Adaptive Path Interpolation Method for Sparse Systems: Application to a Censored Block Model

Jean Barbier, Chun Lam Chan and Nicolas Macris
Laboratoire de Théorie des Communications, Faculté Informatique et Communications,
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Suisse.
{jean.barbier, chunlam.chan, nicolas.macris}@epfl.ch

Abstract

Recently a new adaptive path interpolation method has been developed as a simple and versatile scheme to calculate exactly the asymptotic mutual information (or free energy) of Bayesian inference problems defined on dense factor graphs. These include random linear and generalized estimation, superposition codes, or low rank matrix and tensor estimation. For all these systems the adaptive interpolation method directly proves that the replica symmetric prediction is exact, and this in a simple and unified manner. When the underlying factor graph of the inference problem is sparse the replica prediction is considerably more complicated and rigorous results are often lacking or obtained by rather complicated methods. In this work we show how to extend the adaptive path interpolation method to sparse systems. We concentrate on a Censored Block Model, where hidden variables are measured through a binary erasure channel, for which we fully prove the replica prediction.

CONTENTS

I. Introduction

Much progress has been achieved recently in Bayesian inference of high dimensional problems. It has been possible to develop rigorous methods in order to derive exact "single letter" variational formulas for the mutual information (or free energy) in the asymptotic limit of the number of variables tending to infinity, when the prior and all hyperparameters of the problem are assumed to be known (this is referred as the Bayes-optimal setting). Such formulas have often been first conjectured on the basis of the replica and cavity methods of statistical mechanics of disordered spin systems and are also known as "replica symmetric" formulas [1, 2]. Examples where *full* proofs have been achieved are random linear estimation and compressed sensing [3–5], generalized estimation and learning (for single layer networks) [6], low-rank matrix and tensor estimation [7–11]. Invariably, the Guerra-Toninelli interpolation method [12] has been used to derive one-sided bounds (we note that [5] is exceptional in that it does not seem to rely directly on the same interpolation). For the converse bounds typically other ideas have usually been necessary, such as spatial coupling [3, 4, 7] or the Aizenman-Sims-Starr principle [8–10]. Recently two of us introduced a new interpolation scheme, called *adaptive path interpolation method*,[1] that allows to derive the replica symmetric formulas in a more straightforward and unified manner [13]. The new method is quite generic once the mean field solution has been identified and is directly applicable when the concentration of the "overlap" can be proved. Overlap concentration itself follows from variants of Ghirlanda-Guerra identities [14] adapted to Bayesian inference combined with the so-called Nishimori identities (see [13]).

The successes of the adaptive interpolation method have so far been limited to inference models with a dense underlying factor graph. It is therefore desirable to see to what extent the method can be developed when the factor graph is instead sparse. Typical examples of such systems are Low-Density Parity-Check codes, Low-Density Generator-Matrix codes, or the Stochastic and Censored Block models. It is fair to say that the replica symmetric formulas for the mutual information is much more complicated in such models. On one hand, besides the measurements (or channel outputs), the graph is also random, and on the other hand the single letter variational problem involves a functional over a set of measures (instead of scalars). Existing rigorous derivations of the replica formulas have so far been achieved using a combination of the interpolation method (first developed by [15] for sparse models) and spatial coupling [16] or the Aizenman-Sims-Starr principle [17, 18]. In this work we consider a simple version of the Censored Block model for which we fully develop the adaptive interpolation method. We believe that this constitutes a first step towards an analysis of more complicated models via this relatively simple method.

In the Censored Block model one has a set of n hidden binary variables. One observes products of n/R random K-tuples ($R > 0$ is interpreted as the measurement rate or $1/R$ as the fraction of measurements) through a noisy channel, and the goal is to reconstruct a good estimate of the hidden variables from the noisy observations. This system can also be interpreted as a Low-Density Generator-Matrix code ensemble on a factor graph with degree K factor nodes and variable nodes with Poisson degrees (when $n \to +\infty$), or in the language of statistical mechanics as a K-spin Ising model. The model has been discussed when the measurement channel is a Binary Symmetric Channel in [19] and the replica formula proven in this case [18]. Here, we consider a simpler situation where the measurement channel is the Binary Erasure Channel (BEC) for which the adaptive interpolation method can be completely developed. As we will see this method requires concentration results for overlaps. On the BEC this issue is much simpler because only the lowest order overlap matters, while on other channels one has to take into account an infinite number of them. This is the only aspect of the method that we have not (yet) extended to other channels.

The paper is organized as follows. In Section II-A we give a precise formulation of the model and state the main result of this paper (Theorem II.2). In Section III we review two important tools used throughout our analysis, namely the Nishimori identities and the Griffiths-Kelly-Sherman inequalities. The adaptive interpolation method for the sparse graph models is formulated in Section IV and the core of the proof of Theorem II.2 is also developed. This section contains the main new technical ideas of this paper. Overlap concentration is proved in Section VI and a series of more technical results are found in Section VII and in the appendices.

II. Setting and main result

A. Censored Block Model

We shall denote binary variables by $\sigma_i \in \{-1, +1\}$, $i = 1, \dots, n$ and vectors of such variables by $\underline{\sigma} = (\sigma_1, \dots, \sigma_n) \in \{-1, +1\}^n$. Subsets $S \subset \{1, \dots, n\}$ with at least two elements are always denoted by capital letters. For the product of binary variables in a subset S we use the shorthand notation $\sigma_S \equiv \prod_{i\in S} \sigma_i$. If there is a possible confusion between small and

[1]In its original version on dense graphs the method was called "stochastic interpolation method" but the new terminology is more appropriate.

capital letter subscripts we occasionally use more specific notations. Below the integer $K \geq 2$ and the rate $R \in \mathbb{R}_+$ are fixed independent of n.

In the Censored Block Model n *hidden* binary variables $\sigma_1^0, \ldots, \sigma_n^0$ are i.i.d. according to a uniform prior $P_0(\sigma_i^0) = \frac{1}{2}\delta_{\sigma_i^0,+1} + \frac{1}{2}\delta_{\sigma_i^0,-1}$, where $p_0 \in [0,1]$. A *noiseless measurement* consists in a product $\sigma_{i_1}^0 \sigma_{i_2}^0 \ldots \sigma_{i_K}^0$ of a K-tuple of variables drawn uniformly at random. The K-tuple is identified with a subset $A \equiv \{i_1, \ldots, i_K\} \subset \{1, \ldots, n\}$ and we set $\sigma_A^0 \equiv \sigma_{i_1}^0 \sigma_{i_2}^0 \ldots \sigma_{i_K}^0$. Of course $\sigma_A^0 = \pm 1$. The *true observations* $J_A \in \mathbb{R}$ are noisy versions of these products obtained through a binary input memoryless channel described by some transition probability $Q(J_A|\sigma_A^0)$. For large n the total number of observations m asymptotically follows a Poisson distribution with mean n/R, i.e., $m \sim \text{Poi}(n/R)$. We shall also index the observations as $A = 1, \ldots, m$.

Let us now describe the Bayesian setting used here to determine the information theoretic limits for reconstructing the hidden variables. From the Bayes rule we have

$$P(\underline{\sigma}|\underline{J}) = \frac{\prod_{i=1}^n P_0(\sigma_i) \prod_{A=1}^m Q(J_A|\sigma_A)}{\sum_{\underline{\sigma}\in\{-1,+1\}^n} \prod_{i=1}^n P_0(\sigma_i) \prod_{A=1}^m Q(J_A|\sigma_A)}.$$

Dividing both the numerator and denominator by $\prod_{A=1}^m Q(J_A|\sigma_A = +1)$, the posterior $P(\underline{\sigma}|\underline{J})$ can be rewritten as

$$P(\underline{\sigma}|\underline{\tilde{J}}) = \frac{1}{Z} \exp \sum_{A=1}^m \tilde{J}_A(\sigma_A - 1), \tag{1}$$

where

$$\tilde{J}_A \equiv \frac{1}{2} \ln \frac{Q(J_A|+1)}{Q(J_A|-1)},$$
$$Z \equiv \sum_{\underline{\sigma}\in\{-1,+1\}^n} \exp \sum_{A=1}^m \tilde{J}_A(\sigma_A - 1).$$

We will often use the language and notations of statistical mechanics. The normalization Z is then also called the partition function. The bipartite factor graph $\mathcal{G}$ underlying (1) contains variable nodes $i = 1, \ldots, n$ and constraint nodes $A = 1, \ldots, m$. Each variable node i "carries" variable σ_i and each constraint node A "carries" the log-likelihood-ratio $\tilde{J}_A$ and uniformly connects to K variable nodes $i_1, \ldots, i_K$. As said before we may identify $A \equiv \{i_1, \ldots, i_K\}$. Distribution (1) can be interpreted as the Gibbs distribution of a *random* spin system (or spin glass). The expectation with respect to (1) will be denoted by a bracket $\langle - \rangle$. This distribution is itself random because of the randomness of: i) the factor graph $\mathcal{G}$ ensemble; ii) the observations $\underline{J}$ given the hidden vector $\underline{\sigma}^0$; and iii) the hidden vector $\underline{\sigma}^0$. It is equivalent to work in terms of observations $\underline{J}$ or associated half-log-likelihood ratios $\underline{\tilde{J}}$, which are (formally) distributed according to

$$\prod_{i=1}^n P_0(\sigma_i^0) \prod_{A=1}^m \mathsf{c}(\tilde{J}_A|\sigma_A^0) d\tilde{J}_A = \prod_{i=1}^n P_0(\sigma_i^0) \prod_{A=1}^m Q(J_A|\sigma_A^0) dJ_A. \tag{2}$$

Most of the time it will be more convenient for us to refer directly to half-log-likelihood ratios. The graph, the observations and the hidden vector are called "quenched" random variables (r.v.) because for a given instance of the problem their realization is *fixed*, in contrast with the dynamical (or "annealed") r.v. $\underline{\sigma}$. Expectations with respect to these quenched variables are denoted $\mathbb{E}_{\mathcal{G}}$ and $\mathbb{E}_{\underline{\sigma}^0}\mathbb{E}_{\underline{\tilde{J}}|\underline{\sigma}^0}$.

Let $H(\underline{\sigma}|\underline{\tilde{J}}) \equiv -\sum_{\underline{\sigma}} P(\underline{\sigma}|\underline{\tilde{J}}) \ln P(\underline{\sigma}|\underline{\tilde{J}})$ be the conditional entropy of the hidden variables given fixed observations. It is easy to see that the average conditional entropy (per variable) is given by the average *free entropy*

$$\frac{1}{n}\mathbb{E}_{\mathcal{G}}\mathbb{E}_{\underline{\sigma}^0}\mathbb{E}_{\underline{\tilde{J}}|\underline{\sigma}^0} H(\underline{\sigma}|\underline{\tilde{J}}) = \frac{1}{n}\mathbb{E}_{\mathcal{G}}\mathbb{E}_{\underline{\sigma}^0}\mathbb{E}_{\underline{\tilde{J}}|\underline{\sigma}^0} \ln Z. \tag{3}$$

We refer readers to e.g. [20, Chapter 1], [21] for details. The singularities, as a function of noise, of this limiting quantity when $n \to +\infty$ give us the information theoretic reconstruction thresholds, or the location of "phase transitions" in physics language.

B. The replica symmetric formula for the average conditional entropy

The cavity method [1] predicts that the asymptotic normalized average conditional entropy is accessible from the following "replica symmetric" functional. This functional is an "average form" of the Bethe free entropy expression. Details of the relationship between the replica symmetric functional and Bethe free entropy can be found in [22, Appendix VII].

Definition II.1 (The replica symmetric free entropy functional)**.** *Let V be a r.v. with distribution x, and V_i, $i = 1, \ldots, K$ i.i.d. copies of V. Let*[2]

$$U = \tanh^{-1}\Big(\tanh\tilde{J}\prod_{i=1}^{K-1}\tanh V_i\Big), \tag{4}$$

and U_B, $B = 1, \ldots, l$ are i.i.d. copies of U where l is a Poisson distributed integer $\mathrm{Poi}(K/R)$*. Let $\sigma^0 \in \{-1, +1\}$ uniformly and $\tilde{J}$ be a log-likelihood r.v. distributed as described above. The replica symmetric free entropy functional is defined to be*

$$h_{\mathrm{RS}}(\mathsf{x}) \equiv \mathbb{E}_l\mathbb{E}_{\sigma^0}\mathbb{E}_{\tilde{J}|\sigma^0}\mathbb{E}_{\underline{U}}\mathbb{E}_{\underline{V}}\Big[\ln\Big(\prod_{B=1}^{l}(1+\tanh U_B) + \prod_{B=1}^{l}(1-\tanh U_B)\Big) - \frac{K-1}{R}\ln\Big(1+\tanh\tilde{J}\prod_{i=1}^{K}\tanh V_i\Big) - \frac{1}{R}\ln(1+\tanh\tilde{J})\Big]. \tag{5}$$

While most of our analysis can be extended to general (symmetric) memoryless channels our main result is fully proved for the Binary Erasure Channel. This channel has transition probability $Q(J_A|\sigma_A^0) = (1-q)\delta_{J_A,\sigma_A^0} + q\delta_{J_A,0}$. From (2) we get $\mathsf{c}(\tilde{J}_A|\sigma_A^0) = (1-q)\delta_{\sigma_A^0\tilde{J}_A,+\infty} + q\delta_{\tilde{J}_A,0}$. The set of distributions with point masses at $\{0, +\infty\}$ plays a special role and will be called $\mathcal{B}$. Moreover we adopt the notation (from coding theory) Δ_0 and Δ_∞ for these point masses. Thus any distribution $\mathsf{x} \in \mathcal{B}$ is of the from $\mathsf{x} = x\Delta_0 + (1-x)\Delta_\infty$, $x \in [0, 1]$.

Our main result is the proof, through the use of the adaptive interpolation method for sparse graphs, of the following theorem:

Theorem II.2 (The replica symmetric formula is exact)**.** *For a Censored Block Model with observations obtained through a Binary Erasure Channel as described above we have*

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}_{\mathcal{G}}\mathbb{E}_{\underline{\sigma}^0}\mathbb{E}_{\underline{\tilde{J}}|\underline{\sigma}^0}H(\underline{\sigma}|\underline{\tilde{J}}) = \sup_{\mathsf{x}\in\mathcal{B}} h_{\mathrm{RS}}(\mathsf{x}). \tag{6}$$

This theorem is a direct consequence of two main Propositions IV.4 and IV.9 proved in Section IV.

III. Two preliminary tools

In this section we review standard material [23, 24] which is needed in our analysis.

A. Nishimori identities

1) Tower property: Consider the quantity $\prod_{S\in\mathcal{C}}\sigma_S^0\prod_{S\in\mathcal{C}}\langle\sigma_S\rangle$ for a given graph and any collection $\mathcal{C}$ of subsets $S \subset \{1, \ldots, n\}$. The same subset can occur many times in a collection. From Bayes formula, for a given factor graph $\mathcal{G}$,

$$\mathbb{E}_{\underline{\sigma}^0}\mathbb{E}_{\underline{\tilde{J}}|\underline{\sigma}^0}\Big[\prod_{S\in\mathcal{C}}\sigma_S^0\prod_{S\in\mathcal{C}}\langle\sigma_S\rangle\Big] = \mathbb{E}_{\underline{\tilde{J}}}\mathbb{E}_{\underline{\sigma}^0|\underline{\tilde{J}}}\Big[\prod_{S\in\mathcal{C}}\sigma_S^0\prod_{S\in\mathcal{C}}\langle\sigma_S\rangle\Big]$$

which is nothing else than the identity

$$\mathbb{E}_{\underline{\sigma}^0}\mathbb{E}_{\underline{\tilde{J}}|\underline{\sigma}^0}\Big[\prod_{S\in\mathcal{C}}\sigma_S^0\prod_{S\in\mathcal{C}}\langle\sigma_S\rangle\Big] = \mathbb{E}_{\underline{\sigma}^0}\mathbb{E}_{\underline{\tilde{J}}|\underline{\sigma}^0}\Big[\Big\langle\prod_{S\in\mathcal{C}}\sigma_S\Big\rangle\prod_{S\in\mathcal{C}}\langle\sigma_S\rangle\Big]. \tag{7}$$

This is the "tower property" of conditional expectations.

2) Nishimori identities for symmetric channels: For symmetric channels this identity can be further specialized and yields the so-called Nishimori identities, as we now show. This is specially important for us since the BEC is a symmetric channel. By definition, symmetric channels are those satisfying $Q(J_A|\sigma_A^0) = Q(-J_A|-\sigma_A^0)$ or equivalently $\mathsf{c}(\tilde{J}_A|\sigma_A^0) = \mathsf{c}(-\tilde{J}_A|-\sigma_A^0)$.

