Tight Bounds on the Capacity of Binary Input Random CDMA Systems

Satish Babu Korada and Nicolas Macris

Abstract—**In this paper, we consider code-division multiple-access (CDMA) communication over a binary input additive white Gaussian noise (AWGN) channel using random spreading. For a general class of symmetric distributions for spreading sequences, in the limit of a large number of users, we prove an upper bound to the capacity. The bound matches the formula obtained by Tanaka using the replica method. We also show concentration of various relevant quantities including mutual information and free energy. The mathematical methods are quite general and allow us to discuss extensions to other multiuser scenarios.**

Index Terms—**Binary input, capacity, code-division multiple-access (CDMA), interpolation method, random spreading.**

I. Introduction

CODE-DIVISION MULTIPLE-ACCESS (CDMA) has been a successful scheme for reliable communication between multiple users and a common receiver. The scheme consists of $\mathtt{K}$ users modulating their information sequence by a signature sequence (spreading sequence) of length $\mathtt{N}$ and transmitting the resulting signal. The number $\mathtt{N}$ is sometimes referred to as the spreading gain. The receiver obtains the sum of all the transmitted signals and the noise which is assumed to be white and Gaussian [additive white Gaussian noise (AWGN)].

The capacity region (for real-valued inputs) with input power constraints and optimal decoding has been given in [1]. There it is shown that the achievable rates depend only on the correlation matrix of the spreading sequences. If the spreading sequences are not orthogonal then the complexity of optimum detectors scales exponentially with the number of users. Therefore, in practice, low-complexity receivers like matched filter and linear minimum mean square error (MMSE) estimators are considered [2].

Random spreading sequences were initially considered in [3] and [4]. The authors analyzed the spectral efficiency[1] for optimal decoding in the *large-system limit* $\left(\mathtt{K}\to\infty, \mathtt{N}\to\infty, \frac{\mathtt{K}}{\mathtt{N}}=\beta\right)$ using spectrum of large random matrices. The random spreading allows to obtain nice analytic formulas for various quantities like sum capacity, which provides qualitative insights to the problem. For some scenarios in practice, it is reasonable to assume random spreading. It models the scenario where the spreading sequences are pseudonoise sequences having length much larger than the symbol intervals. Then, the spreading sequence corresponding to each symbol interval behaves as a randomly chosen sequence. Random spreading also models the scenario where the signal is distorted by channel fading. In [5] and [6], the authors analyzed the spectral efficiency for low-complexity detectors like matched filter and MMSE detector. In the large-system limit, they obtained analytical formulas for the spectral efficiency and showed that it concentrates with respect to the randomness in the spreading sequences. These claims follow from known results for the spectrum of large random covariance matrices. We can say that the system is reasonably well understood for Gaussian inputs.

In practice, however, the input of the user is restricted to a constellation, e.g., pulse amplitude modulation (PAM) and quadrature amplitude modulation (QAM). Not much is known in this case. A notable exception is the spectral efficiency for binary input in the case of high signal-to-noise ratio (SNR) [7]. The main reason for this disparity in understanding is that the random matrix tools which played a central role in the analysis for Gaussian inputs are not applicable here.

The work of Tanaka [8] is a breakthrough in the analysis of binary input CDMA system. Using the nonrigorous "replica method," developed for analyzing random spin systems in statistical mechanics, Tanaka computed the spectral efficiency for both optimal and suboptimal detectors and also computed the bit error rate (BER) for uncoded transmission for optimal as well as low-complexity suboptimal detectors. The analysis is extended in [9] to include the case of unequal powers and other constellations. The replica method, though nonrigorous, is believed to yield exact results for some models in statistical mechanics [10], known an mean-field models. The CDMA system can be viewed as one such model and hence the conjectures are believed to be true. Some evidence to this belief is given by Montanari and Tse [11]. Using tools developed for sparse graph codes, they compute the spectral efficiency for a range $\beta < \beta_s$. Here β_s is the maximal value of β such that no phase transition, as a function of SNR, occurs (see Section I-C). The formulas they obtain in this range match with the conjectured formula of Tanaka.

Our main contributions in this paper are twofold. First, we prove that Tanaka's formula is an upper bound to the capacity for all values of the parameters β and SNR (Theorem 6). Second, we prove various useful concentration theorems in the large-system limit. We already know from [11] that this upper

Manuscript received March 10, 2008; revised June 08, 2009. Date of current version October 20, 2010. This work was supported in part by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under Grant 5005-67322.

S. B. Korada was with the EPFL, School of Computer and Communication Sciences, EPFL, Lausanne CH-1015, Switzerland. He is now with Stanford University, Stanford, CA 94305 USA (e-mail: satishbabu.k@gmail.com).

N. Macris is with the School of Computer and Communication Sciences, EPFL, Lausanne CH-1015, Switzerland (e-mail: nicolas.macris@epfl.ch).

Communicated by H. Boche, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2010.2070131

[1]Spectral efficiency is defined as the number of bits per chip that can be reliably transmitted.

bound is an equality when $\beta < \beta_s$. As explained in Section I-C, combining our upper bound with the method of [11], we can extend the proof of the equality to $\beta \geq \beta_s$ at least for SNR larger than the critical value where the phase transition occurs. These remarks, together with the heuristic arguments from the replica method, convincingly suggest that our bound is tight in the sense that it should in fact be an equality for all values of β and SNR.

A. *Statistical Mechanics Approach*

There is a natural connection between various communication systems and statistical mechanics of random spin systems, stemming from the fact that often in both systems there are a large number of degrees of freedom (bits or spins), interacting in a random environment. So far, there have been applications of two important but somewhat complementary approaches of statistical mechanics of random systems.

The first one is the very important but mathematically nonrigorous replica method. The merit of this approach is that it allows one to obtain explicit formulas for quantities of interest such as, conditional entropy, or error probability. The replica method has been applied to many scenarios in communication including channel and source coding using sparse graph codes, multiuser settings like broadcast channel (see, for example, [12]–[14]) and the case of interest here [8]: randomly spread CDMA with binary inputs.

The second type of approach aims at a rigorous understanding of the replica formulas and has its origins in methods stemming from mathematical physics. For systems whose underlying degrees of freedom have Gaussian distribution (Gaussian input symbols or Gaussian spins in continuous spin systems) random matrix methods can successfully be employed. However, when the degrees of freedom are binary (binary information symbols or Ising spins) these seem to fail. Fortunately, in the latter case, the recently developed interpolation method by Guerra [15]–[18] has had a lot of success.[2] The basic idea of the interpolation method is to study a measure which interpolates between the posterior measure of the ideal decoder and a mean-field measure. The latter can be guessed from the replica formulas and from this perspective the replica method is a valuable tool. So far this program has been developed only for linear error correcting codes on sparse graphs and binary input symmetric channels [19], [20].

In this paper, we develop the interpolation method for the randomly spread CDMA system with binary inputs (in the large-system limit). The situation is qualitatively different than the ones mentioned above in that the "underlying graph" is complete. Superficially, one might think that it is similar to the Sherrington–Kirkpatrick model which was the first one treated by the interpolation method. However, as we will see, the analysis of the randomly spread CDMA system is substantially different due to the structure of the interaction between the various degrees of freedom.

[2]Let us point out that, as will be shown later in this paper, the interpolation method can also serve as an alternative to random matrix theory for Gaussian inputs.

B. *Communication Setup*

In the following, we use uppercase letters, e.g., X and Y, to denote random variables and their lowercase counterparts, e.g., x and y to denote the realizations.

The system consists of $\mathtt{K}$ users sending binary information symbols $x_k \in \{\pm 1\}$ to a common receiver. The user k has a random signature sequence $\bar{s}_k = (s_{1k}, \ldots, s_{\mathtt{N}k})^\top$, where s_{ik} is the realization of a random variable S_{ik}. The random variables S_{ik} are independent and identically distributed (i.i.d.) as $S_{ik} \sim p_S$. The random variables S_{ik} are assumed to be symmetric, i.e., $p_S(s) = p_S(-s)$. For each time division (or chip) interval $i = 1, \ldots, \mathtt{N}$, the received signal y_i is given by

$$y_i = \frac{1}{\sqrt{\mathtt{N}}} \sum_{k=1}^{\mathtt{K}} s_{ik} x_k + \sigma n_i$$

where n_i are i.i.d. realizations of $\mathcal{N}(0,1)$. Therefore, the noise power is σ^2. The variance of S is assumed to be 1 and the scaling factor $1/\sqrt{\mathtt{N}}$ is introduced so that the power (per symbol) of each user is normalized to 1.

