_c \chi^{1c} \xrightarrow{M\to\infty} 1 + \sqrt{\lambda^2\rho_L + (1-\lambda)^2\rho_U}\, Z, \tag{C.45}$$

where the two independent Gaussians $Z_L, Z_U$ collapse into a single standard Gaussian $Z\sim\mathcal{N}(0,1)$. The signal-multiplicative Gaussian noise of variance $\langle m\rangle^2(\lambda^2\rho_L+(1-\lambda)^2\rho_U)$ combines with the slow-noise cavity field $\theta\sqrt{\alpha\beta\sum_k d_k\tilde\mu_k\langle p^k\rangle}$ into a single Gaussian of variance

$$\alpha\sum_k d_k\,\frac{\tilde\mu_k^2\,\langle q\rangle}{\big[1-\beta\tilde\mu_k(1-\langle q\rangle)\big]^2} + \langle m\rangle^2(\lambda^2\rho_L + (1-\lambda)^2\rho_U), \tag{C.46}$$

obtained by substituting (C.29) into $\sum_k d_k\tilde\mu_k\langle p^k\rangle$ and identifying

$$\beta^{-1}\sum_k d_k\tilde\mu_k\langle p^k\rangle = \beta\alpha\sum_k d_k\,\frac{\tilde\mu_k^2\,\langle q\rangle}{\big[1-\beta\tilde\mu_k(1-\langle q\rangle)\big]^2}. \tag{C.47}$$

Collecting (C.38), (C.32), (C.29), (C.44), and (C.46), the RS self-consistency system in the large dataset limit takes the compact form of Section 4.3. This is the closed RS system used in Section 4 to draw the phase diagrams of Figs. 3–4.

### C.2. Zero-temperature limit of the RS self-consistency equations

We derive here in detail the closed zero-temperature system (4.45)–(4.46). The starting point is the large-dataset RS self-consistency system (4.41)–(4.43), which we rewrite here for the two independent order parameters $\langle m\rangle$ and $\langle q\rangle$ as

$$\langle m\rangle = \mathbb{E}_z\Bigg[\tanh\Bigg(\beta\langle m\rangle \tag{C.48}$$

$$+\beta z\sqrt{\alpha\sum_k d_k\,\frac{\tilde\mu_k^2\,\langle q\rangle}{\big[1-\beta\tilde\mu_k(1-\langle q\rangle)\big]^2} + \langle m\rangle^2\big(\lambda^2\rho_L+(1-\lambda)^2\rho_U\big)}\;\Bigg)\Bigg], \tag{C.49}$$

$$\langle q\rangle = \mathbb{E}_z\Bigg[\tanh^2\Bigg(\beta\langle m\rangle \tag{C.50}$$

$$+\beta z\sqrt{\alpha\sum_k d_k\,\frac{\tilde\mu_k^2\,\langle q\rangle}{\big[1-\beta\tilde\mu_k(1-\langle q\rangle)\big]^2} + \langle m\rangle^2\big(\lambda^2\rho_L+(1-\lambda)^2\rho_U\big)}\;\Bigg)\Bigg]. \tag{C.51}$$

As $\beta\to\infty$ the argument of the hyperbolic tangent in (C.51) diverges for almost every $z$, so $\tanh^2\to 1$ and the overlap saturates, $\langle q\rangle\to 1$. The rate of this saturation is controlled by the vanishing measure of the values of $z$ for which the argument stays finite: one has $1-\langle q\rangle = O(\beta^{-1})$, so that the product

$$\mathcal{Q} := \beta(1-\langle q\rangle) \tag{C.52}$$

remains finite in the limit. We may then set $\langle q\rangle \to 1$ in each numerator and replace $\beta\tilde{\mu}_k(1-\langle q\rangle) = \tilde{\mu}_k\,\beta(1-\langle q\rangle) \to \tilde{\mu}_k\mathcal{Q}$ in each denominator, obtaining

$$\alpha\sum_k d_k\,\frac{\tilde{\mu}_k^2\,\langle q\rangle}{\left[1-\beta\tilde{\mu}_k(1-\langle q\rangle)\right]^2} \xrightarrow{\beta\to\infty} \alpha\sum_k \frac{d_k\,\tilde{\mu}_k^2}{(1-\tilde{\mu}_k\mathcal{Q})^2}. \tag{C.53}$$

Turning to the magnetization, in (C.49) the hyperbolic tangent steepens into the sign function, $\tanh(\beta x)\to\mathrm{sgn}(x)$, and using (C.53) we obtain

$$\langle m\rangle = \int \mathcal{D}z\ \mathrm{sgn}\Big(\langle m\rangle + z\sqrt{\alpha\sum_k \frac{d_k\,\tilde{\mu}_k^2}{(1-\tilde{\mu}_k\mathcal{Q})^2} + \langle m\rangle^2(\lambda^2\rho_L+(1-\lambda)^2\rho_U)}\Big), \qquad \mathcal{D}z = \frac{dz}{\sqrt{2\pi}}\,e^{-z^2/2}. \tag{C.54}$$