Given $\underline{\sigma}^0$ the Gibbs distribution (1) is (independently of channel symmetry) invariant under the gauge transformation $\sigma_i \to \sigma_i^0\sigma_i$, $\tilde{J}_A \to \sigma_A^0\tilde{J}_A$. Let us denote by $\underline{\sigma}^0 \star \underline{\tilde{J}}$ the "component-wise" product $(\sigma_A^0\tilde{J}_A)_{A=1}^m$. Now we perform a gauge transformation on both sides of (7). For the left hand side we have

$$\mathbb{E}_{\underline{\tilde{J}}|\underline{\sigma}^0}\Big[\prod_{S\in\mathcal{C}}\sigma_S^0\prod_{S\in\mathcal{C}}\langle\sigma_S\rangle\Big] = \mathbb{E}_{\underline{\sigma}^0\star\underline{\tilde{J}}|\underline{\sigma}^0}\Big[\prod_{S\in\mathcal{C}}\langle\sigma_S\rangle\Big]. \tag{8}$$

Moreover, from $\mathsf{c}(\tilde{J}_A|\sigma_A^0) = \mathsf{c}(-\tilde{J}_A|-\sigma_A^0)$ one can see that for a symmetric channel $\mathsf{c}(\sigma_A^0\tilde{J}_A|\sigma_A^0) = \mathsf{c}(\tilde{J}_A|1)$ and therefore in (8) we can replace $\mathbb{E}_{\underline{\sigma}^0\star\underline{\tilde{J}}|\underline{\sigma}^0}$ by $\mathbb{E}_{\underline{\tilde{J}}|\underline{1}}$ and we get

$$\mathbb{E}_{\underline{\tilde{J}}|\underline{\sigma}^0}\Big[\prod_{S\in\mathcal{C}}\sigma_S^0\prod_{S\in\mathcal{C}}\langle\sigma_S\rangle\Big] = \mathbb{E}_{\underline{\tilde{J}}|\underline{1}}\Big[\prod_{S\in\mathcal{C}}\langle\sigma_S\rangle\Big]. \tag{9}$$

[2]Equation (4) corresponds to one of the two *density evolution* fixed point equations associated with the belief propagation algorithm, see [2] for the links between this algorithm and the replica symmetric functional.

The same steps show that the right hand side of (7) also satisfies

$$\mathbb{E}_{\underline{\tilde{J}}|\underline{\sigma}^0}\Big[\langle \prod_{S\in\mathcal{C}} \sigma_S\rangle \prod_{S\in\mathcal{C}} \langle\sigma_S\rangle\Big] = \mathbb{E}_{\underline{\tilde{J}}|\underline{1}}\Big[\langle \prod_{S\in\mathcal{C}} \sigma_S\rangle \prod_{S\in\mathcal{C}} \langle\sigma_S\rangle\Big]. \tag{10}$$

From (9), (10), (7) we get the Nishimori identity

$$\mathbb{E}_{\underline{\tilde{J}}|\underline{1}}\Big[\prod_{S\in\mathcal{C}} \langle\sigma_S\rangle\Big] = \mathbb{E}_{\underline{\tilde{J}}|\underline{1}}\Big[\langle \prod_{S\in\mathcal{C}} \sigma_S\rangle \prod_{S\in\mathcal{C}} \langle\sigma_S\rangle\Big]. \tag{11}$$

An important role is played by the space of distributions of half-log-likehood of symmetric "channels" $\mathcal{X}$. More precisely, a symmetric "channel" is any distribution satisfying $q(J|\sigma^0) = q(-J|-\sigma^0)$, $\sigma^0 \in \{-1,+1\}$, $J\in\mathbb{R}$. The associated half-log-likelihood variable is $h = \frac{1}{2}\ln\frac{q(J|+1)}{q(J|-1)}$. The space $\mathcal{X}$ is the space of distributions formally defined by $\mathsf{x}(dh) = q(J|+1)dJ$. It is easy to deduce from $q(J|\sigma^0) = q(-J|-\sigma^0)$ that a distribution $\mathsf{x}\in\mathcal{X}$ satisfies $\mathsf{x}(-dh) = e^{-2h}\mathsf{x}(dh)$.

There is an important special case of the Nishimori identity (11). Namely the one satisfied by the system constituted by a single hidden variable σ^0, observed through a noisy symmetric "channel". The Gibbs distribution is simply $e^{h\sigma}/(2\cosh h)$ where h is the half-log-likelihood of the "channel". Since $\langle\sigma\rangle = \tanh h$, an application of (11) (where the singleton set is taken k times) yields

$$\int (\tanh h)^{2k-1}\,\mathsf{x}(dh) = \int (\tanh h)^{2k}\,\mathsf{x}(dh), \qquad k\in\mathbb{N}^*. \tag{12}$$

In Appendix A we show in a self-contained way that any $\mathsf{x}\in\mathcal{X}$ satisfies (12).

3) Conditional entropy for symmetric channels: Since the Gibbs distribution is invariant under a gauge transformation, the partition function Z also is, and therefore, for a given graph $\mathcal{G}$ and hidden vector $\underline{\sigma}^0$, we have

$$\mathbb{E}_{\underline{\tilde{J}}|\underline{\sigma}^0}\ln Z = \mathbb{E}_{\underline{\sigma}^0\star\underline{\tilde{J}}|\underline{\sigma}^0}\ln Z. \tag{13}$$

For symmetric channels the r.h.s. equals $\mathbb{E}_{\underline{\tilde{J}}|\underline{1}}\ln Z$ and thus the average conditional entropy (3) becomes

$$\frac{1}{n}\mathbb{E}_{\mathcal{G}}\mathbb{E}_{\underline{\sigma}^0}\mathbb{E}_{\underline{\tilde{J}}|\underline{\sigma}^0} H(\underline{\sigma}|\underline{\tilde{J}}) = \frac{1}{n}\mathbb{E}_{\mathcal{G}}\mathbb{E}_{\underline{\tilde{J}}|\underline{1}}\ln Z. \tag{14}$$

4) Summary: When one is dealing with symmetric measurement channels, in order to compute the average conditional entropy, or *certain* averages, one may assume that $\sigma_i^0 = 1$, $i = 1,\ldots,n$ and that the quenched variables $\underline{\tilde{J}}$ have distribution $\mathsf{c}(\tilde{J}_A|1)$, $A = 1,\ldots,m$. From now on this is understood unless explicitly specified otherwise.

A word about notation: From now on when we take an expectation over *all* quenched variables arising in an expression we simply write $\mathbb{E}$. We indicate subscripts for partial expectations or when some confusion could arise.

B. Griffiths-Kelly-Sherman inequalities for the BEC

The BEC is a symmetric channel so as shown before, without loss of generality for analysis purposes, we assume $\sigma_i^0 = 1$, $i = 1,\ldots,n$ and that $\underline{\tilde{J}}$ have distribution $\mathsf{c}(\tilde{J}_A|1)$, $A = 1,\ldots,m$. Since $\mathsf{c}(\tilde{J}_A|1) = (1-q)\Delta_\infty + q\Delta_0$ the Gibbs distribution (1) has non-negative coupling constants $\tilde{J}_A$, $A = 1,\ldots,m$. Therefore the Gibbs distribution satisfies the Griffiths-Kelly-Sherman (GKS) inequalities [23, 25]: For any subsets of variable indices $S, T\subset\{1\ldots n\}$ we have

$$\langle\sigma_S\rangle \geq 0, \tag{15}$$

$$\langle\sigma_S\sigma_T\rangle - \langle\sigma_S\rangle\langle\sigma_T\rangle \geq 0. \tag{16}$$

These two inequalities play an important role in the proof of Theorem II.2.

IV. THE ADAPTIVE PATH INTERPOLATION METHOD

For $t = 1,\ldots,T$ let $V_i^{(t)}$ be i.i.d. r.v. distributed according to $\mathsf{x}^{(t)}$. We set $\underline{\mathsf{x}} = (\mathsf{x}^{(1)},\ldots,\mathsf{x}^{(T)})$. Consider the r.v.

$$U^{(t)} = \tanh^{-1}\Big(\tanh\tilde{J}\prod_{i=1}^{K-1} V_i^{(t)}\Big) \tag{17}$$

and independent copies denoted $U_B^{(t)}$ where B is a subscript which runs over $l^{(t)}\sim \text{Poi}\big(\frac{K}{RT}\big)$ of these copies. Later on, we call $\tilde{\mathsf{x}}^{(t)}$ the distribution of $U^{(t)}$ (induced by $\mathsf{x}^{(t)}$ and c). Let also H be a r.v. such that $H = 0$ with probability $1-\epsilon$ and $H = \infty$ with probability ϵ. We define

$$\begin{aligned}\tilde{h}_\epsilon(\underline{\mathsf{x}}) \equiv \mathbb{E}\Big[\ln\Big(&\prod_{t=1}^{T}\prod_{B=1}^{l^{(t)}}(1+\tanh U_B^{(t)}) + e^{-2H}\prod_{t=1}^{T}\prod_{B=1}^{l^{(t)}}(1-\tanh U_B^{(t)})\Big)\\ &- \frac{K-1}{RT}\sum_{t=1}^{T}\ln\Big(1+\tanh\tilde{J}\prod_{i=1}^{K}\tanh V_i^{(t)}\Big) - \frac{1}{R}\ln(1+\tanh\tilde{J})\Big].\end{aligned} \tag{18}$$

One can easily check that if $\mathsf{x}^{(t)} = \mathsf{x}$ for all t then $\tilde{h}_{\epsilon=0}(\underline{\mathsf{x}}) = h_{\rm RS}(\mathsf{x})$. More is true as the following lemma, that we prove in Section VII-E, shows:

Lemma IV.1. *Let $\mathcal{B}^T = \mathcal{B} \times \mathcal{B} \times \ldots \times \mathcal{B}$. For $\epsilon = 0$ we have*

$$\sup_{\underline{\mathsf{x}} \in \mathcal{B}^T} \tilde{h}_{\epsilon=0}(\underline{\mathsf{x}}) = \sup_{\mathsf{x} \in \mathcal{B}} h_{\rm RS}(\mathsf{x}). \tag{19}$$

A. The (t, s)–interpolating model

Consider the construction of an interpolating factor graph ensemble $\mathcal{G}_{t,s}$ for $t \in \{1, 2, \ldots, T\}$ and $s \in [0, 1]$ with the following algorithm, which is the sparse graph counterpart of the interpolating ensemble developed for dense graphs in [13]:

Algorithm 1 Construction of $\mathcal{G}_{t,s}$

for $t' = 1, \ldots, t-1$ **do**
  for $i = 1, \ldots, n$ **do**
    draw a random number $e_i^{(t')} \sim \text{Poi}\left(\frac{K}{RT}\right)$
    for $B = 1, \ldots, e_i^{(t')}$ **do**
      connect variable node i with a half edge with weight $U_{B\to i}^{(t')} \sim \tilde{\mathsf{x}}^{(t')}$
for $t' = t+1, \ldots, T$ **do**
  draw a random number $m^{(t')} \sim \text{Poi}\left(\frac{n}{RT}\right)$
  for $A = 1, \ldots, m^{(t')}$ **do**
    associate factor node (t', A) with weight $\tilde{J}_{t',A} \sim \mathsf{c}$
    connect this factor node (t', A) to K variable nodes uniformly and randomly (and the subset of the variable nodes is denoted by (t', A))
for $i = 1, \ldots, n$ **do**
  draw a random number $e_{i,s}^{(t)} \sim \text{Poi}\left(\frac{Ks}{RT}\right)$
  for $B = 1, \ldots, e_{i,s}^{(t)}$ **do**
    connect variable node i with a half edge with weight $U_{B\to i}^{(t)} \sim \tilde{\mathsf{x}}^{(t)}$
draw a random number $m_s^{(t)} \sim \text{Poi}\left(\frac{n(1-s)}{RT}\right)$
for $A = 1, \ldots, m_s^{(t)}$ **do**
  associate factor node (t, A) with weight $\tilde{J}_{t,A} \sim \mathsf{c}$
  uniformly and randomly connect factor node (t, A) to K variable nodes

The distribution of $\tilde{\mathsf{x}}^{(t)}$ is a function of $\mathsf{x}^{(t)}$ and c according to (17). Therefore $\tilde{\mathsf{x}}^{(t)} \in \mathcal{B}$ when $\mathsf{x}^{(t)}, \mathsf{c} \in \mathcal{B}$. The interpolating graph is designed such that $\mathcal{G}_{t,1}$ is statistically equivalent to $\mathcal{G}_{t+1,0}$; in addition, $\mathcal{G}_{t,s}$ maintains the degree distribution of variable nodes invariant for different t and s. The Hamiltonian associated with $\mathcal{G}_{t,s}$ is

$$\begin{aligned}\mathcal{H}_{t,s}(\underline{\sigma}, \underline{\tilde{J}}, \underline{U}, \underline{m}, \underline{e}) = &- \sum_{t'=1}^{t-1} \sum_{i=1}^{n} \sum_{B=1}^{e_i^{(t')}} U_{B\to i}^{(t')}(\sigma_i - 1) - \sum_{t'=t+1}^{T} \sum_{A=1}^{m^{(t')}} \tilde{J}_{t',A}(\sigma_{t',A} - 1) \\ &- \sum_{i=1}^{n} \sum_{B=1}^{e_{i,s}^{(t)}} U_{B\to i}^{(t)}(\sigma_i - 1) - \sum_{A=1}^{m_s^{(t)}} \tilde{J}_{t,A}(\sigma_{t,A} - 1).\end{aligned} \tag{20}$$

We further consider a generalized version of (20) by adding a perturbation H_i to each variable node, where $H_i = 0$ with probability $1 - \epsilon$ and $H_i = +\infty$ with probability ϵ. Thus our final *interpolating Hamiltonian* is

$$\mathcal{H}_{t,s;\epsilon}(\underline{\sigma}, \underline{\tilde{J}}, \underline{U}, \underline{m}, \underline{e}, \underline{H}) \equiv \mathcal{H}_{t,s}(\underline{\sigma}, \underline{\tilde{J}}, \underline{U}, \underline{m}, \underline{e}) - \sum_{i=1}^{n} H_i(\sigma_i - 1). \tag{21}$$

We also define the associated interpolating partition function, Gibbs expectation and free entropy to be

$$Z_{t,s;\epsilon} \equiv \sum_{\underline{\sigma}} e^{-\mathcal{H}_{t,s;\epsilon}(\underline{\sigma}, \underline{\tilde{J}}, \underline{U}, \underline{m}, \underline{e}, \underline{H})}, \tag{22}$$

$$\langle O \rangle_{t,s;\epsilon} \equiv \frac{1}{Z_{t,s;\epsilon}} \sum_{\underline{\sigma}} O(\underline{\sigma}) e^{-\mathcal{H}_{t,s;\epsilon}(\underline{\sigma}, \underline{\tilde{J}}, \underline{U}, \underline{m}, \underline{e}, \underline{H})}, \tag{23}$$

$$h_{t,s;\epsilon} \equiv \frac{1}{n} \mathbb{E} \ln Z_{t,s;\epsilon}. \tag{24}$$

Recall our notation: the expectation $\mathbb{E}$ here carries over *all* quenched variables entering in the interpolating system. Note that Nishimori's identity (11) and GKS inequalities (15), (16) still apply to the Gibbs expectation $\langle - \rangle_{t,s;\epsilon}$.

The following formulas will play an important role, and the details of their derivation are provided in Appendix D.

$$\frac{d}{d\epsilon} h_{t,s;\epsilon} = -\frac{\ln 2}{n} \sum_{i=1}^{n} (1 - \mathbb{E}\langle \sigma_i \rangle_{t,s;\epsilon;\sim H_i}) = -\frac{\ln 2}{n(1-\epsilon)} \sum_{i=1}^{n} (1 - \mathbb{E}\langle \sigma_i \rangle_{t,s;\epsilon}), \tag{25}$$

$$\frac{d^2}{d\epsilon^2} h_{t,s;\epsilon} = \frac{\ln 2}{n(1-\epsilon)^2} \sum_{i \neq j} \mathbb{E}[\langle \sigma_i \sigma_j \rangle_{t,s;\epsilon} - \langle \sigma_i \rangle_{t,s;\epsilon} \langle \sigma_j \rangle_{t,s;\epsilon}], \tag{26}$$

where $\langle \sigma_i \rangle_{t,s;\epsilon;\sim H_i}$ is the Gibbs expectation with H_i fixed to 0. We immediately note that the first equality of (25) and (15) tell us $|\frac{d}{d\epsilon} h_{t,s;\epsilon}| \leq \ln 2$. Moreover from (16) we see that $h_{t,s;\epsilon}$ is convex.