We write $\mathbf{s}$ for the $\mathtt{N} \times \mathtt{K}$ matrix (s_{ik}) and $\mathbf{S}$ for the corresponding random matrix. We use $\bar{x}$ to denote the vector $(x_1, \ldots, x_{\mathtt{K}})^\top$ and $\bar{X}$ to denote the vector of random variables $(X_1, \ldots, X_{\mathtt{K}})^\top$. Similarly, $\bar{y}$ and $\bar{Y}$ denote the $\mathtt{N}$-dimensional vectors $(y_1, \ldots, y_{\mathtt{N}})^\top$ and $(Y_1, \ldots, Y_{\mathtt{N}})^\top$, respectively. The quantity of interest is

$$C_{\mathtt{K}} = \frac{1}{\mathtt{K}} \max_{\prod_{k=1}^{\mathtt{K}} p_k(x_k)} I(\bar{X}; \bar{Y} | \mathbf{S}) \tag{1}$$

in the large-system limit, i.e., $\mathtt{K} \to +\infty$ with $\frac{\mathtt{K}}{\mathtt{N}} = \beta$ fixed. The maximization is over $p_k(x) = \mu_k \delta(x-1) + (1-\mu_k)\delta(x+1)$ and $\mu_k \in [0,1]$, $k = 1, \ldots, \mathtt{K}$. We refer to $C_{\mathtt{K}}$ as the capacity of the CDMA system. The spectral efficiency is related to the capacity as $\beta C_{\mathtt{K}}$.

We will now show that even if the maximization is done over joint distributions $p_{\bar{X}}$, the maximum is attained for a uniform distribution. For any realization of $\mathbf{S} = \mathbf{s}$, the mutual information $I(\bar{X}; \bar{Y} | \mathbf{S} = \mathbf{s})$ is a concave functional of $p_{\bar{X}}$ and thus so is its expected value. Moreover, the latter is invariant under the transformations $p_{\bar{X}}(x_1, x_2, \ldots, x_{\mathtt{K}}) \rightarrow p_{\bar{X}}(\epsilon_1 x_1, \epsilon_2 x_2, \ldots, \epsilon_{\mathtt{K}} x_{\mathtt{K}})$ where $\epsilon_k = \pm 1$. Combining these two facts, we deduce that the maximum in (1) is attained for the convex combination

$$\frac{1}{2^{\mathtt{K}}} \sum_{\epsilon_1, \ldots, \epsilon_{\mathtt{K}}} p_{\bar{X}}(\epsilon_1 x_1, \ldots, \epsilon_{\mathtt{K}} x_{\mathtt{K}}) = \frac{1}{2^{\mathtt{K}}}$$

which is equal to the uniform distribution.

In the next few paragraphs, we discuss various settings for which it is justified to consider (1) as capacity. Following [5], let us consider the case of "long spreading sequences," that is, sequences that extend over many symbol durations. If the spreading sequences are not used at the encoder, i.e., if the input distribution is not allowed to vary with the spreading sequences, then the capacity is given by (1). However, if the input distribution is allowed to depend on the spreading sequences, then the capacity is given by

$$\frac{1}{\mathtt{K}} \max_{\prod_{k=1}^{\mathtt{K}} p_k(x_k|\mathbf{S})} I(\bar{X};\bar{Y}|\mathbf{S}). \tag{2}$$

This requires that every user have the knowledge of the spreading sequences of all the users and have different code books for each spreading sequence realization. Moreover, for the continuous input case [5], it was shown that the gain in capacity due to "dynamic" power allocation is negligible. Therefore, it is reasonable to assume that the input distribution is independent of the spreading sequences and hence the capacity is given by (1).

In the traditional CDMA setting (see, e.g., [2]), the spreading sequences are assigned to each user and do not change from symbol to symbol. Let $\mathbf{s}$ be the spreading matrix. Then, the total capacity per user (or maximal achievable sum rate per user) is given by

$$\frac{1}{\mathtt{K}} \max_{\prod_{k=1}^{\mathtt{K}} p_k(x_k)} I(\bar{X};\bar{Y}|\mathbf{S}=\mathbf{s}). \tag{3}$$

However, if the spreading sequences that are used are generated randomly, then our concentration results state that for any input distribution $p_{\bar{X}}$, $I(\bar{X};\bar{Y}|\mathbf{S}=\mathbf{s})$ concentrates around its average $I(\bar{X};\bar{Y}|\mathbf{S})$. Hence, the capacity $C_{\mathtt{K}}$ plays a crucial role in this scenario also.

At this point, it is interesting to discuss the situation for the continuous input case. There it is known that the maximum of (3) is attained for a Gaussian input distribution independent of the spreading sequence realization [1]. Then, the concentration theorems suffice to prove that in the large-system limit, (3) asymptotically equals (1). It is an open problem to decide if an analogous result holds in the binary input case. More precisely, is it true that the maximum of (3) is attained for the uniform input distribution irrespective of the spreading sequence realization?

Let us mention that (1) can also be interpreted as the capacity of a multiple-input–multiple-output (MIMO) system with $\mathtt{K}$ transmit antennas and $\mathtt{N}$ receive antennas. The matrix $\mathbf{s}$ models the fading coefficients of the $\mathtt{KN}$ paths between the transmitter and the receiver and are known only to the receiver [21], [22]. In this case, the information bits across various transmit antennas can be correlated, because all the antennas are located at the same sender. Therefore, $p_{\bar{X}}$ need not be a product distribution, which implies that the capacity per antenna is given by

$$\frac{1}{\mathtt{K}} \max_{p_{\bar{X}}} I(\bar{X};\bar{Y}|\mathbf{S}). \tag{4}$$

In our previous discussion, we have seen that the maximum is attained for the uniform distribution and hence (4) is equal to (1). Therefore, our results about the capacity of the CDMA system have similar implications to the MIMO channel.

Let us now collect a few formulas that will be useful in the rest of the paper. Let $p_{\bar{X}}(\bar{x}) = \prod_k p_k(x_k)$ denote the input distribution. For any realization of spreading sequences $\mathbf{s}$, and output $\bar{y}$, the posterior probability distribution is given by

$$p(\bar{x} \mid \bar{y}, \mathbf{s}) = \frac{p_{\bar{X}}(\bar{x})}{Z(\bar{y},\mathbf{s})} e^{-\frac{1}{2\sigma^2}\|\bar{y}-\mathtt{N}^{-\frac{1}{2}}\mathbf{s}\bar{x}\|^2} \tag{5}$$

with the normalization factor

$$Z(\bar{y},\mathbf{s}) = \sum_{\bar{x}} p_{\bar{X}}(\bar{x}) e^{-\frac{1}{2\sigma^2}\|\bar{y}-\mathtt{N}^{-\frac{1}{2}}\mathbf{s}\bar{x}\|^2}. \tag{6}$$

The distribution of the output $\bar{Y}$ is given by

$$\begin{aligned} p(\bar{y} \mid \mathbf{s}) &= \sum_{\bar{x}^0} p_X(\bar{x}^0) \frac{e^{-\frac{1}{2\sigma^2}\|\bar{y}-\mathtt{N}^{-\frac{1}{2}}\mathbf{s}\bar{x}^0\|^2}}{(\sqrt{2\pi}\sigma)^{\mathtt{N}}} \\ &= \frac{Z(\bar{y},\mathbf{s})}{(\sqrt{2\pi}\sigma)^{\mathtt{N}}} \end{aligned} \tag{7}$$

where $\bar{x}^0$ is interpreted as the input information vector.

In the language of statistical mechanics, the bits x_i play the role of Ising spins. The Hamiltonian for the state $\bar{x}$ is given by $\mathtt{H}(\bar{x}) = -\frac{1}{2\sigma^2}\|\bar{y}-\mathtt{N}^{-\frac{1}{2}}\mathbf{s}\bar{x}\|^2$.[3] The normalization factor (6) can be interpreted as the partition function. In view of this, it is not surprising that the free energy defined by

$$f(\bar{y},\mathbf{s}) = \frac{1}{\mathtt{K}} \ln Z(\bar{y},\mathbf{s}) \tag{8}$$

plays a crucial role. In fact, the free energy is related to the mutual information as

$$\begin{aligned} \frac{1}{\mathtt{K}} I(\bar{X};\bar{Y}|\mathbf{S}=\mathbf{s}) &= H(\bar{Y}|\mathbf{S}=\mathbf{s}) - H(\bar{Y} \mid \bar{X},\mathbf{S}=\mathbf{s}) \\ &\overset{(a)}{=} -\int p(\bar{y} \mid \mathbf{s}) \ln p(\bar{y} \mid \mathbf{s}) \\ &\quad - \frac{1}{2\beta}\ln(2\pi e\sigma^2) \\ &\overset{(b)}{=} -\frac{1}{2\beta} - \mathbb{E}_{\bar{Y}}[f(\bar{Y},\mathbf{s})]. \end{aligned} \tag{9}$$

The equality (a) follows from the fact that $H(\bar{Y}|\bar{X},\mathbf{S}=\mathbf{s})$ is the entropy of an $\mathtt{N}$-dimensional i.i.d. Gaussian noise vector with variance σ^2. The equality (b) follows by using (7) for $p(\bar{y} \mid \bar{x},\mathbf{s})$. Therefore

$$C_{\mathtt{K}} = -\frac{1}{2\beta} - \min_{p_{\bar{X}}} \mathbb{E}_{\bar{Y},\mathbf{S}}[f(\bar{Y},\mathbf{S})]. \tag{10}$$

Of course, from our previous discussion, we know that the minimum is attained at $p_{\bar{X}}(\bar{x}) = \frac{1}{2^{\mathtt{K}}}$.