The Gaussian integral of a sign function is elementary: for $A\in\mathbb{R}$ and $B>0$, $\mathrm{sgn}(A+Bz)=+1$ for $z>-A/B$ and $-1$ otherwise, hence

$$\int \mathcal{D}z\ \mathrm{sgn}(A+Bz) = \int_{-A/B}^{\infty}\mathcal{D}z - \int_{-\infty}^{-A/B}\mathcal{D}z = 1-2\,\Phi\Big(-\tfrac{A}{B}\Big) = \mathrm{erf}\Big(\frac{A}{\sqrt{2}\,B}\Big), \tag{C.55}$$

where $\Phi$ is the standard normal cumulative distribution and we used $1-2\Phi(-t)=\mathrm{erf}(t/\sqrt{2})$. Applying (C.55) to (C.54) with $A=\langle m\rangle$ and $B=\sqrt{\Sigma_0^2+\langle m\rangle^2(\lambda^2\rho_L+(1-\lambda)^2\rho_U)}$ gives the retrieval equation

$$\langle m\rangle = \mathrm{erf}\left(\frac{\langle m\rangle}{\sqrt{2\big(\alpha\sum_k \frac{d_k\,\tilde{\mu}_k^2}{(1-\tilde{\mu}_k\mathcal{Q})^2} + \langle m\rangle^2(\lambda^2\rho_L+(1-\lambda)^2\rho_U)\big)}}\right), \tag{C.56}$$

i.e. (4.45). The value of $\mathcal{Q}=\beta(1-\langle q\rangle)$ is not fixed by (C.56) and requires an independent closure obtained from (C.51); writing $1-\tanh^2=\mathrm{sech}^2$,

$$1-\langle q\rangle = \int \mathcal{D}z\ \mathrm{sech}^2\Big[\,\beta\,\Big(\,\langle m\rangle + z\sqrt{\alpha\sum_k \frac{d_k\,\tilde{\mu}_k^2}{(1-\tilde{\mu}_k\mathcal{Q})^2} + \langle m\rangle^2(\lambda^2\rho_L+(1-\lambda)^2\rho_U)}\,\Big)\Big], \tag{C.57}$$

and multiplying both sides by $\beta$,

$$\mathcal{Q} = \beta(1-\langle q\rangle) = \int \mathcal{D}z\ \beta\ \mathrm{sech}^2\Big[\,\beta\,\Big(\,\langle m\rangle + z\sqrt{\alpha\sum_k \frac{d_k\,\tilde{\mu}_k^2}{(1-\tilde{\mu}_k\mathcal{Q})^2} + \langle m\rangle^2(\lambda^2\rho_L+(1-\lambda)^2\rho_U)}\,\Big)\Big]. \tag{C.58}$$

The kernel $\beta\,\mathrm{sech}^2(\beta x)$ is a nascent Dirac delta: it is even, peaked at $x=0$ with width $O(\beta^{-1})$, and normalized to

$$\int_{-\infty}^{\infty}\beta\,\mathrm{sech}^2(\beta x)\,dx = \big[\tanh(\beta x)\big]_{-\infty}^{\infty} = 2, \tag{C.59}$$

so that $\beta\,\mathrm{sech}^2(\beta x)\xrightarrow{\beta\to\infty} 2\,\delta(x)$ in the distributional sense. Inserting this limit into (C.58),

$$\mathcal{Q} = \int \mathcal{D}z\ 2\,\delta\Big(\langle m\rangle + z\sqrt{\alpha\sum_k \frac{d_k\,\tilde{\mu}_k^2}{(1-\tilde{\mu}_k\mathcal{Q})^2} + \langle m\rangle^2(\lambda^2\rho_L+(1-\lambda)^2\rho_U)}\Big), \tag{C.60}$$

and, with $A=\langle m\rangle$ and $B=\sqrt{\alpha\sum_k \frac{d_k\,\tilde{\mu}_k^2}{(1-\tilde{\mu}_k\mathcal{Q})^2} + \langle m\rangle^2(\lambda^2\rho_L+(1-\lambda)^2\rho_U)}$, the scaling property $\delta(A+Bz)=|B|^{-1}\delta(z-z^\star)$ selects $z^\star=-A/B=-\langle m\rangle/\sqrt{\alpha\sum_k \frac{d_k\,\tilde{\mu}_k^2}{(1-\tilde{\mu}_k\mathcal{Q})^2} + \langle m\rangle^2(\lambda^2\rho_L+(1-\lambda)^2\rho_U)}$, whence

$$\mathcal{Q} = \frac{2}{\sqrt{\alpha\sum_k \frac{d_k\,\tilde{\mu}_k^2}{(1-\tilde{\mu}_k\mathcal{Q})^2} + \langle m\rangle^2(\lambda^2\rho_L+(1-\lambda)^2\rho_U)}}\,\frac{e^{-z^{\star 2}/2}}{\sqrt{2\pi}} \tag{C.61}$$

$$= \frac{2}{\sqrt{2\pi}\sqrt{\alpha\sum_k \frac{d_k\,\tilde{\mu}_k^2}{(1-\tilde{\mu}_k\mathcal{Q})^2} + \langle m\rangle^2(\lambda^2\rho_L+(1-\lambda)^2\rho_U)}}\,\exp\Big[-\frac{\langle m\rangle^2}{2\,\big(\,\alpha\sum_k \frac{d_k\,\tilde{\mu}_k^2}{(1-\tilde{\mu}_k\mathcal{Q})^2} + \langle m\rangle^2(\lambda^2\rho_L+(1-\lambda)^2\rho_U)}\Big]. \tag{C.62}$$