The connection between the unperturbed and perturbed free energies is given by (see Sec. VII-A for the proof)

Lemma IV.2. *Let* $\mathsf{c} \in \mathcal{B}$ *and* $\underline{\mathsf{x}} \in \mathcal{B}^T$*. For any sequence* $\epsilon_n \to 0_+$ *we have*

$$\lim_{n \to +\infty} |h_{t,s;\epsilon_n} - h_{t,s;\epsilon=0}| = 0. \tag{27}$$

B. Evaluating the free entropy change along the (t,s)*–interpolation*

By interpolating $h_{t,s;\epsilon}$ from $(t=1, s=0)$ to $(t=T, s=1)$, we have

$$h_{1,0;\epsilon} = h_{T,1;\epsilon} + \sum_{t=1}^{T} (h_{t,0;\epsilon} - h_{t,1;\epsilon}) = h_{T,1;\epsilon} - \sum_{t=1}^{T} \int_0^1 ds \frac{dh_{t,s;\epsilon}}{ds}. \tag{28}$$

One can see that

$$\mathcal{H}_{1,0;\epsilon} = -\sum_{t'=1}^{T} \sum_{A=1}^{m^{(t')}} \tilde{J}_{t',A}(\sigma_{t',A} - 1) - \sum_{i=1}^{n} H_i(\sigma_i - 1) \stackrel{?}{=} -\sum_{A=1}^{m} \tilde{J}_A(\sigma_A - 1) - \sum_{i=1}^{n} H_i(\sigma_i - 1)$$

due to the fact that a sum of independent Poisson-distributed r.v. is also Poisson-distributed (with mean equal to the sum of their means). Therefore

$$h_{1,0;\epsilon=0} = \frac{1}{n} \mathbb{E} H(\underline{\sigma} | \underline{\tilde{J}}). \tag{29}$$

On the other hand $h_{T,1;\epsilon}$ corresponds to a part of (18). A subsequent computation (see Section V) on (28) leads to the fundamental sum rule

$$h_{1,0;\epsilon} = \tilde{h}_\epsilon(\underline{\mathsf{x}}) + \frac{1}{RT} \sum_{t=1}^{T} \int_0^1 ds \mathcal{R}_{t,s;\epsilon} \tag{30}$$

where

$$\mathcal{R}_{t,s;\epsilon} = \sum_{p=1}^{\infty} \frac{\mathbb{E}[(\tanh \tilde{J})^{2p}]}{2p(2p-1)} \, \mathbb{E}\big\langle Q_{2p}^K - K q_{2p}^{(t)K-1}(Q_{2p} - q_{2p}^{(t)}) - q_{2p}^{(t)K} \big\rangle_{t,s;\epsilon} \tag{31}$$

with $Q_p \equiv \frac{1}{n} \sum_{i=1}^{n} \sigma_i^{(1)} \cdots \sigma_i^{(p)}$ the *overlap* for p independent *replicas* $\underline{\sigma}^{(1)}, \ldots, \underline{\sigma}^{(p)}$ and $q_p^{(t)} \equiv \mathbb{E}[(\tanh V^{(t)})^p]$. Here the Gibbs average $\langle - \rangle_{t,s;\epsilon}$ over a function of p replicas must be understood as average over the product measure

$$\prod_{\alpha=1}^{p} \frac{1}{Z_{t,s;\epsilon}} e^{-\mathcal{H}_{t,s;\epsilon}(\underline{\sigma}^{(\alpha)}, \underline{\tilde{J}}, \underline{U}, \underline{m}, \underline{e}, \underline{H})}$$

where the quenched variables are the same for all replicas (thus $\mathbb{E}\langle - \rangle_{t,s;\epsilon}$ when applied to a function of different replicas is *not* a product measure). We still denote this Gibbs average by $\langle - \rangle_{t,s;\epsilon}$ for simplicity.

C. Lower bound

In order to show the lower bound we need the following concentration lemma (proven in Sec. VI):

Lemma IV.3 (Concentration of Q_p^K on $\langle Q_p \rangle_{t,s;\epsilon}^K$)**.** *For any* $\mathsf{c} \in \mathcal{B}$*,* $\underline{\mathsf{x}} \in \mathcal{B}^T$*, we have*

$$\lim_{n \to \infty} \int_0^1 d\epsilon \; \mathbb{E}\big\langle \big| Q_p^K - \langle Q_p \rangle_{t,s;\epsilon}^K \big| \big\rangle_{t,s;\epsilon} = 0. \tag{32}$$

Proposition IV.4 (Lower bound)**.** *For* $\mathsf{c} \in \mathcal{B}$ *we have*

$$\liminf_{n\to\infty} \frac{1}{n}\mathbb{E}H(\underline{\sigma}|\underline{\tilde{J}}) \geq \sup_{\mathsf{x}\in\mathcal{B}} h_{\mathrm{RS}}(\mathsf{x}). \tag{33}$$

Proof. Lemma IV.3 implies that there exists a sequence $\epsilon_n \to 0_+, n \to +\infty$ such that[3]

$$\lim_{n\to\infty} \mathbb{E}\big\langle \big|Q_p^K - \langle Q_p\rangle_{t,s;\epsilon_n}^K\big|\big\rangle_{t,s;\epsilon_n} = 0. \tag{34}$$

An elementary proof of this claim is given in Appendix B for completeness. From (30) and (31) this implies

$$\mathcal{R}_{t,s;\epsilon_n} = \sum_{p=1}^{\infty} \frac{\mathbb{E}[(\tanh \tilde{J})^{2p}]}{2p(2p-1)}\, \mathbb{E}\Big[\langle Q_{2p}\rangle_{t,s;\epsilon_n}^K - Kq_{2p}^{(t)K-1}(\langle Q_{2p}\rangle_{t,s;\epsilon_n} - q_{2p}^{(t)}) - q_{2p}^{(t)K}\Big] + o_n(1) \tag{35}$$

where $\lim_{n\to+\infty} o_n(1) = 0$. It is easy to check that $\langle Q_{2p}\rangle_{t,s;\epsilon_n} \geq 0$, so that due to convexity of the function x^K for $x \in \mathbb{R}_+$ we get

$$\langle Q_{2p}\rangle_{t,s;\epsilon_n}^K - q_{2p}^{(t)K} - Kq_{2p}^{(t)K-1}(\langle Q_{2p}\rangle_{t,s;\epsilon_n} - q_{2p}^{(t)}) \geq 0, \qquad \text{and thus} \qquad \mathcal{R}_{t,s;\epsilon_n} \geq o_n(1). \tag{36}$$

Combining (30) and (36) we conclude

$$\liminf_{n\to\infty} h_{1,0;\epsilon_n} \geq \lim_{n\to\infty} \tilde{h}_{\epsilon_n}(\underline{\mathsf{x}}) = \tilde{h}_{\epsilon=0}(\underline{\mathsf{x}}). \tag{37}$$

Recall Lemma IV.2. Since $\liminf$ is the infimum of the set of accumulation points of a sequence, and the only accumulation point of the sequence $h_{1,0;\epsilon_n} - h_{1,0;\epsilon=0}$ is 0, we have

$$\liminf_{n\to\infty} h_{1,0;\epsilon_n} = \liminf_{n\to\infty} (h_{1,0;\epsilon_n} - h_{1,0;\epsilon=0} + h_{1,0;\epsilon=0}) = \liminf_{n\to\infty} h_{1,0;\epsilon=0}. \tag{38}$$

Thus (37) becomes

$$\liminf_{n\to\infty} h_{1,0;\epsilon=0} \geq \tilde{h}_{\epsilon=0}(\underline{\mathsf{x}}). \tag{39}$$

Finally one can take the supremum of the right hand side and use (19) as well as (29) to obtain (33). □

D. Upper bound

In this paragraph we crucially use the specificities of the BEC. In particular we use the following lemma (proven in Sec. VII-B):

Lemma IV.5. *For any* $\mathsf{c} \in \mathcal{B}$*,* $\underline{\mathsf{x}} \in \mathcal{B}^T$*, and any* $A \subseteq \{1,\dots,n\}$ *we have* $\langle \sigma_A\rangle_{t,s;\epsilon} \in \{0,1\}$.

Notice that $\langle Q_p\rangle_{t,s;\epsilon} = \frac{1}{n}\sum_{i=1}^n \langle\sigma_i\rangle_{t,s;\epsilon}^p$. Lemma IV.5 then implies $\langle Q_p\rangle_{t,s;\epsilon} = \langle Q_1\rangle_{t,s;\epsilon}$ for all $p \in \mathbb{N}^*$. We also have $V^{(t)}$ equals either 0 or 1 because $\mathsf{x} \in \mathcal{B}$ thus $q_p^{(t)} = q_1^{(t)}$ for $p \in \mathbb{N}^*$. Finally recall that $\mathsf{c}(\tilde{J}|1) = (1-q)\Delta_\infty + q\Delta_0$. These facts reduce (35) to

$$\mathcal{R}_{t,s;\epsilon_n} = (1-q)\ln(2)\mathbb{E}\Big[\langle Q_1\rangle_{t,s;\epsilon_n}^K - Kq_1^{(t)K-1}(\langle Q_1\rangle_{t,s;\epsilon_n} - q_1^{(t)}) - q_1^{(t)K}\Big] + o_n(1). \tag{40}$$

To further simplify this expression we need the following lemmas (Lemma IV.6 is proven in Sec. VI, Lemma IV.7 in Sec. VII-C).

Lemma IV.6 (Concentration of $\langle Q_1\rangle_{t,s;\epsilon}^K$ on $\mathbb{E}[\langle Q_1\rangle_{t,s;\epsilon}]^K$)**.** *For any* $\mathsf{c} \in \mathcal{B}$ *and* $\underline{\mathsf{x}} \in \mathcal{B}^T$ *we have*

$$\lim_{n\to\infty} \mathbb{E}\big[\big|\langle Q_1\rangle_{t,s;\epsilon}^K - \mathbb{E}[\langle Q_1\rangle_{t,s;\epsilon}]^K\big|\big] = 0 \tag{41}$$

uniformly in $0 < \epsilon < 1$*. In particular this is also true with* ϵ *replaced by the sequence* $\epsilon_n \to 0_+$.

Lemma IV.7 (Weak s-dependence at fixed t)**.** *One can find a numerical constant* $C > 0$ *such that for any* $s \in [0,1]$

$$\big|\mathbb{E}\langle Q_1\rangle_{t,0;\epsilon_n} - \mathbb{E}\langle Q_1\rangle_{t,s;\epsilon_n}\big| \leq \frac{Cn}{T}. \tag{42}$$

Using (41) the expression (40) becomes

$$\mathcal{R}_{t,s;\epsilon_n} = (1-q)\ln(2)\Big(\mathbb{E}[\langle Q_1\rangle_{t,s;\epsilon_n}]^K - Kq_1^{(t)K-1}(\mathbb{E}\langle Q_1\rangle_{t,s;\epsilon_n} - q_1^{(t)}) - q_1^{(t)K}\Big) + o_n(1). \tag{43}$$

Furthermore note that since T is a free parameter (controlling the mean of $e_{i,s}^{(t)}$ and $m_s^{(t)}$) we can set it significantly larger than n when using Lemma IV.7. Therefore, we can remove the s-dependence of $\mathcal{R}_{t,s;\epsilon_n}$ and write it as

$$\mathcal{R}_{t,s;\epsilon_n} = (1-q)\ln(2)\Big(\mathbb{E}[\langle Q_1\rangle_{t,0;\epsilon_n}]^K - Kq_1^{(t)K-1}(\mathbb{E}\langle Q_1\rangle_{t,0;\epsilon_n} - q_1^{(t)}) - q_1^{(t)K}\Big) + o_n(1). \tag{44}$$

[3]For even K we do not need Lemma IV.3 and the proof just follows from convexity of x^K for $x \in \mathbb{R}$ (see e.g. [26]).

The use of Lemma IV.7 is critical because $\mathbb{E}\langle Q_1\rangle_{t,0;\epsilon_n}$ is independent of $\{\mathsf{x}^{(t')}\}_{t'\geq t}$. Also recall $q_1^{(t)} \equiv \mathbb{E}\tanh V^{(t)}$. This allows us to sequentially choose a distribution $\hat{\mathsf{x}}_n^{(t)}$ for $V^{(t)}$ along our interpolation from $t=1$ to T such that the following equation is satisfied:

$$q_1^{(t)} = \mathbb{E}\langle Q_1\rangle_{t,0;\epsilon_n}. \tag{45}$$

In other words the interpolation path is adapted so that (45) holds in order to cancel the remainder (44). It thus allows to choose the "optimal interpolation path". This equation indeed posseses a (unique) solution, see Sec. VII-D for the proof:

Lemma IV.8 (Existence of the "optimal interpolation path")**.** *Equation* (45) *has a unique solution* $\underline{\hat{\mathsf{x}}}_n = \{\hat{\mathsf{x}}_n^{(t)}\}_{t=1,\dots,T} \in \mathcal{B}^T$.

We are now ready to prove the upper bound.

Proposition IV.9 (Upper bound)**.** *For any* $\mathsf{c} \in \mathcal{B}$ *we have*

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}H(\underline{\sigma}|\underline{\tilde{J}}) \leq \sup_{\mathsf{x}\in\mathcal{B}} h_{\rm RS}(\mathsf{x}). \tag{46}$$

Proof. The specific choice $\underline{\hat{\mathsf{x}}}_n$ (essentially) eliminates $\mathcal{R}_{t,s;\epsilon_n}$ and reduces (30) to

$$h_{1,0;\epsilon_n} = \tilde{h}_{\epsilon_n}(\underline{\hat{\mathsf{x}}}_n) + o_n(1) \leq \sup_{\underline{\mathsf{x}}\in\mathcal{B}^T} \tilde{h}_{\epsilon_n}(\underline{\mathsf{x}}) + o_n(1). \tag{47}$$

Directly from (18) we see that $\tilde{h}_{\epsilon_n}(\underline{\mathsf{x}})$ is of the form $\epsilon_n g(\underline{\mathsf{x}}) + \tilde{h}_{\epsilon=0}(\underline{\mathsf{x}})$. Thus

$$\sup_{\underline{\mathsf{x}}\in\mathcal{B}^T} \tilde{h}_{\epsilon_n}(\underline{\mathsf{x}}) \leq \epsilon_n \sup_{\underline{\mathsf{x}}\in\mathcal{B}^T} g(\underline{\mathsf{x}}) + \sup_{\underline{\mathsf{x}}\in\mathcal{B}^T} \tilde{h}_{\epsilon=0}(\underline{\mathsf{x}}) \tag{48}$$

and passing to the limit (and recalling that $\epsilon_n \to 0_+$)

$$\limsup_{n\to\infty} h_{1,0;\epsilon_n} \leq \limsup_{n\to\infty} \sup_{\underline{\mathsf{x}}\in\mathcal{B}^T} \tilde{h}_{\epsilon_n}(\underline{\mathsf{x}}) \leq \sup_{\underline{\mathsf{x}}\in\mathcal{B}^T} \tilde{h}_{\epsilon=0}(\underline{\mathsf{x}}). \tag{49}$$

Because of Lemma IV.2 we have $\limsup_{n\to\infty} h_{1,0;\epsilon_n} = \limsup_{n\to\infty} h_{1,0;\epsilon=0} = \limsup_{n\to\infty} \frac{1}{n}\mathbb{E}H(\underline{\sigma}|\underline{\tilde{J}})$ using (29). This remark, together with equation (19), allows to conclude the proof. □

V. Proof of the fundamental sum rule (30)-(31)

Similar computations go back to [15] and were applied in Nishimori symmetric situations in [21, 26, 27], so we will be relatively brief. We compute $h_{T,1;\epsilon}$ and $\frac{dh_{t,s;\epsilon}}{ds}$ in (28). From the definitions (21), (22), (24) and using the identity $e^{\sigma x} = \cosh x(1+\sigma\tanh x)$ for $\sigma \in \{-1,+1\}$, we can expand $h_{T,1;\epsilon}$ as

$$h_{T,1;\epsilon} = \mathbb{E}\Big[\ln\Big(\prod_{t'=1}^{T}\prod_{B=1}^{e_i^{(t')}}(1+\tanh U_B^{(t')}) + e^{-2H}\prod_{t'=1}^{T}\prod_{B=1}^{e_i^{(t')}}(1-\tanh U_B^{(t')})\Big) - \frac{K}{RT}\sum_{t'=1}^{T}\ln(1+\tanh U^{(t')})\Big]. \tag{50}$$