C. Tanaka's Formula for Binary Inputs

In this section, let us restrict the input distribution to be uniform, i.e., $p_{\bar{X}}(\bar{x}) = \frac{1}{2^{\mathtt{K}}}$. Let the spreading sequence distribution be symmetric with finite fourth moment. Using the formal

[3]In statistical mechanics, the convention is to denote the Hamiltonian as $\mathtt{H}(\bar{x}) = \frac{1}{2\sigma^2}\|\bar{y}-\mathtt{N}^{-\frac{1}{2}}\mathbf{s}\bar{x}\|^2$. Here, for convenience, we include the negative sign also in the Hamiltonian.

replica method, Tanaka conjectured that the capacity of the CDMA system is given by

$$\lim_{\mathtt{K}\to\infty} C_{\mathtt{K}} = \min_{m\in[0,1]} c_{RS}(m) \tag{11}$$

where the "replica symmetric capacity functional" is given by

$$c_{RS}(m) = \frac{\lambda}{2}(1+m) - \frac{1}{2\beta}\ln\lambda\sigma^2 - \int Dz\ln(2\cosh(\sqrt{\lambda}z+\lambda)). \tag{12}$$

The parameter λ is defined by

$$\lambda = \frac{1}{\sigma^2+\beta(1-m)} \tag{13}$$

through a parameter[4] m and Dz is the standard Gaussian measure $Dz \equiv \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}dz$. The minimizer of (11) either belongs to $\{0,1\}$ or it must satisfy the fixed point condition

$$m = \int Dz\tanh(\sqrt{\lambda}z+\lambda) \tag{14}$$

which is obtained from $\frac{d}{dm}c_{RS}(m)=0$. To see this let us compute the derivative as follows:

$$\frac{d}{dm}c_{RS}(m) = \frac{\lambda'}{2}(1+m) - \int Dz\tanh(\sqrt{\lambda}z+\lambda)\Big(\frac{z}{2\sqrt{\lambda}}+1\Big)\lambda'. \tag{15}$$

We now use the following integration by parts formula for Gaussian random variables:

$$\int Dz f(z)z = \int Dz f'(z) \tag{16}$$

which is valid for any continuous function f. Applying the above formula to the integral in (15), we get

$$\int Dz\tanh(\sqrt{\lambda}z+\lambda)\frac{z}{2\sqrt{\lambda}} = \frac{1}{2}\int Dz\,\mathrm{sech}^2(\sqrt{\lambda}z+\lambda).$$

The proof follows by using $\mathrm{sech}^2(x) = 1-\tanh^2(x)$ and

$$\begin{aligned}\int Dz\tanh^2(\sqrt{\lambda}z+\lambda) &= \int dz\frac{(e^z-e^{-z})^2}{(e^z+e^{-z})^2}e^{-\frac{z^2}{2\lambda}-\frac{\lambda}{2}+z}\\ &= \int_0^\infty dz\frac{(e^z-e^{-z})^2}{(e^z+e^{-z})^2}e^{-\frac{z^2}{2\lambda}-\frac{\lambda}{2}}(e^z+e^{-z})\\ &= \int_0^\infty dz\frac{(e^z-e^{-z})^2}{(e^z+e^{-z})}e^{-\frac{z^2}{2\lambda}-\frac{\lambda}{2}}\\ &= \int Dz\tanh(\sqrt{\lambda}z+\lambda).\end{aligned}$$

In the present problem, from statistical physics semiheuristic arguments, one expects *a priori* that replica symmetry is not broken because of a gauge symmetry [23, ch. 4], [24] induced by channel symmetry. For this reason, Tanaka's formula is conjectured to be exact. Our upper bound (Theorem 6) on the capacity precisely coincides with the above formulas and strongly supports this conjecture.

[4]This parameter can be interpreted as the expected value of the MMSE estimate for the information bits

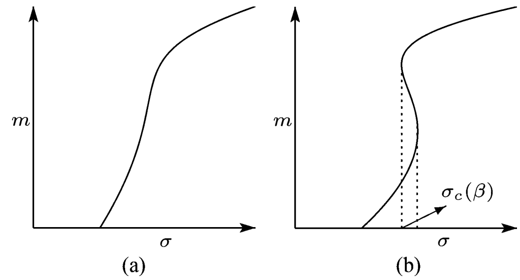

Fig. 1. An illustration of the behavior of solutions of (14). (a) Behavior for $\beta < \beta_s$ (no phase transition), where (14) has only one solution for every σ. (b) $\beta > \beta_s$ (phase transition), where (14) has multiple solutions between the two dashed lines.

The following discussion provides even stronger arguments for the general validity of Tanaka's formula.

The work of Montanari and Tse [11] provides strong support to the conjecture at least in a regime of β without phase transitions (more precisely, for $\beta < \beta_s$ where β_s is the maximal value of β such that the solution of (14) remains unique). Fig. 1 illustrates the behavior of solutions of (14) for the two regimes. The authors first solve the case of sparse signature sequence (using the area theorem and the data processing inequality) in the limit $\mathtt{K}\to\infty$. Then, the dense signature sequence (which is of interest here) is recovered by exchanging the $\mathtt{K}\to\infty$ and *sparse* $\to$ *dense* limits. In this way, they obtain Tanaka's formula for $\beta < \beta_s$.

Let us now consider $\beta \geq \beta_s$. For this regime, one finds a phase transition for a critical value $\sigma_c(\beta)$ (e.g., Fig. 1). As explained above [11], using the data processing inequality, we obtain a bound on the derivative of the mutual information with respect to σ. Integrating this bound from 0 to $\sigma < \sigma_c(\beta)$ yields a lower bound for the capacity which matches our upper bound (Theorem 6).

D. Gaussian Inputs

In the case of continuous inputs $x_k \in \mathbb{R}$, in (6) and (7), $\sum_{\bar{x}}$ are replaced by $\int d\bar{x}$. The capacity is maximized by a Gaussian prior

$$p_{\bar{X}}(\bar{x}) = \frac{e^{-\frac{\|\bar{x}\|^2}{2}}}{(2\pi)^{\mathtt{N}/2}} \tag{17}$$

and one can express it in terms of a determinant involving the correlation matrix of the spreading sequences. Using the exact spectral measure given by random matrix theory, Shamai and Verdu [5] obtained the rigorous result

$$\lim_{\mathtt{K}\to\infty} C_{\mathtt{K}} = \frac{1}{2}\log\left(1+\sigma^{-2}-\frac{1}{4}Q(\sigma^{-2},\beta)\right) + \frac{1}{2\beta}\log\left(1+\sigma^{-2}\beta-\frac{1}{4}Q(\sigma^{-2},\beta)\right) - \frac{Q(\sigma^{-2},\beta)}{8\beta\sigma^{-2}} \tag{18}$$

where

$$Q(x,z) = \left(\sqrt{x(1+\sqrt{z})^2+1}-\sqrt{x(1-\sqrt{z})^2+1}\right)^2.$$

On the other hand, Tanaka applied the formal replica method to this case and found (11) with

$$c_{RS}(m) = \frac{1}{2}\log(1+\lambda) - \frac{1}{2\beta}\log\lambda\sigma^2 - \frac{\lambda}{2}(1-m) \quad (19)$$

where $\lambda = (\sigma^2 + \beta(1-m))^{-1}$. The minimizer satisfies

$$m = \frac{\lambda}{1+\lambda}. \quad (20)$$

Solving (20), we obtain $m = \frac{\sigma^2}{4\beta}Q(\sigma^{-2}, \beta)$ and substituting this in (19) gives the equality between (18) and (19). So at least for the case of Gaussian inputs we are already assured that the replica method finds the correct solution.