Using $2/\sqrt{2\pi}=\sqrt{2/\pi}$ this becomes

$$\mathcal{Q} = \sqrt{\frac{2}{\pi\big(\alpha\sum_k \frac{d_k\,\tilde{\mu}_k^2}{(1-\tilde{\mu}_k\mathcal{Q})^2} + \langle m\rangle^2(\lambda^2\rho_L+(1-\lambda)^2\rho_U)\big)}}\,\exp\Big[-\frac{\langle m\rangle^2}{2\big(\Sigma_0^2+\langle m\rangle^2(\lambda^2\rho_L+(1-\lambda)^2\rho_U)\big)}\Big], \tag{C.63}$$

i.e. (4.46). Eqs. (C.56) and (C.63) form the closed zero-temperature system in $(\langle m\rangle,\mathcal{Q})$.

## Appendix D. Convexity of the quenched pressure in the mixing parameter

We collect here the convexity properties invoked in Section 5, together with the uniform bounds on the pressure on which they rely. Throughout, the energy densities of (2.19) are normalized so that $-\beta\mathcal{H} = \beta N[\lambda\mathcal{E}_L + (1-\lambda)\mathcal{E}_U]$ is extensive, i.e. $\mathcal{E}_L = \frac{1}{2\Gamma_L}\sum_\mu \big(\sum_a n_L^{\mu a}\big)^2$ and $\mathcal{E}_U = \frac{1}{2\Gamma_U}\sum_{\mu,c}(n_U^{\mu c})^2$; since $|n^{\mu a}| \le 1$, this gives the deterministic bounds $0 \le \mathcal{E}_L \le \alpha_N/(2\mathcal{R}_L)$ and $0 \le \mathcal{E}_U \le \alpha_N/(2r^2)$, with $\alpha_N := K/N$.

We first record that the pressure is bounded uniformly in $N$ at any load and temperature. The lower bound follows from $-\beta\mathcal{H}_N \ge 0$, which gives $\mathcal{Z}_{N,\beta} \ge 2^N$ and hence $\mathcal{A}_N \ge \ln 2$; for the upper bound, $-\beta\mathcal{H}_N = \frac{\beta}{2}\boldsymbol{\sigma}^\top J^\lambda \boldsymbol{\sigma} \le \frac{\beta N}{2}\|J^\lambda\|_{\mathrm{op}}$ implies $\mathcal{Z}_{N,\beta} \le 2^N \exp\big(\frac{\beta N}{2}\|J^\lambda\|_{\mathrm{op}}\big)$, so that

$$\begin{aligned}\ln 2 \;\le\; \mathcal{A}_N \;&\le\; \ln 2 + \frac{\beta}{2}\,\mathbb{E}\|J^\lambda\|_{\mathrm{op}} \\ &\le\; \ln 2 + \frac{\beta c_0}{2}\left[\lambda\Big(1+\alpha_N+\sqrt{\tfrac{\alpha_N}{\mathcal{R}_L}}\Big) + \frac{1-\lambda}{r^2}\Big(\frac{1}{M_U}+\alpha_N+\sqrt{\tfrac{\alpha_N}{M_U}}\Big)\right],\end{aligned} \tag{D.1}$$

with $c_0$ a universal constant. The last inequality uses that $J^{(L)} = \frac{1}{N\mathcal{R}_L}\sum_\mu \bar{\boldsymbol{\eta}}^\mu_{(L)}(\bar{\boldsymbol{\eta}}^\mu_{(L)})^\top$ (cf. (4.1)) and $J^{(U)} = \frac{1}{N\Gamma_U}\sum_{\mu,c}\boldsymbol{\eta}^{\mu c}(\boldsymbol{\eta}^{\mu c})^\top$ are Gram matrices of random matrices whose entries, conditionally on $\boldsymbol{\xi}$, are independent, bounded and centered with variances $\mathcal{R}_L$ and 1 respectively, so that their expected squared operator norms are controlled by standard random-matrix estimates.[12]

We now turn to convexity. Writing $-\beta\mathcal{H} = \beta N[\mathcal{E}_U + \lambda(\mathcal{E}_L - \mathcal{E}_U)]$, which is affine in $\lambda$, a first differentiation of $N^{-1}\ln\mathcal{Z}_{N,\beta}$ gives

$$\frac{\partial}{\partial\lambda}\,\frac{\ln\mathcal{Z}_{N,\beta}}{N} = \frac{1}{N}\,\omega\big(\partial_\lambda(-\beta\mathcal{H})\big) = \beta\,\omega\big(\mathcal{E}_L - \mathcal{E}_U\big), \tag{D.2}$$

and a second differentiation, using $\partial_\lambda\omega(\mathcal{O}) = \beta N\,[\omega(\mathcal{O}(\mathcal{E}_L-\mathcal{E}_U)) - \omega(\mathcal{O})\,\omega(\mathcal{E}_L-\mathcal{E}_U)]$ with $\mathcal{O} = \mathcal{E}_L - \mathcal{E}_U$, yields