Note that the first two terms are a part of (18). For $\frac{dh_{t,s;\epsilon}}{ds}$ we use a property of the Poisson distribution, namely that any function $f(X)$ of a r.v. X with Poisson distribution and mean ν satisfies the identity

$$\frac{d\,\mathbb{E}f(X)}{d\nu} = \mathbb{E}f(X+1) - \mathbb{E}f(X). \tag{51}$$

This allows us to write

$$\frac{dh_{t,s;\epsilon}}{ds} = -\frac{1}{RT}\mathbb{E}_{t,s;\epsilon}\mathbb{E}_{B,\tilde{J}_B}\ln\langle e^{\tilde{J}_B(\sigma_B-1)}\rangle_{t,s;\epsilon} + \frac{K}{nRT}\sum_{i=1}^{n}\mathbb{E}_{t,s;\epsilon}\mathbb{E}_{U_i^{(t)}}\ln\langle e^{U_i^{(t)}(\sigma_i-1)}\rangle_{t,s;\epsilon}, \tag{52}$$

where we distinguish the expectation $\mathbb{E}_{t,s;\epsilon}$ with respect to the original interpolating model with Hamiltonian (21) and the expectation with respect to an "extra measurement" and its neighborood $\mathbb{E}_{B,\tilde{J}_B}$ and an "extra field" $\mathbb{E}_{U_i^{(t)}}$. Standard algebra, using again the identity $e^{\sigma x} = \cosh x(1+\sigma\tanh x)$, leads to

$$\begin{aligned}\mathbb{E}\ln\langle e^{\tilde{J}_B(\sigma_B-1)}\rangle_{t,s;\epsilon} &= \mathbb{E}_{t,s;\epsilon}\mathbb{E}_{B,\tilde{J}_B}\ln\big(1+\langle\sigma_B\rangle_{t,s;\epsilon}\tanh\tilde{J}_B\big) - \mathbb{E}_{\tilde{J}_B}\ln(1+\tanh\tilde{J}_B)\\ &= \sum_{p=1}^{\infty}\frac{(-1)^{p+1}}{p}\frac{1}{n^K}\sum_{i_1,\dots,i_K}\mathbb{E}[\langle\sigma_{i_1}\cdots\sigma_{i_K}\rangle_{t,s;\epsilon}^p]\mathbb{E}[(\tanh\tilde{J})^p] - \mathbb{E}\ln(1+\tanh\tilde{J})\end{aligned}$$

and similarly, using (17),

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\ln\langle e^{U_i^{(t)}(\sigma_i-1)}\rangle_{t,s;\epsilon}\\
=\ &\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{t,s;\epsilon}\mathbb{E}_{U_i^{(t)}}\ln\big(1+\langle\sigma_i\rangle_{t,s;\epsilon}\tanh U_i^{(t)}\big)-\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\ln(1+\tanh U_i^{(t)})\\
=\ &\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{t,s;\epsilon}\mathbb{E}_{\tilde J,\underline{V}^{(t)}}\ln\big(1+\langle\sigma_i\rangle_{t,s;\epsilon}\tanh\tilde J\prod_{j=1}^{K-1}\tanh V_j^{(t)}\big)-\mathbb{E}\ln(1+\tanh U^{(t)})\\
=\ &\sum_{p=1}^{\infty}\frac{(-1)^{p+1}}{p}\mathbb{E}[(\tanh\tilde J)^p]\mathbb{E}[(\tanh V^{(t)})^p]^{K-1}\frac{1}{n}\sum_i\mathbb{E}[\langle\sigma_i\rangle_{t,s;\epsilon}^p]-\mathbb{E}\ln(1+\tanh U^{(t)}).
\end{aligned}$$

Recall $Q_p\equiv\frac{1}{n}\sum_{i=1}^n\sigma_i^{(1)}\cdots\sigma_i^{(p)}$ and thus $\langle Q_p\rangle_{t,s;\epsilon}=\frac{1}{n}\sum_i\langle\sigma_i\rangle_{t,s;\epsilon}^p$, and $\langle Q_p^K\rangle_{t,s;\epsilon}=\frac{1}{n^K}\sum_{i_1,\dots,i_K}\langle\sigma_{i_1}\cdots\sigma_{i_K}\rangle_{t,s;\epsilon}^p$. Recall also $q_p^{(t)}\equiv\mathbb{E}[(\tanh V^{(t)})^p]$. Then (52) becomes

$$\begin{aligned}
\frac{dh_{t,s;\epsilon}}{ds}&=-\frac{1}{RT}\sum_{p=1}^{\infty}\frac{(-1)^{p+1}}{p}\mathbb{E}[(\tanh\tilde J)^p]\mathbb{E}[\langle Q_p^K\rangle_{t,s;\epsilon}-Kq_p^{(t)K-1}\langle Q_p\rangle_{t,s;\epsilon}]\\
&+\frac{1}{RT}\mathbb{E}\ln(1+\tanh\tilde J)-\frac{K}{RT}\mathbb{E}\ln(1+\tanh U^{(t)})\\
&=-\frac{1}{RT}\sum_{p=1}^{\infty}\frac{(-1)^{p+1}}{p}\mathbb{E}[(\tanh\tilde J)^p]\mathbb{E}\big\langle Q_p^K-Kq_p^{(t)K-1}(Q_p-q_p^{(t)})-q_p^{(t)K}\big\rangle_{t,s;\epsilon}\\
&+\frac{K-1}{RT}\mathbb{E}\ln\big(1+\tanh\tilde J\prod_{j=1}^{K}\tanh V_j^{(t)}\big)+\frac{1}{RT}\mathbb{E}\ln(1+\tanh\tilde J)-\frac{K}{RT}\mathbb{E}\ln(1+\tanh U^{(t)}).
\end{aligned}\tag{53}$$

Substituting (50) and (53) into (28) gives (30), where

$$\mathcal{R}_{t,s;\epsilon}=\sum_{p=1}^{\infty}\frac{(-1)^{p+1}}{p}\mathbb{E}[(\tanh\tilde J)^p]\mathbb{E}\big\langle Q_p^K-Kq_p^{(t)K-1}(Q_p-q_p^{(t)})-q_p^{(t)K}\big\rangle_{t,s;\epsilon}.\tag{54}$$

An application of (11) yields $\mathbb{E}\langle Q_{2p-1}^m\rangle_{t,s;\epsilon}=\mathbb{E}\langle Q_{2p}^m\rangle_{t,s;\epsilon}$ for $p\ge1$ and any integer m. Similarly an application of (12) yields $q_{2p-1}^{(t)}=q_{2p}^{(t)}$ as well as $\mathbb{E}[(\tanh\tilde J)^{2p-1}]=\mathbb{E}[(\tanh\tilde J)^{2p}]$ for $p\ge1$. Therefore combining the odd and even terms of (54) we obtain the form in (31).

VI. Concentration of overlaps: Proof of Lemmas IV.3 and IV.6

We will prove the following two lemmas which imply Lemmas IV.3 and IV.6.

Lemma VI.1 (Concentration of Q_p on $\langle Q_p\rangle_{t,s;\epsilon}$)**.** *For any* $\mathsf{c}\in\mathcal{B}$ *and* $\underline{\mathsf{x}}\in\mathcal{B}^T$

$$\int_0^1 d\epsilon\ \mathbb{E}\big\langle(Q_p-\langle Q_p\rangle_{t,s;\epsilon})^2\big\rangle_{t,s;\epsilon}=\mathcal{O}\Big(\frac{p}{n}\Big)\tag{55}$$

uniformly in t,s.

Lemma VI.2 (Concentration of $\langle Q_1\rangle_{t,s;\epsilon}$ on $\mathbb{E}\langle Q_1\rangle_{t,s;\epsilon}$)**.** *For* $\mathsf{c}\in\mathcal{B}$, $\underline{\mathsf{x}}\in\mathcal{B}^T$ *and* $0<\eta<1$ *we have*

$$\mathbb{E}\big[(\langle Q_1\rangle_{t,s;\epsilon}-\mathbb{E}\langle Q_1\rangle_{t,s;\epsilon})^2\big]=\mathcal{O}\left(n^{-1+\eta}\right)\tag{56}$$

uniformly in t,s,ϵ.

Let us immediately show that Lemmas IV.3 and IV.6 follow. We first note

$$\begin{aligned}
\mathbb{E}\big\langle|Q_p^K-\langle Q_p\rangle_{t,s;\epsilon}^K|\big\rangle_{t,s;\epsilon}&=\mathbb{E}\Big\langle\Big|(Q_p-\langle Q_p\rangle_{t,s;\epsilon})\sum_{k=0}^{K-1}Q_p^{K-k-1}\langle Q_p\rangle_{t,s;\epsilon}^k\Big|\Big\rangle_{t,s;\epsilon}\\
&\le K\mathbb{E}\big\langle|Q_p-\langle Q_p\rangle_{t,s;\epsilon}|\big\rangle_{t,s;\epsilon}.
\end{aligned}\tag{57}$$

We can apply the Cauchy-Schwarz inequality to get

$$\int_0^1 d\epsilon\ \mathbb{E}\big\langle|Q_p^K-\langle Q_p\rangle_{t,s;\epsilon}^K|\big\rangle_{t,s;\epsilon}\le K\Big\{\int_0^1 d\epsilon\ \mathbb{E}\big\langle(Q_p-\langle Q_p\rangle_{t,s;\epsilon})^2\big\rangle_{t,s;\epsilon}\Big\}^{1/2}\tag{58}$$

and thanks to (55) we obtain

$$\int_0^1 d\epsilon\ \mathbb{E}\big\langle \big|Q_p^K - \langle Q_p\rangle_{t,s;\epsilon}^K\big|\big\rangle_{t,s;\epsilon} = \mathcal{O}\left(K\left(\frac{p}{n}\right)^{1/2}\right).$$

This gives Lemma IV.3. Similarly to (57) and (58), it is easy to show

$$\mathbb{E}\big[\big|\langle Q_1\rangle_{t,s;\epsilon}^K - \mathbb{E}[\langle Q_1\rangle_{t,s;\epsilon}]^K\big|\big] \le K\,\mathbb{E}\big[(\langle Q_1\rangle_{t,s;\epsilon} - \mathbb{E}\langle Q_1\rangle_{t,s;\epsilon})^2\big]^{1/2} \tag{59}$$

which with (56) implies

$$\mathbb{E}\left[\big|\langle Q_1\rangle_{t,s;\epsilon}^K - \mathbb{E}[\langle Q_1\rangle_{t,s;\epsilon}]^K\big|\right] = \mathcal{O}(Kn^{-\frac{1}{2}+\frac{\eta}{2}})$$

and proves Lemma IV.6.

A. Proof of Lemma VI.1

From the definition of Q_p we have

$$\begin{aligned}\mathbb{E}\big\langle (Q_p - \langle Q_p\rangle_{t,s;\epsilon})^2\big\rangle_{t,s;\epsilon} &= \frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}\big[\langle \sigma_i\sigma_j\rangle_{t,s;\epsilon}^p - \langle\sigma_i\rangle_{t,s;\epsilon}^p\langle\sigma_j\rangle_{t,s;\epsilon}^p\big] \\ &= \frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}\Big[(\langle\sigma_i\sigma_j\rangle_{t,s;\epsilon} - \langle\sigma_i\rangle_{t,s;\epsilon}\langle\sigma_j\rangle_{t,s;\epsilon})\sum_{l=0}^{p-1}\langle\sigma_i\sigma_j\rangle_{t,s;\epsilon}^{p-1-l}\langle\sigma_i\rangle_{t,s;\epsilon}^l\langle\sigma_j\rangle_{t,s;\epsilon}^l\Big].\end{aligned} \tag{60}$$

By (16) we have $0 \le \langle\sigma_i\sigma_j\rangle_{t,s;\epsilon} - \langle\sigma_i\rangle_{t,s;\epsilon}\langle\sigma_j\rangle_{t,s;\epsilon}$. This allows us to upper bound (60) as

$$\begin{aligned}\mathbb{E}\big\langle (Q_p - \langle Q_p\rangle_{t,s;\epsilon})^2\big\rangle_{t,s;\epsilon} &\le \frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}\Big[(\langle\sigma_i\sigma_j\rangle_{t,s;\epsilon} - \langle\sigma_i\rangle_{t,s;\epsilon}\langle\sigma_j\rangle_{t,s;\epsilon})\sum_{l=0}^{p-1}\big|\langle\sigma_i\sigma_j\rangle_{t,s;\epsilon}^{p-1-l}\langle\sigma_i\rangle_{t,s;\epsilon}^l\langle\sigma_j\rangle_{t,s;\epsilon}^l\big|\Big] \\ &\le \frac{p}{n^2}\sum_{i,j=1}^n \mathbb{E}\left[\langle\sigma_i\sigma_j\rangle_{t,s;\epsilon} - \langle\sigma_i\rangle_{t,s;\epsilon}\langle\sigma_j\rangle_{t,s;\epsilon}\right].\end{aligned} \tag{61}$$

Hence integrating (61) over $\epsilon \in [0,1]$ and recalling the formula (26), we obtain

$$\begin{aligned}\int_0^1 d\epsilon\ \mathbb{E}\big\langle (Q_p - \langle Q_p\rangle_{t,s;\epsilon})^2\big\rangle_{t,s;\epsilon} &\le p\int_0^1 d\epsilon\ \frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}\left[\langle\sigma_i\sigma_j\rangle_{t,s;\epsilon} - \langle\sigma_i\rangle_{t,s;\epsilon}\langle\sigma_j\rangle_{t,s;\epsilon}\right] \\ &\le \frac{p}{n} + \frac{p}{n\ln 2}\int_0^1 d\epsilon \frac{d^2}{d\epsilon^2}h_{t,s;\epsilon} \\ &= \frac{p}{n} + \frac{p}{n\ln 2}\Big[\frac{d}{d\epsilon}h_{t,s;\epsilon}\Big]_{\epsilon=0}^{\epsilon=1} \\ &\le \frac{3p}{n},\end{aligned}$$

using the first equality of (25) for the last inequality.

B. Proof of Lemma VI.2

Consider a slightly more perturbed Hamiltonian $\mathcal{H}_{t,s;\epsilon+\delta-\epsilon\delta}$ where $H_i = 0$ with probability $(1-\epsilon)(1-\delta)$ and $H_i = +\infty$ with probability $1-(1-\epsilon)(1-\delta) = \epsilon+\delta-\epsilon\delta$. It is equivalent to express $\mathcal{H}_{t,s;\epsilon+\delta-\epsilon\delta}$ as

$$\mathcal{H}_{t,s;\epsilon,\delta}(\underline{\sigma},\underline{\tilde{J}},\underline{U},\underline{m},\underline{e},\underline{H},\underline{\tilde{H}}) \equiv \mathcal{H}_{t,s;\epsilon}(\underline{\sigma},\underline{\tilde{J}},\underline{U},\underline{m},\underline{e},\underline{H}) - \sum_{i=1}^n \tilde{H}_i(\sigma_i - 1) \tag{62}$$

where $\tilde{H}_i = 0$ with probability $1-\delta$ and $\tilde{H}_i = +\infty$ with probability δ. Let $Z_{t,s;\epsilon,\delta}$ be the partition function associated with $\mathcal{H}_{t,s;\epsilon,\delta}$. Furthermore, let

$$\tilde{F}_{t,s;\epsilon}(\delta) \equiv -\frac{1}{n}\mathbb{E}_{\underline{\tilde{H}}}\ln Z_{t,s;\epsilon,\delta}, \tag{63}$$

$$\tilde{f}_{t,s;\epsilon}(\delta) \equiv -\frac{1}{n}\mathbb{E}_{t,s;\epsilon}\mathbb{E}_{\underline{\tilde{H}}}\ln Z_{t,s;\epsilon,\delta} = -\frac{1}{n}\mathbb{E}\ln Z_{t,s;\epsilon,\delta}. \tag{64}$$

The Hamiltonian $\mathcal{H}_{t,s;\epsilon,\delta}$ is not involved in the adaptive interpolation scheme, but is used only in this proof to provide an upper bound on the concentration of $\langle Q_1\rangle_{t,s;\epsilon}$. We work on the expression $\mathcal{H}_{t,s;\epsilon,\delta}$ rather than $\mathcal{H}_{t,s;\epsilon+\delta-\epsilon\delta}$ so that the expectation over $\underline{\tilde{H}}$ in the definition of $\tilde{F}_{t,s;\epsilon}(\delta)$ is well defined.