As we will show in Section VII-C, our methods also work in the case of Gaussian inputs and yield the upper bound.

E. Contributions and Organization of This Work

The main focus and challenge of this work is on the case of binary inputs for the communication setup described above, although the methods also work for many other constellations including Gaussian inputs. The main results are explained in Section II while the remaining sections are devoted to the proofs.

We prove concentration of the mutual information in the limit of $\mathtt{K} \to +\infty$ and $\beta = \frac{\mathtt{K}}{\mathtt{N}}$ fixed (Theorems 1 and 3 in Section II-A). As we will see, the mathematical underpinning of this is the concentration of a more fundamental object, namely, the "free energy" of the associated spin system (Theorem 2). In fact, this turns out to be important in the proof of the bound on capacity. When the spreading coefficients are Gaussian, the main tool used is a powerful theorem [10, Th. 2.2.4], for the concentration of Lipschitz functions of many independent Gaussian variables, and this leads to subexponential concentration bounds. For more general spreading sequence distributions such tools do not suffice and we have to combine them with martingale arguments which lead to weaker algebraic bounds. Since the concentration proofs are mainly technical they are presented in Appendixes I-A and I-B.

Sections III and IV form the core of the paper. They detail the proof of the main Theorem 6 announced in Section II-D, namely the tight upper bound on capacity. We use ideas from the interpolation method combined with a nontrivial concentration theorem for the empirical average of soft bit estimates.

Section V shows that the average capacity is independent of the spreading sequence distribution at least for the case where it is symmetric and decays fast enough (Theorem 4 in Section II-B). This enables us to restrict ourselves to the case of Gaussian spreading sequences which is more amenable to analysis. The existence of the limit $\mathtt{K} \to \infty$ for the capacity is shown in Section VI.

Section VII discusses various extensions of this work. We sketch the treatment for unequal powers for each user as well as colored noise. As alluded to before the bound on capacity for the case of Gaussian inputs can also be obtained by the present method and we give some indications to this effect.

Appendixes I–IV contain the proofs of various technical calculations. Preliminary versions of the results obtained in this paper have been summarized in [25] and [26].

II. Main Results

In this section, we discuss the main results of this paper. For compactness, let us introduce the following notation. Let $\imath(\bar{X};\bar{Y}|\mathbf{s}) = I(\bar{X};\bar{Y}|\mathbf{S}=\mathbf{s})$. We will treat $\imath(\bar{X};\bar{Y}|\mathbf{s})$ as a function of $\mathbf{s}$. Then $\imath(\bar{X};\bar{Y}|\mathbf{S})$ is a random variable depending on $\mathbf{S}$ whose expectation is given by $\mathbb{E}_{\mathbf{S}}[\imath(\bar{X};\bar{Y}|\mathbf{S})] = I(\bar{X};\bar{Y}|\mathbf{S})$.

A. Concentration

In the case of a Gaussian input signal, the concentration can be deduced from the results on the concentration of the spectral density for large random matrices. But this approach breaks down for binary inputs.

Theorem 1 (Concentration of Capacity, Gaussian Spreading Sequences): Consider a CDMA system with binary inputs and let the spreading sequence distribution be the standard Gaussian distribution. Given $\epsilon > 0$, there exists an integer $\mathtt{K}_1 = O(|\ln\epsilon|)$ independent of $p_{\bar{X}}$, such that for all $\mathtt{K} > \mathtt{K}_1$

$$\mathbb{P}[|\imath(\bar{X};\bar{Y}|\mathbf{S}) - \mathbb{E}_{\mathbf{s}}[\imath(\bar{X};\bar{Y}|\mathbf{S})]| \geq \epsilon\mathtt{K}] \leq 3e^{-\alpha_1\mathtt{K}}$$

where $\alpha_1(\beta,\sigma,\epsilon) > 0$ and is independent of $\mathtt{K}$.

The constant α_1 is shown explicitly in the proof. The mathematical underpinning of this result is in fact a more general concentration result for the free energy (8) that will be of some use latter on.

Theorem 2 (Concentration of Free Energy, Gaussian Spreading Sequences): Consider a CDMA system with binary inputs and let the spreading sequence distribution be the standard Gaussian distribution. Given $\epsilon > 0$, there exists an integer $\mathtt{K}_2 = O(|\ln\epsilon|)$ independent of $p_{\bar{X}}$, such that for all $\mathtt{K} \geq \mathtt{K}_2$

$$\mathbb{P}[|f(\bar{Y},\mathbf{S}) - \mathbb{E}_{\bar{Y},\mathbf{S}}[f(\bar{Y},\mathbf{S})]| \geq \epsilon] \leq 3e^{-\alpha_2\sqrt{\mathtt{K}}}$$

where $\alpha_2(\beta,\sigma,\epsilon) > 0$ and is independent of $\mathtt{K}$.

We prove these theorems thanks to powerful probabilistic tools developed by Ledoux and Talagrand for Lipschitz functions of many Gaussian random variables. These tools are briefly reviewed in Appendix I-A for the convenience of the reader and the proofs of the theorems are presented in Appendix I-B. Unfortunately, the same tools do not apply directly to the case of other spreading sequences. However, in this case, the following weaker result can be obtained.

Theorem 3 (General Spreading Sequences): Consider a CDMA system with binary inputs and let the spreading sequence distribution be symmetric with finite fourth moment. There exists an integer $\mathtt{K}_1$ independent of $p_{\bar{X}}$, such that for all $\mathtt{K} > \mathtt{K}_1$

$$\mathbb{P}[|\imath(\bar{X};\bar{Y}|\mathbf{S}) - \mathbb{E}_{\mathbf{S}}[\imath(\bar{X};\bar{Y}|\mathbf{S})]| \geq \epsilon\mathtt{K}] \leq \frac{\alpha}{\mathtt{K}\epsilon^2}$$
$$\mathbb{P}[f(\bar{Y},\mathbf{S}) - \mathbb{E}_{\bar{Y},\mathbf{S}}[f(\bar{Y},\mathbf{S})]| \geq \epsilon] \leq \frac{\alpha}{\mathtt{K}\epsilon^2}$$

where $\alpha(\beta,\sigma) > 0$ and is independent of $\mathtt{K}$.

To prove such estimates it is enough (by Chebycheff) to control second moments. For the mutual information, we simply have to adapt martingale arguments of Pastur and Shcherbina [27] and Shcherbina and Tirozzi [28] whereas the case of free energy is more complicated because of the additional randomness due to the Gaussian noise. We deal with these by combining martingale arguments and Lipschitz function techniques.

We wish to argue here that Theorem 2 suggests a method for proving the concentration of the BER for uncoded communication [8] given by

$$\frac{1}{2}\left(1-\frac{1}{\mathtt{K}}\sum_{k=1}^{\mathtt{K}}x_k^0\hat{x}_k\right). \tag{21}$$

The maximum *a posteriori* (MAP) bit estimate $\hat{x}_k$ is defined through the marginal of (5) as $\hat{x}_k = \text{argmax}_{x_k=\{\pm 1\}}p(x_k \mid \bar{y}, \mathbf{s})$. Note that

$$\hat{x}_k = \text{sign}\langle x_k\rangle$$

where we find it convenient to adopt the statistical mechanics notation $\langle - \rangle$ for the average with respect to the posterior measure (5). For example, the average

$$\langle x_k\rangle = \sum_{\bar{x}} x_k p(\bar{x} \mid \bar{y}, \mathbf{s})$$

(a soft bit estimate or "magnetization") can be obtained from the free energy by adding first an infinitesimal perturbation ("small external magnetic field") to the exponent in (5), namely $h\sum_{k=1}^{\mathtt{K}} x_k^0 x_k$, and then differentiating the perturbed free energy[5]

$$\frac{1}{\mathtt{K}}\sum_{k=1}^{\mathtt{K}} x_k^0\langle x_k\rangle = \lim_{h\to 0}\frac{d}{dh}\frac{1}{\mathtt{K}}\ln Z(\bar{y},\mathbf{s}).$$

However, one really needs to relate $\text{sign}\langle x_k\rangle$ to the derivative of the free energy and this does not appear to be obvious. One way out is to introduce product measure of n copies (also called "real replicas") of the posterior measure

$$p\left(\bar{x}^{(1)} \mid \bar{y},\mathbf{s}\right)p\left(\bar{x}^{(2)} \mid \bar{y},\mathbf{s}\right)\dots p\left(\bar{x}^{(n)} \mid \bar{y},\mathbf{s}\right)$$

and then relate

$$\sum_{k=1}^{\mathtt{K}}\left(x_k^0\left\langle x_k\right\rangle\right)^n = \sum_{k=1}^{\mathtt{K}}\left\langle x_k^0 x_k^{(1)}\cdots x_k^0 x_k^{(n)}\right\rangle_n$$

to a suitable derivative of the replicated free energy. Then, from the set of all moments, one can in principle reconstruct $\text{sign}\langle x_k\rangle$. Thus, one could try to deduce the concentration of the BER from the one for the free energy. However, the completion of this program requires a uniform, with respect the system size, control of the derivative of the free energy precisely at $h=0$, which at the moment is still lacking.[6]

B. Independence With Respect to the Distribution of the Spreading Sequence

The replica method leads to the same formula for the capacity for all symmetric spreading sequence distributions with equal second moment and finite fourth moment. Here we rigorously show a result of that flavor for the following class of distributions.