$$\frac{\partial^2}{\partial\lambda^2}\,\frac{\ln\mathcal{Z}_{N,\beta}}{N} = \beta^2 N\,\mathrm{Var}_\omega\big(\mathcal{E}_L - \mathcal{E}_U\big) \;\ge\; 0, \tag{D.3}$$

a Gibbs variance, hence non-negative for every $N$, every $\beta > 0$ and every realization of the dataset $S$. Convexity is preserved by the quenched average and by pointwise limits, so both $\mathcal{A}_N$ and any thermodynamic limit $\mathcal{A}$ are convex in $\lambda$; the exchange of $\partial_\lambda$ and $\mathbb{E}$ is licit because $|\partial_\lambda \ln\mathcal{Z}| \le \beta N(\mathcal{E}_L + \mathcal{E}_U)$ is uniformly bounded. Convexity in turn gives the chord inequality

$$\mathcal{A}(\beta,\lambda) \;\le\; \lambda\,\mathcal{A}(\beta,1) + (1-\lambda)\,\mathcal{A}(\beta,0), \tag{D.4}$$

so the pressure of the mixture never exceeds the interpolation of the two pure channels.

Two consequences for the mixing parameter follow at once. First, by (D.2) the derivative $d\mathcal{A}/d\lambda = \beta\,\mathbb{E}\,\omega(\mathcal{E}_L - \mathcal{E}_U)$ is non-decreasing, so the equipartition condition (5.5) has a solution set which is a closed interval, reduced to a single point $\lambda_{\mathrm{eq}}$ whenever the convexity is strict. Second, since $\mathcal{A}$ is convex, every interior solution of (5.5) is a global *minimum* of $\lambda \mapsto \mathcal{A}(\beta,\lambda)$, that is, a global *maximum* of the quenched free energy. Finally, the deterministic bounds on $\mathcal{E}_L, \mathcal{E}_U$ give, through (D.2),

$$\Big|\partial_\lambda\mathcal{A}_N\Big| \;\le\; \frac{\beta\alpha_N}{2}\big(\mathcal{R}_L^{-1} + r^{-2}\big) \qquad \text{uniformly in } N, \tag{D.5}$$

so that the family $\{\mathcal{A}_N(\beta,\cdot)\}_N$ is uniformly Lipschitz on $[0,1]$; combined with the uniform bound (D.1), an Arzelà–Ascoli argument on the compact $[0,1]$ upgrades pointwise convergence to uniform convergence.

The physical reading of these facts is discussed in Section 5: the convexity of $\mathcal{A}$ in $\lambda$ is the formal obstruction preventing equilibrium thermodynamics from selecting an interior mixing, and hence the reason why the operational optimum $\lambda^\star$ must be defined through the retrieval observables rather than through the pressure.

## Appendix E. Numerical methods

This appendix collects the details of the numerical work reported in the main text. Three logically distinct procedures are involved, and it is worth keeping them apart: a deterministic one-step dynamics at zero temperature, which is the direct counterpart of the signal-to-noise analysis of Section 3; an equilibrium Monte Carlo dynamics at finite temperature, which probes the thermodynamic phase of the network; and the numerical solution of the replica-symmetric self-consistency equations of Section 4. The first two share the construction of the dataset and of the coupling matrix, and differ only in the update rule, so we present them as a single routine with a branch on the temperature. Each procedure is given below in pseudocode, and we specify which figure relies on which.

### *E.1. Dataset and coupling matrix*

[12] See [54], combined with Talagrand's concentration inequality for convex Lipschitz functions of bounded independent entries; the rank-one conditional means of the $\mu = 1$ columns are absorbed in $c_0$.

**Algorithm 1:** Generation of one realization of the dataset and of the coupling matrix

**Input:** network size $N$, number of archetypes $K$, dataset sizes $M_L, M_U$, quality $r$, mixing $\lambda$
**Output:** coupling $J^\lambda$, target archetype $\boldsymbol{\xi}^1$, example pools $\mathcal{E}_L, \mathcal{E}_U$ of $\boldsymbol{\xi}^1$