As one can verify

$$\ln \frac{Z_{t,s;\epsilon,\delta}}{Z_{t,s;\epsilon}} = \ln\langle e^{\sum_{i=1}^n \tilde{H}_i(\sigma_i - 1)}\rangle_{t,s;\epsilon},$$

$$e^{\tilde{H}_i\sigma_i} = e^{\tilde{H}_i}\frac{1+\sigma_i \tanh \tilde{H}_i}{1+\tanh \tilde{H}_i}.$$

Thus

$$\tilde{F}_{t,s;\epsilon}(\delta) = \tilde{F}_{t,s;\epsilon}(0) - \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\tilde{H}_i} \ln \frac{1+\langle\sigma_i\rangle_{t,s;\epsilon}\tanh \tilde{H}_i}{1+\tanh \tilde{H}_i}.$$

Note that $\tanh \tilde{H}_i$ and $\langle\sigma_i\rangle_{t,s;\epsilon}$ equal either 0 or 1, therefore we can further write

$$\begin{aligned}\tilde{F}_{t,s;\epsilon}(\delta) &= \tilde{F}_{t,s;\epsilon}(0) - \frac{1}{n}\ln 2\sum_{i=1}^n \mathbb{E}_{\tilde{H}_i}\big[\langle\sigma_i\rangle_{t,s;\epsilon}\tanh \tilde{H}_i - \tanh \tilde{H}_i\big] \\ &= \tilde{F}_{t,s;\epsilon}(0) - \frac{1}{n}\delta\ln 2\sum_{i=1}^n \langle\sigma_i\rangle_{t,s;\epsilon} + \frac{1}{n}\delta \ln 2. \qquad (65)\end{aligned}$$

A similar computation applied to $\tilde{f}_{t,s;\epsilon}(\delta)$ gives

$$\tilde{f}_{t,s;\epsilon}(\delta) = \tilde{f}_{t,s;\epsilon}(0) - \frac{1}{n}\delta\ln 2\sum_{i=1}^n \mathbb{E}\langle\sigma_i\rangle_{t,s;\epsilon} + \frac{1}{n}\delta \ln 2. \qquad (66)$$

Now, subtracting (65) from (66) yields

$$\begin{aligned}(\tilde{f}_{t,s;\epsilon}(\delta) - \tilde{F}_{t,s;\epsilon}(\delta)) - (\tilde{f}_{t,s;\epsilon}(0) - \tilde{F}_{t,s;\epsilon}(0)) &= \frac{1}{n}\delta\ln 2\sum_{i=1}^n (\langle\sigma_i\rangle_{t,s;\epsilon} - \mathbb{E}\langle\sigma_i\rangle_{t,s;\epsilon}) \\ &= \delta\ln 2(\langle Q_1\rangle_{t,s;\epsilon} - \mathbb{E}\langle Q_1\rangle_{t,s;\epsilon}). \qquad (67)\end{aligned}$$

Squaring and taking expectations one deduces from (67)

$$\mathbb{E}\left[(\langle Q_1\rangle_{t,s;\epsilon} - \mathbb{E}\langle Q_1\rangle_{t,s;\epsilon})^2\right] \le \frac{2}{(\delta\ln 2)^2}\big\{\mathbb{E}\big[(\tilde{F}_{t,s;\epsilon}(\delta) - \tilde{f}_{t,s;\epsilon}(\delta))^2\big] + \mathbb{E}\big[(\tilde{F}_{t,s;\epsilon}(0) - \tilde{f}_{t,s;\epsilon}(0))^2\big]\big\}. \qquad (68)$$

In Appendix E we prove a concentration property for the free energy, namely:

Lemma VI.3 (Free energy concentration)**.** *For any* $s, \epsilon, \delta \in [0,1]$, $t = 1, \dots, T$ *we have*

$$\mathbb{E}\big[(\tilde{F}_{t,s;\epsilon}(\delta) - \tilde{f}_{t,s;\epsilon}(\delta))^2\big] = \mathcal{O}(n^{-1}). \qquad (69)$$

Substituting (69) into (68) and choosing $\delta = \Theta(n^{-\eta/2})$ where $0 < \eta < 1$, we finally get

$$\mathbb{E}\big[(\langle Q_1\rangle_{t,s;\epsilon} - \mathbb{E}\langle Q_1\rangle_{t,s;\epsilon})^2\big] = \mathcal{O}(n^{-1+\eta}). \qquad (70)$$

This ends the proof of Lemma VI.2.

VII. Proofs of technical lemmas

A. *Proof of Lemma IV.2*

We remarked below equation (25) that $|\frac{d}{d\epsilon}h_{t,s;\epsilon}| \le \ln 2$ (for any symmetric channel). Thus by the mean value theorem

$$|h_{t,s;\epsilon_n} - h_{t,s;\epsilon=0}| \le \epsilon_n \ln 2 \qquad (71)$$

so that Lemma IV.2 is true.

B. *Proof of Lemma IV.5*

The first GKS inequality (15) implies

$$\langle\sigma_A\rangle_{t,s;\epsilon}(1 - \langle\sigma_A\rangle_{t,s;\epsilon}) \ge 0. \qquad (72)$$

Moreover, Nishimori's identity (11) implies

$$\mathbb{E}\langle\sigma_A\rangle_{t,s;\epsilon} = \mathbb{E}[\langle\sigma_A\rangle^2_{t,s;\epsilon}] \qquad (73)$$

which can be written as

$$\mathbb{E}[\langle\sigma_A\rangle_{t,s;\epsilon}(1 - \langle\sigma_A\rangle_{t,s;\epsilon})] = 0. \qquad (74)$$

As a result of (72) and (74), we have $\langle\sigma_A\rangle_{t,s;\epsilon}$ equal to either 0 or 1.

C. Proof of Lemma IV.7

Using the fundamental theorem of calculus, the desired difference has an integral form

$$\begin{aligned}\mathbb{E}\langle Q_1\rangle_{t,0;\epsilon_n} - \mathbb{E}\langle Q_1\rangle_{t,s';\epsilon_n} &= \frac{1}{n}\int_{s'}^{0} ds \sum_{i=1}^{n} \frac{d}{ds}\mathbb{E}\langle \sigma_i\rangle_{t,s;\epsilon_n} \\ &= \frac{1}{n}\int_{s'}^{0} ds \sum_{i=1}^{n} \Big\{ \frac{K}{RT}\sum_{j=1}^{n}\big(\mathbb{E}\langle\sigma_i\rangle_{e_{j,s}^{(t)}+1} - \mathbb{E}\langle\sigma_i\rangle_{e_{j,s}^{(t)}}\big) - \frac{n}{RT}\big(\mathbb{E}\langle\sigma_i\rangle_{m_s^{(t)}+1} - \mathbb{E}\langle\sigma_i\rangle_{m_s^{(t)}}\big)\Big\} \end{aligned} \tag{75}$$

where (75) follows from the Poisson property (51). Since $|\langle\sigma_i\rangle|$ is bounded by 1 we see that (75) is bounded by $\mathcal{O}(n/T)$.

D. Proof of Lemma IV.8

For each n, we seek distributions $\mathsf{x}^{(t)} \in \mathcal{B}$ for $V^{(t)}$, $t = 1, \dots, T$ which solve equation (45). By symmetry between vertices $\mathbb{E}\langle Q_1\rangle_{t,0;\epsilon} = \mathbb{E}\langle\sigma_1\rangle_{t,0;\epsilon}$ so the equation becomes

$$\mathbb{E}\tanh V^{(t)} = \mathbb{E}\langle\sigma_1\rangle_{t,0;\epsilon}. \tag{76}$$

Recall that in our interpolation scheme the right hand side depends only on $\{\mathsf{x}^{(t')}\}_{t'<t}$ and is thus independent of $\mathsf{x}^{(t)}$. Thus it suffices to choose $V^{(t)}$ for $t = 1, \dots, T$ as follows: $V^{(t)} = +\infty$ with probability $\mathbb{E}\langle\sigma_1\rangle_{t,0;\epsilon}$ and $V^{(t)} = 0$ with probability $1 - \mathbb{E}\langle\sigma_1\rangle_{t,0;\epsilon}$. These are the distributions $\hat{\mathsf{x}}_n^{(t)} \in \mathcal{B}$ of the Lemma. It is clear that this solution is unique.

E. Proof of Lemma IV.1

Let us first outline the strategy. We will show that the stationarity condition for $\tilde{h}_\epsilon(\underline{\mathsf{x}})$ implies all $\mathsf{x}^{(t)}$ are equal for $t = 1, \dots, T$. Recall also that $\tilde{h}_{\epsilon=0}(\underline{\mathsf{x}}) = h_{\mathrm{RS}}(\mathsf{x})$ if $\mathsf{x}^{(1)} = \dots = \mathsf{x}^{(T)} = \mathsf{x}$. These two facts together tell us that, if the supremum is attained at a stationary point, then we have $\sup_{\underline{\mathsf{x}}\in\mathcal{B}^T}\tilde{h}_{\epsilon=0}(\underline{\mathsf{x}}) = \sup_{\mathsf{x}\in\mathcal{B}} h_{\mathrm{RS}}(\mathsf{x})$. Finally we will check that the supremum of $\tilde{h}_\epsilon(\underline{\mathsf{x}})$ cannot be attained at boundary points unless they are stationary.

In order to facilitate computations, we rewrite $\tilde{h}_\epsilon$ more explicitly in term of the distribution $\underline{\mathsf{x}}$ with the following formalism, already used in a coding theoretic context (see e.g. [22]). We define an entropy functional[4] $H : \mathcal{X} \to \mathbb{R}$ as

$$H(\mathsf{x}_1) \equiv \int \ln(1 + e^{-2a})\mathsf{x}_1(da).$$

Two convolution operators $\circledast, \boxast : \mathcal{X}\times\mathcal{X} \to \mathcal{X}$ are defined for $a_1 \sim \mathsf{x}_1, a_2 \sim \mathsf{x}_2$ such that $\mathsf{x}_1 \circledast \mathsf{x}_2$ is the distribution of $a_1 + a_2$ and $\mathsf{x}_1 \boxast \mathsf{x}_2$ is the distribution of $\tanh^{-1}(\tanh a_1 \tanh a_2)$. Therefore, the entropies of convolutions are

$$H(\circledast_{i=1}^k \mathsf{x}_i) = \int \ln(1 + e^{-2\sum_{i=1}^k a_i}) \prod_{i=1}^k \mathsf{x}_i(da_i),$$

$$H(\boxast_{i=1}^k \mathsf{x}_i) = -\int \ln\Big[\frac{1}{2}\Big(1 + \prod_{i=1}^k \tanh a_i\Big)\Big] \prod_{i=1}^k \mathsf{x}_i(da_i).$$

We define $\mathsf{x}_1^{\circledast 0} \equiv \Delta_0$, where Δ_0 is the identity of $\circledast$ and it is a distribution with solely a point mass at 0. We also define $\Lambda^{\circledast}(\mathsf{x}_1) \equiv \sum_{l=0}^{\infty}\Lambda_l \mathsf{x}_1^{\circledast l}$, where $\Lambda_l = \frac{(K/R)^l}{l!}e^{-K/R}$ denotes the probability that a variable node has degree l, and $\lambda^{\circledast}(\mathsf{x}_1) \equiv \sum_{l=1}\lambda_l \mathsf{x}_1^{\circledast(l-1)}$, where $\lambda_l = \frac{l\Lambda_l}{\Lambda'(1)} = \frac{(K/R)^{l-1}}{(l-1)!}e^{-K/R}$ denotes the probability that an edge is connected to a variable node of degree l. Let h be the distribution of the field H. One can check that (see Appendix C)

$$\begin{aligned}\tilde{h}_\epsilon(\underline{\mathsf{x}}) = &-\frac{K}{R}H\Big(\frac{1}{T}\sum_{t=1}^{T}\mathsf{c}\boxast(\mathsf{x}^{(t)})^{\boxast(K-1)}\Big) + H\Big(\mathsf{h}\circledast\Lambda^{\circledast}\Big(\frac{1}{T}\sum_{t=1}^{T}\mathsf{c}\boxast(\mathsf{x}^{(t)})^{\boxast(K-1)}\Big)\Big) \\ &+ \frac{1}{R}H(\mathsf{c}) + \frac{K-1}{RT}\sum_{t=1}^{T}H\big(\mathsf{c}\boxast(\mathsf{x}^{(t)})^{\boxast K}\big). \end{aligned} \tag{77}$$

The directional derivative of a function $F : \mathcal{X} \to \mathbb{R}$ at x_1 in the direction $\Delta\mathsf{x}_1$ is

$$d_{\mathsf{x}_1}F(\mathsf{x}_1)[\Delta\mathsf{x}_1] \equiv \lim_{\delta\to 0}\frac{F(\mathsf{x}_1 + \Delta\mathsf{x}_1) - F(\mathsf{x}_1)}{\delta}.$$

The stationary condition is then given by $d_{\mathsf{x}^{(t)}}\tilde{h}_\epsilon(\underline{\mathsf{x}})[\Delta\mathsf{x}^{(t)}] = 0$.

To compute $d_{\mathsf{x}^{(t)}}\tilde{h}_\epsilon(\underline{\mathsf{x}})[\Delta\mathsf{x}^{(t)}]$, we employ the following computational rules.

[4]The notation H for the entropy should not be confused with the notation H for the perturbation field in the model.

Lemma VII.1 ([22, Propositions 14 and 15]). *Let $F : \mathcal{X} \to \mathbb{R}$ be a linear functional, and $*$ be either $\circledast$ or $\boxast$. Then for $k \in \mathbb{Z}^+$, $\mathsf{x}_1, \mathsf{x}_2, \mathsf{x}_2' \in \mathcal{X}$ and letting $\mathsf{y} = \mathsf{x}_2 - \mathsf{x}_2'$, we have*

$$d_{\mathsf{x}_1} F(\mathsf{x}_1^{*k})[\mathsf{y}] = kF(\mathsf{x}_1^{*(k-1)} * \mathsf{y}).$$

For any polynomials p, q, we have

$$d_{\mathsf{x}_1} F\big(p^{\circledast}(q^{\boxast}(\mathsf{x}))\big)[\mathsf{y}] = F\big(p'^{\circledast}(q^{\boxast}(\mathsf{x}_1) \circledast (q'^{\boxast}(\mathsf{x}_1) \boxast \mathsf{y})\big).$$

Lemma VII.2 ([24, Theorem 4.41]). *For any $\mathsf{x}_1, \mathsf{x}_1', \mathsf{x}_2 \in \mathcal{X}$, we have*

$$H((\mathsf{x}_1 - \mathsf{x}_1') \circledast \mathsf{x}_2) + H((\mathsf{x}_1 - \mathsf{x}_1') \boxast \mathsf{x}_2) = H(\mathsf{x}_1 - \mathsf{x}_1').$$

Let $\mathsf{T}(\mathsf{h}, \mathsf{c}, \underline{\mathsf{x}}) \equiv \mathsf{h} \circledast \lambda^{\circledast}(\frac{1}{T}\sum_{t=1}^{T} \mathsf{c} \boxast (\mathsf{x}^{(t)})^{\boxast(K-1)})$. Using Lemma VII.1, $d_{\mathsf{x}^{(t)}} \tilde{h}(\underline{\mathsf{x}})[\Delta\mathsf{x}^{(t)}]$ is the sum of the following three terms:

$$-d_{\mathsf{x}^{(t)}} \frac{K}{R} H\Big(\frac{1}{T}\sum_{t=1}^{T} \mathsf{c} \boxast (\mathsf{x}^{(t)})^{\boxast(K-1)}\Big)[\Delta\mathsf{x}^{(t)}] = -\frac{K(K-1)}{RT} H\big(\mathsf{c} \boxast (\mathsf{x}^{(t)})^{\boxast(K-2)} \boxast \Delta\mathsf{x}^{(t)}\big), \tag{78}$$

$$d_{\mathsf{x}^{(t)}} H\Big(\mathsf{h} \circledast \Lambda^{\circledast}\Big(\frac{1}{T}\sum_{t=1}^{T} \mathsf{c} \boxast (\mathsf{x}^{(t)})^{\boxast(K-1)}\Big)\Big)[\Delta\mathsf{x}^{(t)}] = \frac{K(K-1)}{RT} H\big(\mathsf{T}(\mathsf{h}, \mathsf{c}, \underline{\mathsf{x}}) \circledast \big(\mathsf{c} \boxast (\mathsf{x}^{(t)})^{\boxast(K-2)} \boxast \Delta\mathsf{x}^{(t)}\big)\big) \tag{79}$$

$$-d_{\mathsf{x}^{(t)}} \frac{K-1}{RT} \sum_{t=1}^{T} H\big(\mathsf{c} \boxast (\mathsf{x}^{(t)})^{\boxast K}\big)[\Delta\mathsf{x}^{(t)}] = -\frac{K(K-1)}{RT} H\big(\mathsf{c} \boxast (\mathsf{x}^{(t)})^{\boxast(K-1)} \boxast \Delta\mathsf{x}^{(t)}\big). \tag{80}$$

In addition, we use Lemma VII.2 to rewrite (79) as

$$\frac{K(K-1)}{RT} \Big\{ H\left(\mathsf{c} \boxast (\mathsf{x}^{(t)})^{\boxast(K-2)} \boxast \Delta\mathsf{x}^{(t)}\right) - H\left(\mathsf{T}(\mathsf{h}, \mathsf{c}, \underline{\mathsf{x}}) \boxast \left(\mathsf{c} \boxast (\mathsf{x}^{(t)})^{\boxast(K-2)} \boxast \Delta\mathsf{x}^{(t)}\right)\right) \Big\}. \tag{81}$$

Putting (78), (80) and (81) together, we have

$$d_{\mathsf{x}^{(t)}} \tilde{h}_{\epsilon}(\underline{\mathsf{x}})[\Delta\mathsf{x}^{(t)}] = \frac{K(K-1)}{RT} H\left(\left(\mathsf{x}^{(t)} - \mathsf{T}(\mathsf{h}, \mathsf{c}, \underline{\mathsf{x}})\right) \boxast \left(\mathsf{c} \boxast (\mathsf{x}^{(t)})^{\boxast(K-2)} \boxast \Delta\mathsf{x}^{(t)}\right)\right), \tag{82}$$

which implies that $\mathsf{x}^{(t)}$ is a stationnary point of $\tilde{h}_{\epsilon}(\underline{\mathsf{x}})[\Delta\mathsf{x}^{(t)}]$ if and only if it satisfies the equation

$$\mathsf{x}^{(t)} = \mathsf{T}(\mathsf{h}, \mathsf{c}, \underline{\mathsf{x}}), \qquad t = 1, \ldots, T. \tag{83}$$

In particular we have $\mathsf{x}^{(1)} = \ldots \mathsf{x}^{(T)}$ as claimed at the beginning of the paragraph.