Class A: The distribution $p_S(s)$ is symmetric, i.e.,

$$p_S(s) = p_S(-s)$$

and has a rapidly decaying tail. More precisely, there exist positive constants s_0 and A such that $\forall s \geq s_0$

$$\Pr(S\geq s)\leq e^{-As^2}.$$

In particular, the Gaussian and binary cases are included in this class, and also any compactly supported distribution. We believe that a better approximation of some of the error terms in our proofs would widen the class of distributions to the one predicted by replica method.

Theorem 4 (Independence of the Capacity): Consider a CDMA system with binary inputs. Let $C_{\mathtt{K}}$ denote the capacity for a spreading sequence distribution belonging to Class A. Let $C_{\mathtt{K}}^g$ denote the capacity for Gaussian spreading sequence distribution having the same second moment. Then

$$\lim_{\mathtt{K}\to+\infty}(C_{\mathtt{K}}-C_{\mathtt{K}}^g)=0.$$

This theorem turns out to be very useful in order to obtain the bound on capacity because it allows us to make use of convenient integration by parts identities that have no clear counterpart in the non-Gaussian case. The proof of the theorem is given in Section V.

C. Existence of the Limit $\mathtt{K}\to+\infty$

The interpolation method can be used to show the existence of the limit $\mathtt{K}\to+\infty$ for $C_{\mathtt{K}}$.

Theorem 5 (Existence of the Limit): Consider a CDMA system with binary inputs and let the spreading sequence distribution belong to Class A. Let $C_{\mathtt{K}}$ denote its capacity. Then

$$\lim_{\mathtt{K}\to\infty} C_{\mathtt{K}} \text{ exists.} \tag{22}$$

The proof of this theorem is given in Section VI for Gaussian spreading sequences. The general case then follows from Theorem 4.

D. Tight Upper Bound on the Capacity

The main result of this paper is that Tanaka's formula (12) is an upper bound to the capacity for all values of β.

Theorem 6 (Upper Bound on the Capacity): Consider a CDMA system with binary inputs and let the spreading sequence distribution belong to Class A. Let $C_{\mathtt{K}}$ denote its capacity. Then

$$\lim_{\mathtt{K}\to\infty} C_{\mathtt{K}} \leq \min_{m\in[0,1]} c_{RS}(m) \tag{23}$$

where $c_{RS}(m)$ is given by (12).

If we combine this result with an inequality in [11], one can deduce that the equality holds for some regime of noise smaller than $\sigma_c(\beta)$, shown in Fig. 1. This value corresponds to the smallest noise variance for which (14) has multiple solutions.

[5]We do not write explicitly the h dependence in the perturbed free energy

[6]However, this can be done for Lebesgue almost every h

Note that this equality is valid for all β, whereas in [11], the equality holds only for $\beta < \beta_s$.

Since the proof is rather complicated, we find it useful to give the main ideas in an informal way. The integral term in (12) suggests that we can replace the original system with a simpler system where the user bits are sent through $\mathtt{K}$ independent Gaussian channels[7] given by

$$y'_k = x_k + \frac{1}{\sqrt{\lambda}} w_k \tag{24}$$

where $w_k \sim \mathcal{N}(0,1)$ and λ is an effective SNR. Of course, this argument is a bit naive because this effective system does not account for the extra terms in (12), but it has the merit of identifying the correct interpolation.

We introduce an interpolating parameter $t \in [0,1]$ such that the independent Gaussian channels correspond to $t=0$ and the original CDMA system corresponds to $t=1$ (see Fig. 2). It is convenient to denote the SNR of the original Gaussian channel as B (that is $B = \sigma^{-2}$). Then, (13) becomes

$$\lambda = \frac{B}{1+\beta B(1-m)}.$$

We introduce two interpolating SNR functions $\lambda(t)$ and $B(t)$ such that

$$\lambda(0) = \lambda, \quad B(0) = 0 \quad \text{and} \quad \lambda(1) = 0,\ B(1) = B \tag{25}$$

and

$$\frac{B(t)}{1+\beta B(t)(1-m)} + \lambda(t) = \frac{B}{1+\beta B(1-m)}. \tag{26}$$

Here the parameter m is the same as in (12). It is to be considered, *a priori*, as fixed to any arbitrary value in $[0,1]$: all the subsequent calculations are valid for any value. In the final step, one optimizes over m in order to tighten the final bound.

The meaning of (26) is the following. In the interpolating t-system, the effective SNR seen by each user has an effective t-CDMA part and an independent channel part $\lambda(t)$ chosen such that the total SNR is fixed to the effective SNR of the CDMA system. There is a whole class of interpolating functions satisfying the above conditions but it turns out that we do not need to specify them more precisely except for the fact that $B(t)$ is increasing and $\lambda(t)$ is decreasing and with continuous first derivatives. Subsequent calculations are independent of the particular choices of functions.

We now have two sets of channel outputs $\bar{y}$ (from the CDMA with noise variance $B(t)^{-1}$) and $\bar{y}'$ (from the independent channels with noise variance $\lambda(t)^{-1}$) and the interpolating communication system has a posterior distribution

$$p_t(\bar{x}|\bar{y},\bar{y}',\mathbf{s}) = \frac{1}{2^{\mathtt{K}} Z(\bar{y},\bar{y}',\mathbf{s})} e^{-\frac{B(t)}{2}\|\bar{y}-\mathtt{N}^{-\frac{1}{2}}\mathbf{s}\bar{x}\|^2 - \frac{\lambda(t)}{2}\|\bar{y}'-\bar{x}\|^2}. \tag{27}$$

Note that here we take without loss of generality $p_{\bar{X}}(\bar{X}) = \frac{1}{2^{\mathtt{K}}}$. By analyzing the mutual information $I_t(\bar{X};\bar{Y},\bar{Y}'|\mathbf{S})$ of the interpolating system we can relate $I(\bar{X};\bar{Y}|\mathbf{S})$ (the $t=1$ value) to the easily computed entropy $I_0(\bar{X};\bar{Y}'|\mathbf{S})$ of the independent channel limit. The average over $(\bar{Y},\bar{Y}')$ is now performed with respect to

$$p_t(\bar{y},\bar{y}' \mid \mathbf{s}) = \frac{1}{2^{\mathtt{K}}} \sum_{\bar{x}^0} \frac{1}{(\sqrt{2\pi B(t)^{-1}})^{\mathtt{N}} (\sqrt{2\pi\lambda(t)^{-1}})^{\mathtt{K}}} \times e^{-\frac{B(t)}{2}\|\bar{y}-\mathtt{N}^{-\frac{1}{2}}\mathbf{s}\bar{x}^0\|^2 - \frac{\lambda(t)}{2}\|\bar{y}'-\bar{x}^0\|^2}. \tag{28}$$

These equations completely define the interpolating communication system.

In order to carry out this program successfully, it turns out that we need a concentration result on empirical average of the "magnetization"

$$m_1 = \frac{1}{\mathtt{K}} \sum_{k=1}^{\mathtt{K}} x_k^0 x_k$$

which, as explained in Section II-A, is closely related to the BER. Informally speaking, we need to prove that the fluctuations of $\mathbb{E}\langle|m_1 - \mathbb{E}\langle m_1\rangle|\rangle$ are small. This involves the control of two types of fluctuations, $\mathbb{E}\langle|m_1 - \langle m_1\rangle|\rangle$ and $\mathbb{E}|\langle m_1\rangle - \mathbb{E}\langle m_1\rangle|$ (by the triangle inequality). The control of these fluctuations is the object of Theorem 7 in Section III-C. There are technical complications that we have to deal with because such control of fluctuations is only possible away from phase transitions. For this reason, we have to add small appropriate perturbations to the measure (27) and give almost sure statements with respect to the strength of the perturbation. By being sufficiently careful with the order of limits the extra perturbation terms can be removed at the end of the calculations.