**for** $\mu = 1$ **to** $K$ **do**
  draw $\xi_i^\mu \in \{-1,+1\}$ i.i.d. with $\mathbb{P}(\xi_i^\mu = \pm 1) = 1/2$, $i = 1,\dots,N$
  **for** $a = 1$ **to** $M_L$ **do**
    draw $\chi_i^{\mu a}$ i.i.d. with $\mathbb{P}(\chi_i^{\mu a} = +1) = (1+r)/2$, $\eta_i^{\mu a} \leftarrow \xi_i^\mu \chi_i^{\mu a}$ ▷ *labeled*
  **for** $c = 1$ **to** $M_U$ **do**
    draw $\chi_i^{\mu c}$ i.i.d. with $\mathbb{P}(\chi_i^{\mu c} = +1) = (1+r)/2$, $\eta_i^{\mu c} \leftarrow \xi_i^\mu \chi_i^{\mu c}$ ▷ *unlabeled*
$\Gamma_L \leftarrow M_L^2\big[r^2 + (1-r^2)/M_L\big]$, $\quad \Gamma_U \leftarrow M_U r^2$
$J_{ij}^{(L)} \leftarrow \dfrac{1}{N\Gamma_L} \displaystyle\sum_{\mu=1}^{K} \Big(\sum_{a=1}^{M_L} \eta_i^{\mu a}\Big)\Big(\sum_{b=1}^{M_L} \eta_j^{\mu b}\Big)$ ▷ *one collective variable per archetype*
$J_{ij}^{(U)} \leftarrow \dfrac{1}{N\Gamma_U} \displaystyle\sum_{\mu=1}^{K}\sum_{c=1}^{M_U} \eta_i^{\mu c}\eta_j^{\mu c}$ ▷ *one outer product per example*
$J^\lambda \leftarrow \lambda J^{(L)} + (1-\lambda) J^{(U)}$
$J_{ii}^\lambda \leftarrow 0$ for every $i$ ▷ *no self-interaction, cf. Remark 2*
$\mathcal{E}_L \leftarrow \{\boldsymbol{\eta}^{1a}\}_{a \le M_L}$, $\quad \mathcal{E}_U \leftarrow \{\boldsymbol{\eta}^{1c}\}_{c \le M_U}$
**return** $J^\lambda$, $\boldsymbol{\xi}^1$, $\mathcal{E}_L$, $\mathcal{E}_U$

One realization of the quenched disorder is generated as in Algorithm 1: the archetypes are drawn as i.i.d. Rademacher vectors according to (2.1), the examples are obtained by the multiplicative corruption (2.2), and the two Hebbian kernels are assembled and normalized as in (2.6)–(2.8). In agreement with Remark 2, the self-couplings are set *exactly* to zero, $J_{ii}^\lambda = 0$ for every $i = 1,\dots,N$; in the routines that never build the matrix explicitly, but only its action on a configuration, the same is achieved by subtracting $\sigma_i J_{ii}^\lambda$ from the local field. The disorder is regenerated from scratch at every realization, and the same realization is shared by all the values of $\lambda$ that are compared within a figure, so that the curves of the different protocols are correlated and their comparison is not blurred by sample-to-sample fluctuations.

### *E.2. Monte Carlo estimate of the retrieval magnetization*

Both numerical dynamics used in the paper are collected in Algorithm 2, which branches on the temperature.

The branch $T = 0$ is the deterministic one-step protocol that tests the signal-to-noise predictions of Section 3. The network is initialized on the archetype to be retrieved, $\boldsymbol{\sigma} = \boldsymbol{\xi}^1$, and all the spins are updated once, *synchronously*, so that the returned quantity is the one-step Mattis magnetization $m^{(1)}$ of (3.8): it measures the local stability of the archetype under a single update, not the phase in which the network eventually settles, and it must not be read as an equilibrium magnetization. This branch produces the symbols of Fig. 2 and the heat map in the left panel of Fig. 6, as well as the inset of the right panel of Fig. 3.

The branch $T > 0$ is an equilibrium Monte Carlo simulation. The dynamics is the *asynchronous* single-spin-flip Glauber dynamics: a site is drawn uniformly at random and updated with the heat-bath probability, the local fields of the remaining sites being updated incrementally after each accepted flip, and one sweep consists of $N$ such elementary steps, so that on average every spin is addressed once per sweep. The initial configuration is one example of the archetype to be retrieved, drawn from the labeled or from the unlabeled pool with equal probability, so that the network is asked to reconstruct the archetype starting from a noisy example of it. Each run consists of $n_{\mathrm{eq}} = 2000$ equilibration sweeps, which are discarded, followed by $n_{\mathrm{meas}} = 200$ measurements separated by $t_{\mathrm{skip}} = 3$ sweeps, whose average is the estimate for that realization. This branch produces the symbols of the central and right panels of Fig. 5 and of the central panel of Fig. 6, always at $T = 0.001$.

In both branches the estimate is averaged over $n_{\mathrm{real}}$ independent realizations of the dataset, and the error bars reported in the figures are the standard error of the mean over realizations: $n_{\mathrm{real}} = 50$ for the equilibrium simulations, 32 for the heat map of Fig. 6 and 30 for the bisection of Fig. 2.

### *E.3. Numerical solution of the RS equations*

Every continuous curve and every heat map of Figs. 3, 4, 5 and 6 which is not explicitly marked as Monte Carlo data comes from Algorithm 3, namely from the numerical solution of the self-consistency Eqs. (4.41)–(4.44) in the large-dataset form used throughout the paper, or of their zero-temperature limit (4.45)–(4.46).