Now we check that if the supremum is attained at the boundary then it must be a stationary point. Let Δ_∞ is a distribution with solely a point mass at ∞. Recall that Δ_0 is a distribution with solely a point mass at 0. Any distribution $\mathsf{x} \in \mathcal{B}$ has the form $\mathsf{x} = x\Delta_0 + (1-x)\Delta_\infty$ for some $x \in [0, 1]$. A boundary point $\underline{\mathsf{x}} \in \mathcal{B}^T$ is a point which contains $\mathsf{x}^{(t)}$ for some $t \in \{1, \ldots, T\}$ equal to either Δ_0 or Δ_∞. For $\mathsf{x}_1, \mathsf{x}_1' \in \mathcal{B}$ we denote $\mathsf{x}_1' \succ \mathsf{x}_1$ if $x_1' > x_1$; in other words $\mathsf{x}_1' - \mathsf{x}_1 = \alpha\Delta_0 - \alpha\Delta_\infty$ for some $\alpha \in (0, 1]$.

Lemma VII.3 ([22, Proposition 8]). *Let $\mathsf{x}_1, \mathsf{x}_1', \mathsf{x}_2, \mathsf{x}_2' \in \mathcal{B}$ with $\mathsf{x}_1' \succ \mathsf{x}_1$ and $\mathsf{x}_2' \succ \mathsf{x}_2$. Let $\mathsf{y}_1 = \mathsf{x}_1' - \mathsf{x}_1$ and $\mathsf{y}_2 = \mathsf{x}_2' - \mathsf{x}_2$. We have*

$$H(\mathsf{y}_1 \boxast \mathsf{y}_2) < 0.$$

Any boundary point $\underline{\mathsf{x}}$ which is not stationary belongs to one of the following two cases. We will show that substituting such $\underline{\mathsf{x}}$ into (82) gives $d_{\mathsf{x}^{(t)}} \tilde{h}_{\epsilon}(\underline{\mathsf{x}})[\Delta\mathsf{x}^{(t)}] > 0$ for some direction $\Delta\mathsf{x}^{(t)}$, which means that this $\underline{\mathsf{x}}$ is not the supremum point. The positivity of the directional derivative is a result of Lemma VII.3.

- Let $\underline{\mathsf{x}} \in \mathcal{B}^T$ be a vector of distributions containing at least one component $t \in \{1, \ldots, T\}$ with $\mathsf{x}^{(t)} = \Delta_\infty$ and at least one component $t' \in \{1, \ldots, T\} \setminus t$ with $\mathsf{x}^{(t')} = x^{(t')}\Delta_0 + (1 - x^{(t')})\Delta_\infty$, $x^{(t')} \in (0, 1]$. We have $\mathsf{T}(\mathsf{h}, \mathsf{c}, \underline{\mathsf{x}}) \succ \Delta_\infty$. Furthermore, $\Delta\mathsf{x}^{(t)}$ can only be in the form $\mathsf{x}_1' - \Delta_\infty$ with $\mathsf{x}_1' \succ \Delta_\infty$. Let

$$\mathsf{y}_1 \equiv \mathsf{T}(\mathsf{h}, \mathsf{c}, \underline{\mathsf{x}}) - \Delta_\infty,$$
$$\mathsf{y}_2 \equiv \mathsf{c} \boxast (\mathsf{x}^{(t)})^{\boxast(K-2)} \boxast \Delta\mathsf{x}^{(t)}.$$

By Lemma VII.3 we have

$$d_{\mathsf{x}^{(t)}}\tilde{h}_\epsilon(\underline{\mathsf{x}})[\Delta\mathsf{x}^{(t)}] = -\frac{K(K-1)}{RT}H(\mathsf{y}_1 \circledast \mathsf{y}_2) > 0. \tag{84}$$

- Let $\underline{\mathsf{x}} \in \mathcal{B}^T$ be a vector of distributions containing at least one component $t \in \{1,\dots,T\}$ with $\mathsf{x}^{(t)} = \Delta_0$ and at least one component $t' \in \{1,\dots,T\}\setminus t$ with $\mathsf{x}^{(t')} = x^{(t')}\Delta_0 + (1-x^{(t')})\Delta_\infty$, $x^{(t')} \in [0,1)$. We have $\mathsf{T}(\mathsf{h},\mathsf{c},\underline{\mathsf{x}}) \prec \Delta_0$. Furthermore, $\Delta\mathsf{x}^{(t)}$ can only be in the form $\mathsf{x}_1' - \Delta_0$ with $\mathsf{x}_1' \prec \Delta_0$. Let

$$\begin{aligned}
\mathsf{y}_1 &\equiv \Delta_0 - \mathsf{T}(\mathsf{h},\mathsf{c},\underline{\mathsf{x}}),\\
\mathsf{y}_2 &\equiv -\mathsf{c} \circledast (\mathsf{x}^{(t)})^{\circledast(K-2)} \circledast \Delta\mathsf{x}^{(t)}.
\end{aligned}$$

By Lemma VII.3 we have

$$d_{\mathsf{x}^{(t)}}\tilde{h}_\epsilon(\underline{\mathsf{x}})[\Delta\mathsf{x}^{(t)}] = -\frac{K(K-1)}{RT}H(\mathsf{y}_1 \circledast \mathsf{y}_2) > 0. \tag{85}$$

APPENDIX A
DIRECT PROOF OF IDENTITY (12) FOR SYMMETRIC DISTRIBUTIONS

If $\mathsf{x}(-dh) = e^{-2h}\mathsf{x}(dh)$ holds, then we have

$$\begin{aligned}
&\int_{-\infty}^{\infty}(\tanh h)^{2k-1}\,\mathsf{x}(dh) = \int_0^{\infty}(\tanh h)^{2k-1}\,\mathsf{x}(dh) - \int_0^{\infty}(\tanh h)^{2k-1}\,\mathsf{x}(-dh)\\
&= \int_0^{\infty}(\tanh h)^{2k-1}(1-e^{-2h})\,\mathsf{x}(dh) = \int_0^{\infty}(\tanh h)^{2k}(1+e^{-2h})\,\mathsf{x}(dh)\\
&= \int_0^{\infty}(\tanh h)^{2k}\,\mathsf{x}(dh) + \int_0^{\infty}(\tanh h)^{2k}\,\mathsf{x}(-dh) = \int_{-\infty}^{\infty}(\tanh h)^{2k}\,\mathsf{x}(dh).
\end{aligned}$$

APPENDIX B
REMOVING THE ϵ-PERTURBATION: PROOF OF EQUATION (34)

Lemma B.1. *Let $F_n : [0,1] \to \mathbb{R}_+$ be a sequence of non-negative continuous functions. Suppose that*

$$\lim_{n\to+\infty}\int_0^1 d\epsilon F_n(\epsilon) = 0.$$

Given any $\bar\epsilon \in [0,1]$ we can find a sequence $\epsilon_n \to \bar\epsilon$, $n \to +\infty$ such that $\lim_{n\to+\infty} F_n(\epsilon_n) = 0$.

Proof. Set $a_n = \sqrt{\int_0^1 d\epsilon F_n(\epsilon)}$. First we consider the case $\bar\epsilon \in]0,1[$. Note that $\lim_{n\to+\infty} a_n = 0$ so n large enough $[\bar\epsilon - a_n, \bar\epsilon + a_n] \subset [0,1]$ thus

$$\int_0^1 d\epsilon F_n(\epsilon) \geq \int_{\bar\epsilon-a_n}^{\bar\epsilon+a_n} d\epsilon F_n(\epsilon).$$

The mean value theorem tells us that there exists $\epsilon_n \in [\bar\epsilon - a_n, \bar\epsilon + a_n]$ such that the right hand side equals $2a_nF_n(\epsilon_n)$. Therefore

$$0 \leq F_n(\epsilon_n) \leq \frac{1}{2}\sqrt{\int_0^1 d\epsilon F_n(\epsilon)}$$

which implies the claim for $\bar\epsilon \in]0,1[$. Now we consider $\bar\epsilon = 0$. Similarly as before we have for n large enough

$$\int_0^1 d\epsilon F_n(\epsilon) \geq \int_0^{a_n} d\epsilon F_n(\epsilon) = a_nF_n(\epsilon_n).$$

for some ϵ_n by the mean value theorem. This implies $0 \leq F_n(\epsilon_n) \leq \sqrt{\int_0^1 d\epsilon F_n(\epsilon)}$ and the claim follows. The case $\bar\epsilon = 1$ is treated in the same way.

□

In our application we take $F_n(\epsilon) = \mathbb{E}\langle|Q_p^K - \langle Q_p\rangle_{t,s;\epsilon}^K|\rangle_{t,s;\epsilon}$. The whole point of this lemma is that although the functions F_n are uniformly bounded we do not a priori know if their pointwise limit exists almost everywhere and cannot use Lebesgue's dominated convergence theorem.

Appendix C
Rewriting the replica formula: Proof of equation (77)

We copy again

$$\tilde{h}_\epsilon(\underline{\mathsf{x}}) = \mathbb{E}\Big[\ln\Big(\prod_{t=1}^{T}\prod_{b=1}^{l}(1+\tanh U_b^{(t)}) + e^{-2H}\prod_{t=1}^{T}\prod_{b=1}^{l}(1-\tanh U_b^{(t)})\Big) - \frac{K-1}{RT}\sum_{t=1}^{T}\ln\Big(1+\tanh\tilde{J}\prod_{i=1}^{K}\tanh V_i^{(t)}\Big) - \frac{1}{R}\ln(1+\tanh\tilde{J})\Big]. \tag{86}$$

The first term can be rewritten as

$$\begin{aligned}
&\mathbb{E}\ln\Big(\prod_{t=1}^{T}\prod_{b=1}^{l}(1+\tanh U_b^{(t)}) + e^{-2H}\prod_{t=1}^{T}\prod_{b=1}^{l}(1-\tanh U_b^{(t)})\Big)\\
&= \mathbb{E}\ln\Big(\prod_{t=1}^{T}\prod_{b=1}^{l}(1+\tanh U_b^{(t)})\Big) + \mathbb{E}\ln\Big(1+e^{-2H}\prod_{t=1}^{T}\prod_{b=1}^{l}\frac{1-\tanh U_b^{(t)}}{1+\tanh U_b^{(t)}}\Big)\\
&= \mathbb{E}\ln\Big(\prod_{t=1}^{T}\prod_{b=1}^{l}(1+\tanh U_b^{(t)})\Big) + \mathbb{E}\ln\Big(1+e^{-2(\sum_{t=1}^{T}\sum_{b=1}^{l}U_b^t+H)}\Big)\\
&= -\frac{K}{R}H\Big(\frac{1}{T}\sum_{t=1}^{T}\mathsf{c}\boxast(\mathsf{x}^{(t)})^{\boxast(K-1)}\Big) + \frac{K}{R}\ln 2 + H\Big(\mathsf{h}\circledast\Lambda^{\circledast}\Big(\frac{1}{T}\sum_{t=1}^{T}\mathsf{c}\boxast(\mathsf{x}^{(t)})^{\boxast(K-1)}\Big)\Big).
\end{aligned}$$

The second term can be easily seen to be equal to

$$-\frac{K-1}{RT}\sum_{t=1}^{T}\ln(1+\tanh\tilde{J}\prod_{i=1}^{K}\tanh V_i^{(t)}) = \frac{K-1}{RT}\sum_{t=1}^{T}H\big(\mathsf{c}\boxast(\mathsf{x}^{(t)})^{\boxast K}\big) - \frac{K-1}{R}\ln 2.$$

The remaining term is

$$-\frac{1}{R}\mathbb{E}\ln(1+\tanh\tilde{J}) = \frac{1}{R}(H(\mathsf{c})-\ln 2).$$

Appendix D
Derivatives of the conditional entropy: Proof of formulas (25) and (26)

This appendix is essentially an adaptation of [21] (see also [20, Chapter 2]).

A. *Proof of* (25)

It will be convenient to work with random fields H_i, $i=1,\dots,n$ which take value 0 with probability $1-\epsilon_i$ and value $+\infty$ with probability ϵ_i. At the end of our calculations we set all $\epsilon_i=\epsilon$ to recover the original interpolated system. We set $\underline{\epsilon}=(\epsilon_1,\dots,\epsilon_n)$ and the corresponding entropy denoted $h_{t,s;\underline{\epsilon}}$. To ease the notations we drop the subscripts $t,s;\underline{\epsilon}$ in the Gibbs brackets.

Let

$$Z_{t,s;\underline{\epsilon}}^{\sim i} \equiv \sum_{\underline{\sigma}\in\{-1,+1\}^n}\exp\Big(-\mathcal{H}_{t,s}(\underline{\sigma},\underline{\tilde{J}},\underline{U},\underline{m},\underline{e}) + \sum_{\substack{k=1\\k\neq i}}^{n}H_k(\sigma_k-1)\Big)$$

be the partition function associated with the Gibbs expectation $\langle -\rangle_{\sim H_i}$ where we have set $H_i=0$. With the identities

$$\ln\frac{Z_{t,s;\underline{\epsilon}}}{Z_{t,s;\underline{\epsilon}}^{\sim i}} = \ln\langle e^{H_i(\sigma_i-1)}\rangle_{\sim H_i},\qquad e^{H_i(\sigma_i-1)} = \frac{1+\sigma_i\tanh H_i}{1+\tanh H_i}, \tag{87}$$

we have

$$h_{t,s;\underline{\epsilon}} = \frac{1}{n}\mathbb{E}\ln Z_{t,s;\underline{\epsilon}}^{\sim i} + \frac{1}{n}\mathbb{E}\ln\frac{1+\langle\sigma_i\rangle_{\sim H_i}\tanh H_i}{1+\tanh H_i} \tag{88}$$

where $\mathbb{E}_{\sim H_i}$ is the expectation over all random variables except H_i. As $\tanh H_i$ and $\langle\sigma_i\rangle_{\sim H_i}$ equal either 0 or 1, (88) simplifies down to

$$h_{t,s;\underline{\epsilon}} = \frac{1}{n}\mathbb{E}\ln Z^{\sim i}_{t,s;\underline{\epsilon}} - \frac{1}{n}\epsilon_i \ln 2\,(1-\mathbb{E}\langle\sigma_i\rangle_{\sim H_i})\,. \tag{89}$$

Therefore, we have

$$\frac{d}{d\epsilon}h_{t,s;\epsilon} = \sum_{i=1}^{n}\frac{d}{d\epsilon_i}h_{t,s;\underline{\epsilon}}\Big|_{\epsilon_1,\dots,\epsilon_n=\epsilon} = -\frac{\ln 2}{n}\sum_{i=1}^{n}(1-\mathbb{E}\langle\sigma_i\rangle_{\sim H_i}) \tag{90}$$

which is the first equality of (25).