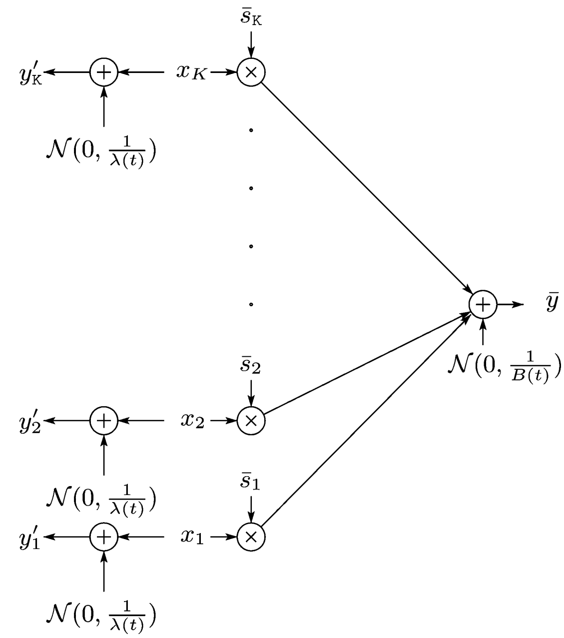

Fig. 2. The information bits x_k are transmitted through the normal CDMA channel with variance $\frac{1}{B(t)}$ and through individual Gaussian channels with noise $\frac{1}{\lambda(t)}$.

[7]These are the single user decoupled channels discussed in [8] and [9]

III. Proof of Bound on Capacity: Theorem 6

A. Preliminaries

The interpolating communication system defined by the measure (27) allows us to compare the original CDMA system with the independent channel system. The distribution of $\bar{y}$, $\bar{y}'$ is given by (28). This distribution consists of a summation of $2^{\mathtt{K}}$ terms, each corresponding to different possible input sequence. Each of these terms contributes equally to the capacity (free energy). The reader can explicitly check this by making the change of variables $x_k \to x_k^0 x_k$, $s_{ik} \to s_{ik} x_k^0$, and $w_k \to w_k x_k^0$, which leave all standard Gaussians invariant. Hence, we can assume that a particular input sequence, say $\bar{x}^0$, is transmitted. The distribution of the received vectors with this assumption is

$$p_t(\bar{y}, \bar{y}' \mid \mathbf{s}) = \frac{1}{(\sqrt{2\pi B(t)^{-1}})^{\mathtt{N}}(\sqrt{2\pi\lambda(t)^{-1}})^{\mathtt{K}}} \times e^{-\frac{B(t)}{2}\|\bar{y}-\mathtt{N}^{-\frac{1}{2}}\mathbf{s}\bar{x}^0\|^2-\frac{\lambda(t)}{2}\|\bar{y}'-\bar{x}^0\|^2}. \quad (29)$$

For technical reasons that will become clear only in the next section, we consider a slightly more general interpolation system where the perturbation term

$$h_u(\bar{x}) = \sqrt{u}\sum_{k=1}^{\mathtt{K}} h_k x_k + u\sum_{k=1}^{\mathtt{K}} x_k^0 x_k - \sqrt{u}\sum_{k=1}^{\mathtt{K}} |h_k| \quad (30)$$

is added in the exponent of the measure (27). Here h_k are realizations of i.i.d. random variables H_k which are distributed as $H_k \sim \mathcal{N}(0,1)$. For the moment $u \geq 0$ is arbitrary, but later, we will take $u \to 0$. The choice of the perturbation term is not unique but this one suits our purpose. The important conditions that we need for the latter proofs is that it preserves the Nishimori (gauge) symmetry and that the perturbed free energy is a convex function of u (see Section IV). This time it is convenient to perform a new change of variables $\bar{y} = B(t)^{-1/2}\bar{n} + \mathtt{N}^{-1/2}\mathbf{s}\bar{x}^0$ and $\bar{y}' = \lambda(t)^{-1/2}\bar{w} + \bar{x}^0$, where $N_i, W_i \sim \mathcal{N}(0,1)$ and we set $\langle - \rangle_{t,u}$ for the average corresponding to the posterior measure

$$p_{t,u}(\bar{x}|\bar{n}, \bar{w}, \bar{h}, \mathbf{s}) = \frac{1}{Z_{t,u}} \times e^{-\frac{1}{2}\|\bar{n}+N^{-\frac{1}{2}}B(t)^{\frac{1}{2}}\mathbf{s}(\bar{x}^0-\bar{x})\|^2-\frac{1}{2}\|\bar{w}+\lambda(t)^{\frac{1}{2}}(\bar{x}^0-\bar{x})\|^2+h_u(\bar{x})} \quad (31)$$

with the obvious normalization factor $Z_{t,u}$. We define the free energy

$$f_{t,u}(\bar{n}, \bar{w}, \bar{h}, \mathbf{s}) = \frac{1}{\mathtt{K}}\ln Z_{t,u}. \quad (32)$$

For $t = 1$, we recover the original free energy

$$\mathbb{E}[f(\bar{Y}, \mathbf{S})] = \frac{1}{2} + \lim_{u\to 0}\mathbb{E}[f_{1,u}(\bar{N}, \bar{W}, \bar{H}, \mathbf{S})]$$

while for $t = 0$ the statistical sums decouple and we have the explicit result[8]

$$\frac{1}{2} + \lim_{u\to 0}\mathbb{E}[f_{0,u}(\bar{N}, \bar{W}, \bar{H}, \mathbf{S})] = -\frac{1}{2\beta} - \lambda + \int Dz \ln(\cosh(\sqrt{\lambda}z + \lambda)) \quad (33)$$

[8]It is also straightforward to compute the full u dependence and see that it is $O(\sqrt{u})$, uniformly in $\mathtt{K}$

where $\mathbb{E}$ denotes the collective expectation over the random variables appearing in the expression. In view of (9), in order to obtain the average capacity, it is sufficient to compute

$$\lim_{\mathtt{K}\to+\infty}\lim_{u\to 0}\mathbb{E}[f_{1,u}(\bar{N}, \bar{W}, \bar{H}, \mathbf{S})] + \frac{1}{2}. \quad (34)$$

Note that to obtain the above formulas, we have exchanged $\lim_{u\to 0}$ and $\mathbb{E}$. This is allowed by dominated convergence since (30), and thus the free energies, are bounded uniformly in u for $0 \leq u \leq 1$. There is no loss in generality in setting

$$x_k^0 = 1 \quad (35)$$

for the input symbols. From now on in Sections III, IV, and VI, we stick to (35). We also use the shorthand notations

$$z_k = x_k^0 - x_k = 1 - x_k, \quad f_{t,u}(\bar{n}, \bar{w}, \bar{h}, \mathbf{s}) = f_{t,u}.$$

Using $|h_u(\bar{x})| \leq 2\sqrt{u}\sum_k |h_k| + \mathtt{K}u$, it easily follows that (u small)

$$|\mathbb{E}[f_{t,u}] - \mathbb{E}[f_{t,0}]| \leq 2\sqrt{u}\mathbb{E}[|H_k|] + u. \quad (36)$$

Therefore, we can permute the two limits in (34) and compute

$$\lim_{u\to 0}\lim_{\mathtt{K}\to+\infty}\mathbb{E}[f_{1,u}] + \frac{1}{2}.$$

From now on we keep the limits in that order. By the fundamental theorem of calculus

$$\mathbb{E}[f_{1,u}] = \mathbb{E}[f_{0,u}] + \int_0^1 dt\frac{d}{dt}\mathbb{E}[f_{t,u}]. \quad (37)$$

Our task is now reduced to estimating

$$\lim_{u\to 0}\lim_{\mathtt{K}\to+\infty}\int_0^1 dt\frac{d}{dt}\mathbb{E}[f_{t,u}].$$

This is done in Sections III-D and III-E. This requires a few preliminary results that are the object of Sections III-B and III-C.

B. Nishimori Identities

As already alluded to in the introduction the "magnetization" plays an important role

$$m_1 = \frac{1}{\mathtt{K}}\sum_{k=1}^{\mathtt{K}} x_k. \quad (38)$$

A closely related quantity is the "overlap parameter"

$$q_{12} = \frac{1}{\mathtt{K}}\sum_{k=1}^{\mathtt{K}} x_k^{(1)}x_k^{(2)} \quad (39)$$

where $x_k^{(1)}$ and $x_k^{(2)}$ are independent copies ("replicas") of the x_k. This means that the joint distribution of $\left(x_k^{(1)}, x_k^{(2)}\right)$ is the product measure

$$p_t\left(\bar{x}^{(1)}|\bar{n}, \bar{w}, \bar{h}, \mathbf{s}\right)p_t\left(\bar{x}^{(2)}|\bar{n}, \bar{w}, \bar{h}, \mathbf{s}\right).$$

The average with respect to this joint distribution is denoted (by a slight abuse of notation) with the same bracket $\langle - \rangle_{t,u}$. The important thing to notice is that the replicas are "coupled" through the common randomness $(\bar{N}, \bar{W}, \bar{H}, \mathbf{S})$.