Three points of the algorithm deserve a comment. First, since the eigen-channel decomposition (4.20) makes the slow noise a sum of terms with denominators $1 - \beta\tilde{\mu}_k(1 - \langle q_{12}\rangle)$, the overlap is clamped throughout the iteration above the value at which the largest

**Algorithm 2:** Monte Carlo estimate of the retrieval magnetization: deterministic one-step dynamics at $T = 0$, asynchronous Glauber dynamics at $T > 0$

**Input:** $J^{\lambda}, \boldsymbol{\xi}^1, \mathcal{E}_L, \mathcal{E}_U$ from Algorithm 1; temperature $T$; $n_{\mathrm{eq}}, n_{\mathrm{meas}}, t_{\mathrm{skip}}$
**Output:** Mattis magnetization $m$ of the realization

**if** $T = 0$ **then**
  $\boldsymbol{\sigma} \leftarrow \boldsymbol{\xi}^1$
  $h_i \leftarrow \sum_j J^{\lambda}_{ij}\sigma_j$ for every $i$
  $\sigma_i \leftarrow \mathrm{sign}(h_i)$ for every $i$, with $\mathrm{sign}(0) := +1$ ▷ *one synchronous update*
  $m \leftarrow \left|\sum_i \sigma_i \xi^1_i\right|/N$ ▷ *one-step magnetization* (3.8)
**else**
  $\beta \leftarrow 1/T$
  draw $\boldsymbol{\sigma}$ uniformly from $\mathcal{E}_L$ or from $\mathcal{E}_U$, each with probability $1/2$
  $h_i \leftarrow \sum_j J^{\lambda}_{ij}\sigma_j$ for every $i$
  **for** $t = 1$ **to** $n_{\mathrm{eq}}$ **do**
    Sweep$(\boldsymbol{\sigma}, \boldsymbol{h}, \beta)$ ▷ *equilibration, discarded*
  **for** $k = 1$ **to** $n_{\mathrm{meas}}$ **do**
    **for** $t = 1$ **to** $t_{\mathrm{skip}}$ **do**
      Sweep$(\boldsymbol{\sigma}, \boldsymbol{h}, \beta)$
    $m_k \leftarrow \left|\sum_i \sigma_i \xi^1_i\right|/N$
  $m \leftarrow \frac{1}{n_{\mathrm{meas}}}\sum_k m_k$ ▷ *time average*
**return** $m$
**Procedure** Sweep$(\boldsymbol{\sigma}, \boldsymbol{h}, \beta)$**:**
  **for** $s = 1$ **to** $N$ **do**
    draw $i$ uniformly in $\{1, \dots, N\}$
    $\sigma_i^{\mathrm{old}} \leftarrow \sigma_i$
    $\sigma_i \leftarrow +1$ with probability $\left[1 + \tanh(\beta h_i)\right]/2$, else $\sigma_i \leftarrow -1$
    **if** $\sigma_i \neq \sigma_i^{\mathrm{old}}$ **then**
      $h_j \leftarrow h_j + (\sigma_i - \sigma_i^{\mathrm{old}})J^{\lambda}_{ij}$ for every $j$ ▷ *incremental field update*

**Algorithm 3:** Numerical solution of the replica-symmetric self-consistency equations at one point $(\alpha, \beta, \lambda)$

**Input:** load $\alpha$, inverse temperature $\beta$, mixing $\lambda$, dataset parameters $r, \rho_L, \rho_U$; damping $\omega = 0.4$, tolerance $\varepsilon_{\mathrm{tol}} = 10^{-9}$, warm start $(m_0, q_0)$
**Output:** $\langle m \rangle, \langle q_{12} \rangle$

compute the spectrum $\{(\tilde{\mu}_k, d_k)\}$ of $SW$ from (4.20) and $R(\lambda)$ from (3.14)
$q_{\min} \leftarrow \max\{0,\ 1 - 1/(\beta\tilde{\mu}_{\max})\}$ ▷ *keeps every denominator positive*
$m \leftarrow m_0, \quad q \leftarrow \max(q_0, q_{\min})$
**repeat**
  $\Sigma^2 \leftarrow \alpha \sum_k d_k \dfrac{\tilde{\mu}_k^2\, q}{\left[1 - \beta\tilde{\mu}_k(1-q)\right]^2}, \qquad \sigma \leftarrow \sqrt{\Sigma^2 + m^2 R(\lambda)}$
  **if** $\beta\sigma > 30$ **then**
    $m^{\mathrm{new}} \leftarrow \mathrm{erf}\left(m/(\sigma\sqrt{2})\right), \qquad q^{\mathrm{new}} \leftarrow 1 - \dfrac{1}{\beta\sigma}\sqrt{\dfrac{2}{\pi}}\, e^{-m^2/(2\sigma^2)}$ ▷ $T \to 0$ *branch*
  **else**
    $m^{\mathrm{new}} \leftarrow \displaystyle\int \mathcal{D}z\, \tanh\left(\beta m + \beta\sigma z\right), \qquad q^{\mathrm{new}} \leftarrow 1 - \frac{1}{\beta\sigma}\int \mathcal{D}z\, z\, \tanh\left(\beta m + \beta\sigma z\right)$
  $q^{\mathrm{new}} \leftarrow \max(q^{\mathrm{new}}, q_{\min})$
  $m \leftarrow (1-\omega)m + \omega\, m^{\mathrm{new}}, \qquad q \leftarrow (1-\omega)q + \omega\, q^{\mathrm{new}}$ ▷ *damped Picard step*
**until** $|m^{\mathrm{new}} - m| < \varepsilon_{\mathrm{tol}}$ ***and*** $|q^{\mathrm{new}} - q| < \varepsilon_{\mathrm{tol}}$
**return** $m, q$

of them would change sign, so that the slow noise stays positive along the whole iteration path. Second, the Gaussian averages are discretized on a uniform grid of 300 nodes spanning $z \in [-12, 12]$, weighted by the Gaussian measure, and when $\beta\sigma$ exceeds 30 the iteration switches to the closed-form zero-temperature expressions, which avoids evaluating tanh and $\mathrm{sech}^2$ in a regime where they

are numerically saturated. Third, the retrieval branch is followed by continuation in the load: the iteration at a given $\alpha$ is initialized at the solution found at the previous, smaller value of $\alpha$, so that the retrieval solution is tracked as long as it exists rather than being re-selected from scratch at every point.