For obtaining the second equality from there, simply notice that as $1-\langle\sigma_i\rangle_{t,s;\epsilon}=0$ when $H_i=+\infty$ (which happens with probability ϵ), performing the expectation over H_i in the following expression we get

$$1-\mathbb{E}\langle\sigma_i\rangle_{t,s;\epsilon} = (1-\epsilon)(1-\mathbb{E}\langle\sigma_i\rangle_{\sim H_i})$$

which leads to the second equality when plugged in (90).

B. Proof of (26)

Let

$$Z^{\sim i,j}_{t,s;\underline{\epsilon}} \equiv \sum_{\underline{\sigma}\in\{-1,+1\}^n}\exp\Big(-\mathcal{H}_{t,s}(\underline{\sigma},\underline{\tilde{J}},\underline{U},\underline{m},\underline{e})+\sum_{\substack{k=1\\k\neq i,j}}^{n}H_k(\sigma_k-1)\Big)$$

be the partition function associated with the Gibbs expectation $\langle\cdot\rangle_{\sim H_i,H_j}$. Using again (87) on the identity

$$\ln\frac{Z_{t,s;\underline{\epsilon}}}{Z^{\sim i,j}_{t,s;\underline{\epsilon}}} = \ln\langle e^{H_i(\sigma_i-1)+H_j(\sigma_j-1)}\rangle_{\sim H_i,H_j},$$

we have

$$\begin{aligned}
h_{t,s;\underline{\epsilon}} &= \frac{1}{n}\mathbb{E}\ln Z^{\sim i,j}_{t,s;\underline{\epsilon}} + \frac{1}{n}\mathbb{E}\ln\frac{1+\langle\sigma_i\rangle_{\sim H_i,H_j}\tanh H_i+\langle\sigma_j\rangle_{\sim H_i,H_j}\tanh H_j+\langle\sigma_i\sigma_j\rangle_{\sim H_i,H_j}\tanh H_i\tanh H_j}{1+\tanh H_i+\tanh H_j+\tanh H_i\tanh H_j}\\
&= \frac{1}{n}\mathbb{E}\ln Z^{\sim i,j}_{t,s;\underline{\epsilon}} + \frac{\epsilon_i\epsilon_j}{n}\mathbb{E}_{\sim H_i,H_j}\ln\frac{1+\langle\sigma_i\rangle_{\sim H_i,H_j}+\langle\sigma_j\rangle_{\sim H_i,H_j}+\langle\sigma_i\sigma_j\rangle_{\sim H_i,H_j}}{4}\\
&\quad+\frac{\epsilon_i(1-\epsilon_j)}{n}\mathbb{E}_{\sim H_i,H_j}\ln\frac{1+\langle\sigma_i\rangle_{\sim H_i,H_j}}{2}+\frac{(1-\epsilon_i)\epsilon_j}{n}\mathbb{E}_{\sim H_i,H_j}\ln\frac{1+\langle\sigma_j\rangle_{\sim H_i,H_j}}{2},
\end{aligned} \tag{91}$$

where (91) follows from taking the expectation over H_i and H_j. From (89) one can deduce that $\frac{d^2}{d\epsilon_i^2}h_{t,s;\underline{\epsilon}}=0$. Therefore

$$\frac{d^2}{d\epsilon^2}h_{t,s;\epsilon} = \sum_{i,j=1}^{n}\frac{d^2}{d\epsilon_j d\epsilon_i}h_{t,s;\underline{\epsilon}}\Big|_{\epsilon_1,\dots,\epsilon_n=\epsilon} = \sum_{i\neq j}\frac{d^2}{d\epsilon_j d\epsilon_i}h_{t,s;\underline{\epsilon}}\Big|_{\epsilon_1,\dots,\epsilon_n=\epsilon}.$$

The derivatives $\frac{d^2}{d\epsilon_j d\epsilon_i}h_{t,s;\underline{\epsilon}}$ can be readily obtained from (91). This provides

$$\frac{d^2}{d\epsilon^2}h_{t,s;\epsilon} = \frac{1}{n}\sum_{i\neq j}\mathbb{E}_{\sim H_i,H_j}\ln\frac{1+\langle\sigma_i\rangle_{\sim H_i,H_j}+\langle\sigma_j\rangle_{\sim H_i,H_j}+\langle\sigma_i\sigma_j\rangle_{\sim H_i,H_j}}{1+\langle\sigma_i\rangle_{\sim H_i,H_j}+\langle\sigma_j\rangle_{\sim H_i,H_j}+\langle\sigma_i\rangle_{\sim H_i,H_j}\langle\sigma_j\rangle_{\sim H_i,H_j}}. \tag{92}$$

We now simplify each term in the sum (92). Given that $\langle\sigma_S\rangle_{\sim H_i,H_j}$ equals either 0 or 1 for any subsets $S\subset\{1\dots n\}$, one can verify that the numerator and denominator of (92) can be written as

$$\begin{aligned}
&\ln\big(1+\langle\sigma_i\rangle_{\sim H_i,H_j}+\langle\sigma_j\rangle_{\sim H_i,H_j}+\langle\sigma_i\sigma_j\rangle_{\sim H_i,H_j}\big) = (\langle\sigma_i\rangle_{\sim H_i,H_j}+\langle\sigma_j\rangle_{\sim H_i,H_j}+\langle\sigma_i\sigma_j\rangle_{\sim H_i,H_j})\ln 2\\
&\quad+(\langle\sigma_i\rangle_{\sim H_i,H_j}\langle\sigma_j\rangle_{\sim H_i,H_j}+\langle\sigma_i\rangle_{\sim H_i,H_j}\langle\sigma_i\sigma_j\rangle_{\sim H_i,H_j}+\langle\sigma_j\rangle_{\sim H_i,H_j}\langle\sigma_i\sigma_j\rangle_{\sim H_i,H_j})(\ln 3-2\ln 2)\\
&\quad+\langle\sigma_i\rangle_{\sim H_i,H_j}\langle\sigma_j\rangle_{\sim H_i,H_j}\langle\sigma_i\sigma_j\rangle_{\sim H_i,H_j}(5\ln 2-3\ln 3),
\end{aligned} \tag{93}$$

and

$$\ln\big(1+\langle\sigma_i\rangle_{\sim H_i,H_j}+\langle\sigma_j\rangle_{\sim H_i,H_j}+\langle\sigma_i\rangle_{\sim H_i,H_j}\langle\sigma_j\rangle_{\sim H_i,H_j}\big) = (\langle\sigma_i\rangle_{\sim H_i,H_j}+\langle\sigma_j\rangle_{\sim H_i,H_j})\ln 2.$$

Special cases of the Nishimori identities (11) in the form

$$\begin{aligned}
\mathbb{E}_{\sim H_i,H_j}[\langle\sigma_i\rangle_{\sim H_i,H_j}\langle\sigma_j\rangle_{\sim H_i,H_j}] &= \mathbb{E}_{\sim H_i,H_j}[\langle\sigma_i\rangle_{\sim H_i,H_j}\langle\sigma_j\rangle_{\sim H_i,H_j}\langle\sigma_i\sigma_j\rangle_{\sim H_i,H_j}],\\
\mathbb{E}_{\sim H_i,H_j}[\langle\sigma_i\rangle_{\sim H_i,H_j}\langle\sigma_i\sigma_j\rangle_{\sim H_i,H_j}] &= \mathbb{E}_{\sim H_i,H_j}[\langle\sigma_i\rangle_{\sim H_i,H_j}\langle\sigma_j\rangle_{\sim H_i,H_j}\langle\sigma_i\sigma_j\rangle_{\sim H_i,H_j}],\\
\mathbb{E}_{\sim H_i,H_j}[\langle\sigma_j\rangle_{\sim H_i,H_j}\langle\sigma_i\sigma_j\rangle_{\sim H_i,H_j}] &= \mathbb{E}_{\sim H_i,H_j}[\langle\sigma_i\rangle_{\sim H_i,H_j}\langle\sigma_j\rangle_{\sim H_i,H_j}\langle\sigma_i\sigma_j\rangle_{\sim H_i,H_j}],
\end{aligned}$$

can now be used to simplify (93) so that each term in the sum (92) becomes

$$\ln(2)\,\mathbb{E}_{\sim H_i,H_j}[\langle\sigma_i\sigma_j\rangle_{\sim H_i,H_j} - \langle\sigma_i\rangle_{\sim H_i,H_j}\langle\sigma_j\rangle_{\sim H_i,H_j}]. \tag{94}$$

Moreover, as $\langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle = 0$ when H_i and/or H_j equals $+\infty$, we obtain

$$\mathbb{E}[\langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle] = (1-\epsilon_i)(1-\epsilon_j)\mathbb{E}_{\sim H_i,H_j}[\langle\sigma_i\sigma_j\rangle_{\sim H_i,H_j} - \langle\sigma_i\rangle_{\sim H_i,H_j}\langle\sigma_j\rangle_{\sim H_i,H_j}]. \tag{95}$$

Finally, from (92), (94) and (95) we obtain (26).

APPENDIX E
CONCENTRATION OF FREE ENERGY

Let $\mathcal{J}$ collect both the realization of $\underline{\tilde{J}}$ and the graph realization of all the factor nodes carrying elements in $\underline{\tilde{J}}$. Let $\mathcal{U}$ collect both the realization of $\underline{U}$ and the graph realization of all the half edges carrying elements in $\underline{U}$. The proof of Lemma VI.3 can be decomposed into the following three lemmas. We stress that the three Lemmas E.1, E.2, E.3 are under the condition that $\underline{\tilde{J}}, \underline{U}$ are non-negative such that we can make use of the consequence $\langle\sigma_S\rangle_{t,s;\epsilon,\delta} \geq 0$ where S is any subset of $\{1,\dots,n\}$. Finally recall definitions (63) and (64).

Lemma E.1 (Concentration w.r.t. $\underline{H}$)**.** *For any $s,\epsilon,\delta \in [0,1]$, $t = 1,\dots,T$, $\nu > 0$ and any realization $\underline{H}$ we have*

$$\mathbb{P}(|\tilde{F}_{t,s;\epsilon}(\delta) - \mathbb{E}_{\underline{H}}\tilde{F}_{t,s;\epsilon}(\delta)| \geq \nu/3) \leq 2\exp\Big(-\frac{2n\nu^2}{(3\ln 2)^2}\Big). \tag{96}$$

Lemma E.2 (Concentration w.r.t. $\mathcal{J}$)**.** *For any $s,\epsilon,\delta \in [0,1]$, $t = 1,\dots,T$, $\nu > 0$ and any realization $\mathcal{J}$ there exists a constant $C_1 > 0$ such that*

$$\mathbb{P}(|\mathbb{E}_{\underline{H}}\tilde{F}_{t,s;\epsilon}(\delta) - \mathbb{E}_{\underline{H},\mathcal{J}}\tilde{F}_{t,s;\epsilon}(\delta)| \geq \nu/3) \leq 3\exp(-n\nu^2 C_1). \tag{97}$$

Lemma E.3 (Concentration w.r.t. $\mathcal{U}$)**.** *For any $s,\epsilon,\delta \in [0,1]$, $t = 1,\dots,T$, $\nu > 0$ and any realization $\mathcal{U}$ there exists a constant $C_2 > 0$ such that*

$$\mathbb{P}(|\mathbb{E}_{\underline{H},\mathcal{J}}\tilde{F}_{t,s;\epsilon}(\delta) - \tilde{f}_{t,s;\epsilon}(\delta)| \geq \nu/3) \leq 3\exp(-n\nu^2 C_2). \tag{98}$$

Lemmas E.1 to E.3 are consequences of McDiarmid's inequality, which states that if $X_1,\dots,X_N$ are independent variables and g is a function satisfying

$$|g(x_1,\dots,x_i,\dots,x_N) - g(x_1,\dots,x_i',\dots,x_N)| \leq d_i \qquad \forall\, i = 1,\dots,N$$

then for any $\nu > 0$ we have

$$\mathbb{P}(|g(\underline{X}) - \mathbb{E}_{\underline{X}}g(\underline{X})| \geq \nu) \leq 2\exp\Big(-\frac{2\nu^2}{\sum_{i=1}^N d_i^2}\Big).$$

We provide the proof of those three lemmas at the end of this section.

From the triangle inequality and the union bound we have

$$\begin{aligned}\mathbb{P}(|\tilde{F}_{t,s;\epsilon}(\delta) - \tilde{f}_{t,s;\epsilon}(\delta)| \geq \nu) \leq{}& \mathbb{P}(|\tilde{F}_{t,s;\epsilon}(\delta) - \mathbb{E}_{\underline{H}}\tilde{F}_{t,s;\epsilon}(\delta)| \geq \nu/3) + \mathbb{P}(|\mathbb{E}_{\underline{H}}\tilde{F}_{t,s;\epsilon}(\delta) - \mathbb{E}_{\underline{H},\mathcal{J}}\tilde{F}_{t,s;\epsilon}(\delta)| \geq \nu/3)\\ &+ \mathbb{P}(|\mathbb{E}_{\underline{H},\mathcal{J}}\tilde{F}_{t,s;\epsilon}(\delta) - \tilde{f}_{t,s;\epsilon}(\delta)]| \geq \nu/3).\end{aligned} \tag{99}$$

Substituting (96), (97) and (98) followed by setting $C \equiv \min\{\frac{2}{(3\ln 2)^2}, C_1, C_2\}$, we have

$$\mathbb{P}(|\tilde{F}_{t,s;\epsilon}(\delta) - \tilde{f}_{t,s;\epsilon}(\delta)| \geq \nu) \leq 8\exp(-n\nu^2 C). \tag{100}$$

Let $D \equiv |\tilde{F}_{t,s;\epsilon}(\delta) - \tilde{f}_{t,s;\epsilon}(\delta)|$. We have

$$\int_0^\infty d\nu\,\nu\mathbb{P}(D \geq \nu) = \int_0^\infty d\nu\,\nu\,\mathbb{E}_D\mathbb{I}(D \geq \nu) = \mathbb{E}_D\int_0^\infty d\nu\,\nu\,\mathbb{I}(D \geq \nu) = \mathbb{E}_D\int_0^D d\nu\,\nu = \frac{1}{2}\mathbb{E}_D[D^2]. \tag{101}$$

Substituting (100) into (101), we have the required bound for Lemma VI.3:

$$\mathbb{E}\big[(\tilde{F}_{t,s;\epsilon}(\delta) - \tilde{f}_{t,s;\epsilon}(\delta))^2\big] = 2\int_0^\infty d\nu\,\nu P(|\tilde{F}_{t,s;\epsilon}(\delta) - \tilde{f}_{t,s;\epsilon}(\delta)| \geq \nu) \leq 16\int_0^\infty d\nu\,\nu\,e^{-n\nu^2 C} = \frac{8}{Cn}.$$

A. *Proof of Lemma E.1*

Consider $g(H_1, \dots, H_n) \equiv \tilde{F}_{t,s;\epsilon}(\delta)$ with $H_i \in \{0, \infty\}$ (note that $\tilde{F}_{t,s;\epsilon}(\delta)$ given by (63) is averaged over $\underline{\tilde{H}}$, not $\underline{H}$). As for all $i = 1, \dots, n$ the function g satisfies

$$
\begin{aligned}
|g(H_1, \dots, H_i, \dots, H_n) - g(H_1, \dots, H_i', \dots, H_n)| &= \Big|\frac{1}{n}\mathbb{E}_{\underline{\tilde{H}}} \ln\langle e^{H_i(\sigma_i - 1)}\rangle_{t,s;\epsilon,\delta}\Big| \\
&= \Big|\frac{1}{n}\mathbb{E}_{\underline{\tilde{H}}} \ln(1 + \langle\sigma_i\rangle_{t,s;\epsilon,\delta}\tanh H_i) - \frac{1}{n}\mathbb{E}_{\underline{\tilde{H}}}\ln(1+\tanh H_i)\Big| \\
&\leq \frac{\ln 2}{n}.
\end{aligned}
$$

McDiarmid's inequality immediately gives the lemma.