Lemma 1 (Nishimori Identity [24]): Consider the distributions of m_1 and q_{12} defined as

$$\mathbb{P}_{m_1}(x) = \mathbb{E}\langle\delta(x - m_1)\rangle_{t,u}$$
$$\mathbb{P}_{q_{12}}(x) = \mathbb{E}\langle\delta(x - q_{12})\rangle_{t,u}.$$

Then

$$\mathbb{P}_{m_1}(x) = \mathbb{P}_{q_{12}}(x).$$

The above lemma in particular implies

$$\mathbb{E}[\langle m_1\rangle_{t,u}] = \mathbb{E}[\langle q_{12}\rangle_{t,u}]. \quad (40)$$

Such identities are known as Nishimori identities in the statistical physics literature and are a consequence of a gauge symmetry satisfied by the measure $\mathbb{E}\langle - \rangle_{t,u}$. They have also been used in the context of communications (see [12] and [19]). For completeness, a sketch of the proof is given in Appendix III.

The next two identities also follow from similar considerations.

Lemma 2 (Nishimori Identity): Let

$$\bar{\mathcal{Z}} = \bar{N} + \sqrt{\frac{B(t)}{\mathtt{N}}}\mathbf{S}\bar{z}.$$

Consider two replicas $\bar{\mathcal{Z}}^{(\alpha)}$, $\alpha = 1, 2$ corresponding to $z_k^{(\alpha)} = 1 - x_k^{(\alpha)}$. We then have

$$\frac{1}{\mathtt{N}}\mathbb{E}[\langle\|\bar{\mathcal{Z}}\|^2\rangle_{t,u}] = 1 \quad (41)$$

and

$$\mathbb{E}[\langle(\bar{N}\cdot\bar{\mathcal{Z}}^{(2)})(\bar{z}^{(1)}\cdot\bar{z}^{(2)})\rangle_{t,u}] = \sum_k \mathbb{E}[\langle(\bar{N}\cdot\bar{\mathcal{Z}})z_k\rangle_{t,u}]. \quad (42)$$

C. Concentration of Magnetization

A crucial feature of the calculation in the next paragraph is that m_1 (and q_{12}) concentrate.

Theorem 7 (Concentration of Magnetization): Fix any $\epsilon > 0$. For Lebesgue almost every $u > \epsilon$

$$\lim_{\mathtt{N}\to\infty}\int_0^1 dt\mathbb{E}\langle|m_1 - \mathbb{E}\langle m_1\rangle_{t,u}|\rangle_{t,u} = 0.$$

The proof of this theorem, which is the point where the careful tuning of the perturbation is needed, has an interest of its own and is presented in Section IV. Similar statements in the spin glass literature have been obtained by Talagrand [10, Sec. 2.12]. The usual signature of replica symmetry breaking is the absence of concentration for the overlap parameter q_{12}. This theorem combined with the Nishimori identity "explains" why the replica symmetry is not broken.

We will also need the following corollary.

Corollary 1: The following holds:

$$\frac{1}{\mathtt{N}^{3/2}}\mathbb{E}\langle(\bar{N}\cdot\mathbf{S}\bar{z})(1 - m_1)\rangle_{t,u} = \frac{1}{\mathtt{N}^{3/2}}\mathbb{E}\langle\bar{N}\cdot\mathbf{S}\bar{z}\rangle_{t,u}(1 - \mathbb{E}\langle m_1\rangle_{t,u}) + o_{\mathtt{N}}(1)$$

with $\lim_{\mathtt{N}\to+\infty} o_{\mathtt{N}}(1) = 0$ for almost every $u > 0$.

Proof: Using Cauchy–Schwartz inequality, we get

$$\frac{1}{\mathtt{N}^{3/2}}\mathbb{E}\langle(\bar{N}\cdot\mathbf{S}\bar{z})(\mathbb{E}\langle m_1\rangle_{t,u} - m_1)\rangle_{t,u} \le \frac{1}{\mathtt{N}^{3/2}} \times(\mathbb{E}\langle(\bar{N}\cdot\mathbf{S}\bar{z})^2\rangle_{t,u})^{1/2}(\mathbb{E}\langle(\mathbb{E}\langle m_1\rangle_{t,u} - m_1)^2\rangle_{t,u})^{1/2}.$$

Because of the concentration of the magnetization m_1 (Theorem 7), it suffices to prove that

$$\mathbb{E}\left\langle\left(\mathtt{N}^{-\frac{3}{2}}\sum_{i,l}N_i S_{il} z_l\right)^2\right\rangle_{t,u} \le D \quad (43)$$

for some constant D independent of $\mathtt{N}$. The proof is given in Appendix IV. □

D. Computation of $\frac{d}{dt}\mathbb{E}[f_{t,u}]$

Let $z_k = 1 - x_k$

$$\frac{d}{dt}\mathbb{E}[f_{t,u}] = T_1 + T_2 \quad (44)$$

where

$$T_1 = -\frac{\lambda'(t)}{2\sqrt{\lambda(t)}\mathtt{K}}\mathbb{E}\langle\bar{W}\cdot\bar{z}\rangle_{t,u} - \frac{\lambda'(t)}{2\mathtt{K}}\mathbb{E}\langle\bar{z}\cdot\bar{z}\rangle_{t,u},$$

and

$$T_2 = -\frac{1}{\mathtt{K}\sqrt{\mathtt{N}}}\frac{B'(t)}{2\sqrt{B(t)}}\mathbb{E}\langle\bar{\mathcal{Z}}\cdot\mathbf{S}\bar{z}\rangle_{t,u}.$$

Note that this computation involves an exchange of the t-derivative with an integral, which is permitted because for finite N and K the following standard criterion can be applied. Let $g(t,y)$ be a function that is jointly continuous in $(t,y) \in [0,1]\times\mathbb{R}^n$ and whose (possibly indefinite) integral $\int dy\, g(t,y)$ converges for each t. If $\frac{\partial}{\partial t}g(t,y)$ is jointly continuous in (t,y) and $\int dy\,\frac{\partial}{\partial t}g(t,y)$ converges uniformly in t, then we have $\frac{d}{dt}\int dy\, g(t,y) = \int dy\,\frac{\partial}{\partial t}g(t,y)$.

1) Transforming T_1: The first term of T_1 is given by

$$\mathbb{E}\langle\bar{W}\cdot\bar{z}\rangle_{t,u} = \sum_k \mathbb{E}\left[\frac{\sum_{\bar{x}} e^{\mathtt{H}_{t,u}(\bar{x})}z_k}{\sum_{\bar{x}} e^{\mathtt{H}_{t,u}(\bar{x})}}W_k\right].$$

Recall that the terms $\mathtt{H}_{t,u}(\bar{x})$ are dependent on W_k. Using the integration by parts formula (16) with respect to W_k leads to

$$\mathbb{E}\langle\bar{W}\cdot\bar{z}\rangle_{t,u} = \sum_k \mathbb{E}\left[\frac{\sum_{\bar{x}} e^{\mathtt{H}_{t,u}(\bar{x})}(-(W_k + \sqrt{\lambda(t)}z_k))z_k}{\sum_{\bar{x}} e^{\mathtt{H}_{t,u}(\bar{x})}} - \frac{(\sum_{\bar{x}} e^{\mathtt{H}_{t,u}(\bar{x})}(-(W_k + \sqrt{\lambda(t)}z_k)))(\sum_{\bar{x}} e^{\mathtt{H}_{t,u}(\bar{x})}z_k)}{(\sum_{\bar{x}} e^{\mathtt{H}_{t,u}(\bar{x})})^2}\right].$$

The first term is equal to

$$-\mathbb{E}\left\langle\sum_k (W_k + \sqrt{\lambda(t)}z_k)z_k\right\rangle_{t,u} = -\mathbb{E}\langle(\bar{W} + \sqrt{\lambda(t)}\bar{z})\cdot\bar{z}\rangle_{t,u}.$$