The critical load $\alpha_c$ is then defined as the largest $\alpha$ carrying a retrieval solution, and it is located by scanning 120 values of $\alpha$ until the retrieval solution is lost, followed by 30 bisection steps between the last value with $\langle m \rangle > 0.05$ and the first one without, which brackets $\alpha_c$ to within $10^{-5}$.

## Data availability

No data was used for the research described in the article.

## References


[1] O. Chapelle, B. Schölkopf, A. Zien, Semi-supervised learning, IEEE Trans. Neural Netw. 20 (3) (2009) 542–542.
[2] J.E. Van Engelen, H.H. Hoos, A survey on semi-supervised learning, Mach. Learn. 109 (2) (2020) 373–440.
[3] Y. Grandvalet, Y. Bengio, Semi-supervised learning by entropy minimization, in: Advances in Neural Information Processing Systems, vol. 17, MIT Press, 2005, pp. 529–536.
[4] D.P. Kingma, D.J. Rezende, S. Mohamed, M. Welling, Semi-supervised learning with deep generative models, in: Advances in Neural Information Processing Systems, vol. 27, 2014, pp. 3581–3589.
[5] D.J. Amit, Modeling Brain Function: The World of Attractor Neural Networks, Cambridge University Press, 1989.
[6] M. Mézard, G. Parisi, M.A. Virasoro, Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications, vol. 9, World Scientific Publishing Company, 1987.
[7] J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. USA 79 (1982) 2554–2558.
[8] D. Sherrington, S. Kirkpatrick, Solvable model of a spin-glass, Phys. Rev. Lett. 35 (26) (1975) 1792.
[9] A.C.C. Coolen, R. Kühn, P. Sollich, Theory of Neural Information Processing Systems, OUP Oxford, 2005.
[10] H. Nishimori, Statistical Physics of Spin Glasses and Information Processing: an Introduction, no. 111, Clarendon Press, 2001.
[11] J. Hertz, A. Krogh, R.G. Palmer, H. Horner, Introduction to the Theory of Neural Computation, American Institute of Physics, 1991.
[12] D.J. Amit, H. Gutfreund, H. Sompolinsky, Storing infinite numbers of patterns in a spin-glass model of neural networks, Phys. Rev. Lett. 55 (1985) 1530–1533.
[13] D. Krotov, J.J. Hopfield, Dense associative memory for pattern recognition, Adv. Neural Inf. Process. Syst. (2016) 1180–1188.
[14] D. Krotov, J. Hopfield, Dense associative memory is robust to adversarial inputs, Neural Comput. 30 (2018) 3151–3167.
[15] P. Baldi, S.S. Venkatesh, Number of stable points for spin-glasses and neural networks of higher orders, Phys. Rev. Lett. 58 (1987).
[16] E. Gardner, Multiconnected neural network models, J. Phys. A: Gen. Phys. 20 (1987).
[17] H. Bao, R. Zhang, Y. Mao, The capacity of the dense associative memory networks, Neurocomputing 469 (2022) 198–208.
[18] A. Bovier, B. Niederhauser, The spin-glass phase-transition in the hopfield model with p-spin interactions, Adv. Theor. Math. Phys. 5 (2001) 1001–1046.
[19] E. Agliari, F. Alemanno, A. Barra, A. Fachechi, Generalized Guerra's interpolation schemes for dense associative neural networks, Neural Netw. 128 (2020) 254–267.
[20] L. Albanese, F. Alemanno, A. Alessandrelli, A. Barra, Replica symmetry breaking in dense hebbian neural networks, J. Stat. Phys. 189 (2) (2022) 1–41.
[21] F. Alemanno, M. Aquaro, I. Kanter, A. Barra, E. Agliari, Supervised hebbian learning, Europhys. Lett. 141 (1) (2023) 11001.
[22] E. Agliari, F. Alemanno, A. Barra, G.D. Marzo, The emergence of a concept in shallow neural networks, Neural Netw. 148 (2022) 232–253.
[23] E. Agliari, G.D. Marzo, Tolerance versus synaptic noise in dense associative memories, Eur. Phys. J. Plus 135 (2020) 11.
[24] L. Albanese, A. Barra, P. Bianco, F. Durante, D. Pallara, Hebbian learning from first principles, J. Math. Phys. 65 (11) (2024) 113302.
[25] S. Shalev-Shwartz, S. Ben-David, Understanding Machine Learning: From Theory To Algorithms, Cambridge University Press, 2014.
[26] T. Fujii, H. Ito, S. Miyoshi, Statistical–mechanical analysis connecting supervised learning and semi-supervised learning, J. Phys. Soc. Japan 86 (6) (2017) 064003.
[27] G. Getz, N. Shental, E. Domany, Semi-supervised learning – a statistical physics approach, 2005, arXiv preprint cs/0604011. 2006.Presented at the 22nd ICML Workshop on Learning with Partially Classified Training Data, Bonn.
[28] G. Ver Steeg, A. Galstyan, A.E. Allahverdyan, Statistical mechanics of semi-supervised clustering in sparse graphs, J. Stat. Mech. Theory Exp. 2011 (08) (2011) P08009.
[29] P. Zhang, C. Moore, L. Zdeborová, Phase transitions in semisupervised clustering of sparse networks, Phys. Rev. E 90 (5) (2014) 052802.
[30] H. Cui, L. Saglietti, L. Zdeborová, Large deviations of semisupervised learning in the stochastic block model, Phys. Rev. E 105 (3) (2022) 034108.
[31] A. Bertoni, M. Frasca, G. Valentini, COSNet: A cost sensitive neural network for semi-supervised learning in graphs, in: Machine Learning and Knowledge Discovery in Databases, ECML PKDD, in: Lecture Notes in Computer Science, vol. 6911, Springer, 2011, pp. 219–234.
[32] M. Frasca, A. Bertoni, M. Re, G. Valentini, A neural network algorithm for semi-supervised node label learning from unbalanced data, Neural Netw. 43 (2013) 84–98.
[33] V. Tutevych, R. Memmesheimer, L. Eichler, D. Pavlichenko, F. Schilke, R. Krudewig, S. Behnke, Efficient image annotation via semi-supervised object segmentation with label propagation, 2026, arXiv preprint arXiv:2604.22992.
[34] F. Guerra, Broken replica symmetry bounds in the mean field spin glass model, Comm. Math. Phys. 233 (2003) 1–12.
[35] A. Barra, G. Genovese, F. Guerra, The replica symmetric approximation of the analogical neural network, J. Stat. Phys. 140 (2010) 784–796.
[36] F. Guerra, F.L. Toninelli, The thermodynamic limit in mean field spin glass models, Comm. Math. Phys. 230 (2002) 71–79.
[37] P. Carmona, Y. Hu, Universality in sherrington-kirkpatrick's spin glass model, Ann. L'Institut Henri Poincare (B) Probab. Stat. 42 (2006).
[38] G. Genovese, Universality in bipartite mean field spin glasses, J. Math. Phys. 53 (2012).
[39] E. Agliari, L. Albanese, F. Alemanno, A. Alessandrelli, A. Barra, F. Giannotti, D. Lotito, D. Pedreschi, Dense hebbian neural networks: A replica symmetric picture of unsupervised learning, Phys. A 627 (2023) 129143.
[40] E. Agliari, L. Albanese, F. Alemanno, A. Alessandrelli, A. Barra, F. Giannotti, D. Lotito, D. Pedreschi, Dense hebbian neural networks: A replica symmetric picture of supervised learning, Phys. A 626 (2023) 129076.
[41] M. Talagrand, Rigorous results for mean field models for spin glasses, Theoret. Comput. Sci. 265 (1–2) (2001) 69–77.
[42] E. Agliari, A. Barra, P. Bianco, A. Fachechi, D. Pallara, The thermodynamic limit in mean field neural networks, 2024, arXiv preprint arXiv:2409.10145.
[43] J.R.L. de Almeida, D.J. Thouless, Stability of the Sherrington-Kirkpatrick solution of a spin glass model, J. Phys. A: Math. Gen. 11 (5) (1978) 983–990.
[44] L. Albanese, A. Alessandrelli, A. Annibale, A. Barra, About the de almeida–thouless line in neural networks, Phys. A 633 (2024) 129372.
[45] M. Mézard, A. Montanari, Information, Physics, and Computation, Oxford University Press, 2009.
[46] H. Hellmann, Einführung in die Quantenchemie, Franz Deuticke, Leipzig, 1937.