B. *Proof of Lemma E.2*

Let $|\underline{\tilde{J}}|$ be the number of components of the vector $\underline{\tilde{J}}$. From the construction of $\mathcal{G}_{t,s}$ in Section IV-A, we have $\mathbb{E}[|\underline{\tilde{J}}|] = \frac{n}{RT}(T - t + 1 - s) \leq \frac{n}{R}$. Set $m_{\max} = (1+\gamma)\frac{n}{R}$ for $\gamma > 0$. The event $|\underline{\tilde{J}}| > m_{\max}$ can be bounded by a relaxed form of the Chernoff bound as follows.

Lemma E.4 (Chernoff bound, [28, Theorem 4.4])**.** *Let $X = \sum_{i=1}^N X_i$ where $X_i = 1$ with probability p_i and $X_i = 0$ with probability $1 - p_i$, and all X_i are independent. Let $\mu = \mathbb{E}[X] = \sum_{i=1}^N p_i$. Then for all $\gamma > 0$ we have*

$$
\mathbb{P}(X > (1+\gamma)\mu) \leq \exp\Big(-\frac{\mu}{3}\min\{\gamma, \gamma^2\}\Big).
$$

By the Chernoff bound we have

$$
\mathbb{P}(|\underline{\tilde{J}}| > m_{\max}) \leq \exp\Big(-\frac{n}{3R}\min\{\gamma,\gamma^2\}\Big). \tag{102}
$$

Conditioned on $|\underline{\tilde{J}}| \leq m_{\max}$, we can have the representation $\mathcal{J} = (c_1, \dots, c_{m_{\max}})$ where for $a = 1, \dots, m_{\max}$ the profile $c_a \equiv (A_a, \tilde{J}_a)$ encodes that a factor node with weight $\tilde{J}_a$ is connected to a K-tuple identified by A_a. For $m < a \leq m_{\max}$ we denote $c_a = (\emptyset, 0)$.

Now consider $g(c_1, \dots, c_{m_{\max}}) \equiv \mathbb{E}_{\underline{H}}\tilde{F}_{t,s;\epsilon}(\delta)$ and pick a c_a for a given a. Let $c_a' \equiv \big(A_a', \tilde{J}_a'\big)$ be a new profile with either $A_a \neq A_a'$ or $\tilde{J}_a \neq \tilde{J}_a'$. Also let $c_a'' \equiv (A_a, 0)$ and $c_a''' \equiv (A_a', 0)$. Note that $g(c_1, \dots, c_a'', \dots, c_{m_{\max}}) = g(c_1, \dots, c_a''', \dots, c_{m_{\max}})$. We then have

$$
\begin{aligned}
&|g(c_1, \dots, c_a, \dots, c_{m_{\max}}) - g(c_1, \dots, c_a', \dots, c_{m_{\max}})| \\
&= |g(c_1, \dots, c_a, \dots, c_{m_{\max}}) - g(c_1, \dots, c_a'', \dots, c_{m_{\max}}) + g(c_1, \dots, c_a''', \dots, c_{m_{\max}}) - g(c_1, \dots, c_a', \dots, c_{m_{\max}})| \\
&\leq |g(c_1, \dots, c_a, \dots, c_{m_{\max}}) - g(c_1, \dots, c_a'', \dots, c_{m_{\max}})| + |g(c_1, \dots, c_a''', \dots, c_{m_{\max}}) - g(c_1, \dots, c_a', \dots, c_{m_{\max}})| \\
&= \Big|\frac{1}{n}\mathbb{E}_{\underline{\tilde{H}},\underline{H}}\ln\langle e^{\tilde{J}_a(\sigma_{A_a}-1)}\rangle_{t,s;\epsilon,\delta}\Big| + \Big|\frac{1}{n}\mathbb{E}_{\underline{\tilde{H}},\underline{H}}\ln\langle e^{\tilde{J}_a'(\sigma_{A_a'}-1)}\rangle_{t,s;\epsilon,\delta}\Big| \\
&= \Big|\frac{1}{n}\mathbb{E}_{\underline{\tilde{H}},\underline{H}}\ln(1 + \langle\sigma_{A_a}\rangle_{t,s;\epsilon,\delta}\tanh\tilde{J}_a) - \frac{1}{n}\mathbb{E}_{\underline{\tilde{H}},\underline{H}}\ln(1+\tanh\tilde{J}_a)\Big| \\
&\qquad + \Big|\frac{1}{n}\mathbb{E}_{\underline{\tilde{H}},\underline{H}}\ln(1 + \langle\sigma_{A_a'}\rangle_{t,s;\epsilon,\delta}\tanh\tilde{J}_a') - \frac{1}{n}\mathbb{E}_{\underline{\tilde{H}},\underline{H}}\ln(1+\tanh\tilde{J}_a')\Big| \\
&\leq \frac{2\ln 2}{n}
\end{aligned}
$$

where we used an identity of the form (87) for the last equality. This allows the use of McDiarmid's inequality to obtain

$$
\mathbb{P}(|\mathbb{E}_{\underline{H}}\tilde{F}_{t,s;\epsilon}(\delta) - \mathbb{E}_{\underline{H},\mathcal{J}}\tilde{F}_{t,s;\epsilon}(\delta)| \geq \nu/3 \mid |\underline{\tilde{J}}| \leq m_{\max}) \leq 2\exp\left(-\frac{n\nu^2R}{18(\ln 2)^2}\right). \tag{103}
$$

Finally, we take the union bound based on (102) and (103):

$$
\mathbb{P}(|\mathbb{E}_{\underline{H}}\tilde{F}_{t,s;\epsilon}(\delta) - \mathbb{E}_{\underline{H},\mathcal{J}}\tilde{F}_{t,s;\epsilon}(\delta)| \geq \nu/3) \leq 2\exp\left(-\frac{n\nu^2R}{18(\ln 2)^2}\right) + \exp\left(-\frac{n}{3R}\min\{\gamma,\gamma^2\}\right).
$$

Choosing $\nu^2 = \min\{\gamma, \gamma^2\}$ and $C_1 = \min\{\frac{R}{18(\ln 2)^2}, \frac{1}{3R}\}$, we obtain the lemma.

C. *Proof of Lemma E.3*

This proof can adopt the same presentation as in the proof of Lemma E.2 by noting that in the construction of $\mathcal{G}_{t,s}$ the Poisson process of adding half edges with weight $U_{a\to i}^{(t')}$ can be rephrased as follows:

1) (*Create all the messages without specifying their location*): We draw the random numbers $e_i^{(t')}$, $e_{i,s}^{(t)}$ and create the associated number of copies of $U^{(t')}$ for $t' = 1, \ldots, t$. We collect all $U^{(t')}$ to form a set $\{U_k\}_{k=1,\ldots,w}$, where w follows a Poisson distribution with mean $\frac{nK}{RT}(t-1+s) \leq \frac{nK}{R}$.
2) (*Specify the location of the messages*): Given the number w and the set $\{U_k\}$, we attach each U_k to variable node i chosen randomly and uniformly.

Let $w_{\max} = (1+\gamma)\frac{nK}{R}$. The Chernoff bound (Lemma E.4) provides that

$$\mathbb{P}(w > w_{\max}) \leq \exp\left(-\frac{nK}{3R}\min\{\gamma, \gamma^2\}\right). \tag{104}$$

Conditioned on $w \leq w_{\max}$, we have the representation $\mathcal{U} = (u_1, \ldots, u_{w_{\max}})$ where for $k = 1, \ldots, w_{\max}$ the profile $u_k = (i_k, U_k)$ represents that a half edge with weight U_k is connected to variable node i_k. For $w < k \leq w_{\max}$ we denote $u_k = (\emptyset, 0)$.

Now consider $g(u_1, \ldots, u_{m_{\max}}) \equiv \mathbb{E}_{\underline{H},\mathcal{J}}\tilde{F}_{t,s;\epsilon}(\delta)$ and pick any u_k. Let $u_k' \equiv (i_k', U_k)$ be a new profile with either $i_k \neq i_k'$ or $U_k \neq U_k'$. Also let $u_k'' = (i_k, 0)$ and $u_k'' = (i_k', 0)$. Note that $g(u_1, \ldots, u_k'', \ldots, u_{m_{\max}}) = g(u_1, \ldots, u_k''', \ldots, u_{w_{\max}})$. We then have

$$\begin{aligned}
&|g(u_1, \ldots, u_k, \ldots, u_{w_{\max}}) - g(u_1, \ldots, u_k', \ldots, u_{w_{\max}})| \\
&= |g(u_1, \ldots, u_k, \ldots, u_{w_{\max}}) - g(u_1, \ldots, u_k'', \ldots, u_{w_{\max}}) + g(u_1, \ldots, u_k''', \ldots, u_{w_{\max}}) - g(u_1, \ldots, u_k', \ldots, u_{w_{\max}})| \\
&\leq |g(u_1, \ldots, u_k, \ldots, u_{w_{\max}}) - g(u_1, \ldots, u_k'', \ldots, u_{w_{\max}})| + |g(u_1, \ldots, u_k''', \ldots, u_{w_{\max}}) - g(u_1, \ldots, u_k', \ldots, u_{w_{\max}})| \\
&= \left|\frac{1}{n}\mathbb{E}_{\underline{\tilde{H}},\underline{H},\mathcal{J}} \ln\langle e^{U_k(\sigma_{i_k}-1)}\rangle_{t,s;\epsilon,\delta}\right| + \left|\frac{1}{n}\mathbb{E}_{\underline{\tilde{H}},\underline{H},\mathcal{J}} \ln\langle e^{U_k'(\sigma_{i_k'}-1)}\rangle_{t,s;\epsilon,\delta}\right| \\
&= \left|\frac{1}{n}\mathbb{E}_{\underline{\tilde{H}},\underline{H},\mathcal{J}} \ln(1 + \langle\sigma_{i_k}\rangle_{t,s;\epsilon,\delta}\tanh U_k) - \frac{1}{n}\mathbb{E}_{\underline{\tilde{H}},\underline{H},\mathcal{J}} \ln(1+\tanh U_k)\right| \\
&\qquad + \left|\frac{1}{n}\mathbb{E}_{\underline{\tilde{H}},\underline{H},\mathcal{J}} \ln(1 + \langle\sigma_{i_k'}\rangle_{t,s;\epsilon,\delta}\tanh U_k') - \frac{1}{n}\mathbb{E}_{\underline{\tilde{H}},\underline{H},\mathcal{J}} \ln(1+\tanh U_k')\right| \\
&\leq \frac{2\ln 2}{n}.
\end{aligned}$$

McDiarmid's inequality is used to obtain

$$\mathbb{P}(|\mathbb{E}_{\underline{H},\mathcal{J}}\tilde{F}_{t,s;\epsilon}(\delta) - \tilde{f}_{t,s;\epsilon}(\delta)| \geq \nu/3 \mid w \leq w_{\max}) \leq 2\exp\left(-\frac{n\nu^2 R}{18(\ln 2)^2 K}\right). \tag{105}$$

Finally, we take the union bound based on (104) and (105):

$$\mathbb{P}(|\mathbb{E}_{\underline{H},\mathcal{J}}\tilde{F}_{t,s;\epsilon}(\delta) - \tilde{f}_{t,s;\epsilon}(\delta)]| \geq \nu/3) \leq 2\exp\left(-\frac{n\nu^2 R}{18(\ln 2)^2 K}\right) + \exp\left(-\frac{nK}{3R}\min\{\gamma,\gamma^2\}\right).$$

Choosing $\nu^2 = \min\{\gamma, \gamma^2\}$ and $C_2 = \min\{\frac{R}{18(\ln 2)^2 K}, \frac{K}{3R}\}$, we obtain the lemma.

Acknowledgments

Jean Barbier and Chun Lam Chan acknowledge the SNSF grant no. 200021-156672.

References

[1] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin-Glass Theory and Beyond*, ser. Lecture Notes in Physics. Singapore: World Scientific, 1987, vol. 9.

[2] M. Mézard and A. Montanari, *Information, Physics and Computation*. Oxford Press, 2009.

[3] J. Barbier, M. Dia, N. Macris, and F. Krzakala, "The Mutual Information in Random Linear Estimation," in *in the 54th Annual Allerton Conference on Communication, Control, and Computing*, September 2016.

[4] J. Barbier, N. Macris, M. Dia, and F. Krzakala, "Mutual Information and Optimality of Approximate Message-Passing in Random Linear Estimation." [Online]. Available: https://arxiv.org/pdf/1701.05823v1.pdf

[5] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for compressed sensing with gaussian matrices is exact," 2016. [Online]. Available: http://arxiv.org/abs/1607.02524

[6] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová, "Phase transitions, optimal errors and optimality of message-passing in generalized linear models," *arXiv preprint arXiv:1708.03395*, 2017.

[7] J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborová, "Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula," in *Advances in Neural Information Processing Systems (NIPS) 29*, 2016, pp. 424–432.

[8] L. Miolane, "Fundamental limits of low-rank matrix estimation: The non-symmetric case," *ArXiv e-prints*, Feb. 2017.

[9] T. Lesieur, L. Miolane, M. Lelarge, F. Krzakala, and L. Zdeborová, "Statistical and computational phase transitions in spiked tensor estimation," *ArXiv e-prints*, Jan. 2017.

[10] M. Lelarge and L. Miolane, "Fundamental limits of symmetric low-rank matrix estimation," *ArXiv e-prints*, Nov. 2016.

[11] J. Barbier, N. Macris, and L. Miolane, “The Layered Structure of Tensor Estimation and its Mutual Information,” in *47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2017.

[12] F. Guerra and F. L. Toninelli, “The thermodynamic limit in mean field spin glass models,” *Communications in Mathematical Physics*, vol. 230, no. 1, pp. 71–79, 2002.

[13] J. Barbier and N. Macris, “The stochastic interpolation method: A simple scheme to prove replica formulas in bayesian inference,” *CoRR*, vol. abs/1705.02780, 2017. [Online]. Available: http://arxiv.org/abs/1705.02780

[14] S. Ghirlanda and F. Guerra, “General properties of overlap probability distributions in disordered spin systems. towards parisi ultrametricity,” *Journal of Physics A: Mathematical and General*, vol. 31, no. 46, p. 9149, 1998. [Online]. Available: http://stacks.iop.org/0305-4470/31/i=46/a=006

[15] S. Franz and M. Leone, “Replica bounds for optimization problems and diluted spin systems,” *Journal of Statistical Physics*, vol. 111, no. 3, pp. 535–564, 2003.

[16] A. Giurgiu, N. Macris, and R. Urbanke, “Spatial coupling as a proof technique and three applications,” *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5281–5295, Oct 2016.

[17] M. Aizenman, R. Sims, and S. L. Starr, “Extended variational principle for the Sherrington-Kirkpatrick spin-glass model,” *Physical Review B*, vol. 68, no. 21, p. 214403, 2003.

[18] A. Coja-Oghlan, F. Krzakala, W. Perkins, and L. Zdeborova, “Information-theoretic thresholds from the cavity method,” in *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, ser. STOC 2017. New York, NY, USA: ACM, 2017, pp. 146–157. [Online]. Available: http://doi.acm.org/10.1145/3055399.3055420

[19] E. Abbe, A. S. Bandeira, A. Bracher, and A. Singer, “Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery,” *CoRR*, vol. abs/1404.4749, 2014. [Online]. Available: http://arxiv.org/abs/1404.4749

[20] S. Kudekar, “Statistical physics methods for sparse graph codes,” Ph.D. dissertation, IC, Lausanne, 2009.

[21] S. Kudekar and N. Macris, “Sharp bounds for optimal decoding of low-density parity-check codes,” *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4635–4650, Oct 2009.

[22] S. Kumar, A. J. Young, N. Macris, and H. D. Pfister, “Threshold saturation for spatially coupled ldpc and ldgm codes on bms channels,” *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7389–7415, Dec 2014.

[23] N. Macris, “Griffith-Kelly-Sherman correlation inequalities: A useful tool in the theory of error correcting codes,” *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 664–683, Feb 2007.

[24] T. Richardson and R. Urbanke, *Modern Coding Theory*. New York, NY, USA: Cambridge University Press, 2008.

[25] R. B. Griffiths, *Phase transitions and critical phenomena volume 1*, ser. Phase Transitions and Critical Phenomena, C. Domb and M. S. Green, Eds. Academic Press, 1972, vol. 1.

[26] A. Montanari, “Tight bounds for ldpc and ldgm codes under map decoding,” *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3221–3246, Sept 2005.

[27] N. Macris, “Griffith–kelly–sherman correlation inequalities: A useful tool in the theory of error correcting codes,” *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 664–683, Feb 2007.

[28] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. New York, NY, USA: Cambridge University Press, 2005.