[47] R.P. Feynman. Forces in molecules, Phys. Rev. 56 (1939) 340–343.
[48] P. Güttinger, Das verhalten von atomen im magnetischen drehfeld, Z. Phys. 73 (1932) 169–184.
[49] E. Agliari, A. Alessandrelli, A. Barra, M.S. Centonze, F. Ricci-Tersenghi, Generalized hetero-associative neural networks, J. Stat. Mech. Theory Exp. 2025 (1) (2025) 013302.
[50] E. Agliari, A. Alessandrelli, A. Barra, M. Centonze, F. Ricci-Tersenghi, Networks of hebbian networks: more is different, Neural Netw. 194 (2026) 108181.
[51] A. Alessandrelli, A. Barra, A. Ladiana, A. Lepre, F. Ricci-Tersenghi, Supervised and unsupervised protocols for hetero-associative neural networks, 2025, p. 130871,
[52] A. Alessandrelli, A. Barra, A. Ladiana, A. Lepre, F. Ricci-Tersenghi, Beyond disorder: unveiling cooperativeness in multidirectional associative memories, 2025, arXiv preprint arXiv:2503.04454.
[53] E. Agliari, A. Fachechi, P.D. Mourao, The beneficial role of noises for disentanglement tasks in modular hebbian networks, Phys. A 682 (2026) 131134.
[54] R. Latala, Some estimates of norms of random matrices, Proc. Amer. Math. Soc. 133 (5) (2005) 1273–1282.