The second term can be expressed as

$$\mathbb{E}\left[\frac{\sum_{\bar{x}^{(1)},\bar{x}^{(2)}} e^{\mathtt{H}_{t,u}(\bar{x}^{(1)})+\mathtt{H}_{t,u}(\bar{x}^{(2)})} \sum_k z_k^{(1)}(W_k + \sqrt{\lambda(t)} z_k^{(2)})}{\sum_{\bar{x}^{(1)},\bar{x}^{(2)}} e^{\mathtt{H}_{t,u}(\bar{x}^{(1)})+\mathtt{H}_{t,u}(\bar{x}^{(2)})}}\right] = \mathbb{E}\langle \bar{z}^{(1)} \cdot (\bar{W} + \sqrt{\lambda(t)}\bar{z}^{(2)})\rangle_{t,u}.$$

Therefore

$$\begin{aligned} T_1 &= \frac{\lambda'(t)}{2\sqrt{\lambda(t)}\mathtt{K}} \mathbb{E}\langle (\bar{W} + \sqrt{\lambda(t)}\bar{z}) \cdot \bar{z}\rangle_{t,u} \\ &\quad - \frac{\lambda'(t)}{2\sqrt{\lambda(t)}\mathtt{K}} \mathbb{E}\langle \bar{z}^{(1)} \cdot (\bar{W} + \sqrt{\lambda(t)}\bar{z}^{(2)})\rangle_{t,u} \\ &\quad - \frac{\lambda'(t)}{2\mathtt{K}} \mathbb{E}\langle \bar{z} \cdot \bar{z}\rangle_{t,u} \\ &\overset{(a)}{=} -\frac{\lambda'(t)}{2} \mathbb{E}\langle 1 - 2m_1 + q_{12}\rangle_t \\ &\overset{(b)}{=} -\frac{\lambda'(t)}{2} \mathbb{E}\langle 1 - m_1\rangle_{t,u}. \end{aligned}$$

The equality (a) follows from the fact that the $\bar{W}$ terms cancel and the equality (b) follows from (40). From the relation between $\lambda(t)$ and $B(t)$ given in (26), T_1 can be rewritten in the form

$$T_1 = \frac{B'(t)}{2(1 + \beta(1-m)B(t))^2} \mathbb{E}\langle 1 - m_1\rangle_{t,u}. \tag{45}$$

2) Transforming T_2: The term T_2 can be rewritten as

$$T_2 = -\frac{B'(t)}{2\beta B(t)\mathtt{N}} \mathbb{E}\langle \|\bar{\mathcal{Z}}\|^2\rangle_{t,u} + \frac{B'(t)}{2\beta B(t)\mathtt{N}} \mathbb{E}\|\bar{N}\|^2 + \frac{B'(t)}{2\sqrt{B(t)}\mathtt{K}\sqrt{\mathtt{N}}} \mathbb{E}\langle \bar{N} \cdot \mathbf{S}\bar{z}\rangle_{t,u}.$$

Because of (41), the first two terms cancel, which results in

$$T_2 = \frac{B'(t)}{2\sqrt{B(t)}\mathtt{K}\sqrt{\mathtt{N}}} \mathbb{E}\langle \bar{N} \cdot \mathbf{S}\bar{z}\rangle_{t,u}. \tag{46}$$

Using integration by parts with respect to S_{ik}, we get

$$T_2 = -\frac{B'(t)}{2\mathtt{KN}} \mathbb{E}\langle (\bar{N} \cdot \bar{\mathcal{Z}})(\bar{z} \cdot \bar{z})\rangle_{t,u} + \frac{B'(t)}{2\mathtt{KN}} \mathbb{E}\langle (\bar{N} \cdot \bar{\mathcal{Z}}^{(2)})(\bar{z}^{(1)} \cdot \bar{z}^{(2)})\rangle_{t,u}.$$

Now applying the Nishimori identity (42) implies

$$\begin{aligned} T_2 &= -\frac{B'(t)}{2\mathtt{KN}} \sum_k \mathbb{E}\langle (\bar{N} \cdot \bar{\mathcal{Z}}) z_k\rangle_{t,u} \\ &= -\frac{B'(t)}{2} \frac{1}{\mathtt{NK}} \sum_k \mathbb{E}[\|\bar{N}\|^2 \langle z_k\rangle_{t,u}] \\ &\quad - \frac{B'(t)\sqrt{B(t)}}{2\mathtt{KN}^{3/2}} \sum_k \mathbb{E}\langle .(\bar{N} \cdot \mathbf{S}\bar{z})(1 - x_k)\rangle_{t,u}. \end{aligned}$$

Since $\frac{1}{\mathtt{N}}\|\bar{N}\|^2 = \frac{1}{\mathtt{N}}\sum_i N_i^2$ concentrates on 1, we get

$$T_2 = -\frac{B'(t)}{2} \mathbb{E}\langle 1 - m_1\rangle_{t,u} + o_{\mathtt{N}}(1) - \frac{\beta B'(t)\sqrt{B(t)}}{2\mathtt{KN}^{1/2}} \mathbb{E}\langle (\bar{N} \cdot \mathbf{S}\bar{z})(1 - m_1)\rangle_{t,u}.$$

Applying Corollary 1 to the last expression for T_2 together with (46), we obtain a closed affine equation for the latter, whose solution is

$$T_2 = -\frac{B'(t)\mathbb{E}\langle 1 - m_1\rangle_{t,u}}{2(1 + \beta B(t)\mathbb{E}\langle 1 - m_1\rangle_{t,u})} + o_{\mathtt{N}}(1). \tag{47}$$

E. End of Proof

We add and subtract the term $\frac{1}{2\beta}\ln(1 + \beta B(1-m))$ from (37) and use the integral representation

$$\frac{1}{2\beta}\ln(1 + \beta B(1-m)) = \frac{1}{2\beta}\int_0^1 dt \frac{\beta B'(t)(1-m)}{1 + \beta B(t)(1-m)}$$

to obtain

$$\mathbb{E}[f_{1,u}] = \mathbb{E}[f_{0,u}] - \frac{1}{2\beta}\ln(1 + \beta B(1-m)) + \int_0^1 dt \left(\frac{d}{dt}\mathbb{E}[f_{t,u}] + \frac{B'(t)(1-m)}{2(1 + \beta B(t)(1-m))}\right).$$

From (45), we can express T_1 as

$$T_1 = \frac{B'(t)\mathbb{E}\langle m - m_1\rangle_{t,u}}{2(1 + \beta(1-m)B(t))^2} + \frac{B'(t)(1-m)}{2(1 + \beta(1-m)B(t))^2}. \tag{48}$$

Using (47), we get

$$T_2 + \frac{B'(t)(1-m)}{2(1 + \beta B(t)(1-m))} = \frac{B'(t)\mathbb{E}\langle m_1 - m\rangle_{t,u}}{2(1 + \beta B(t)\mathbb{E}\langle 1 - m_1\rangle_{t,u})(1 + \beta B(t)(1-m))} + o_{\mathtt{N}}(1). \tag{49}$$

Using (44), (48), and (49), we get

$$\begin{aligned} \frac{d}{dt}\mathbb{E}[f_{t,u}] &+ \frac{B'(t)(1-m)}{2(1 + \beta B(t)(1-m))} \\ &= \frac{B'(t)\mathbb{E}\langle m - m_1\rangle_{t,u}}{2(1 + \beta(1-m)B(t))^2} + \frac{B'(t)(1-m)}{2(1 + \beta(1-m)B(t))^2} \\ &\quad + \frac{B'(t)\mathbb{E}\langle m_1 - m\rangle_{t,u}}{2(1 + \beta B(t)\mathbb{E}\langle 1 - m_1\rangle_{t,u})(1 + \beta B(t)(1-m))} \\ &\quad + o_{\mathtt{N}}(1) \\ &= R(t) + \frac{B'(t)(1-m)}{2(1 + \beta B(t)(1-m))^2} + o_{\mathtt{N}}(1) \end{aligned}$$

where

$$R(t) = \frac{\beta B'(t)B(t)(\mathbb{E}\langle m_1 - m\rangle_{t,u})^2}{2(1 + \beta B(t)(1-m))^2(1 + \beta B(t)\mathbb{E}\langle 1 - m_1\rangle_{t,u})}.$$

So the integral has a positive contribution $\int_0^1 dtR(t) \geq 0$ plus a computable contribution equal to $\frac{B(1-m)}{2(1+\beta B(1-m))} = \frac{\lambda}{2}(1-m)$. Finally, thanks to (33), we get

$$\frac{1}{2} + \mathbb{E}[f_{1,u}] = \int Dz \ln(2\cosh(\sqrt{\lambda}z + \lambda)) - \frac{1}{2\beta} - \frac{1}{2\beta}\ln(1+\beta B(1-m)) - \frac{\lambda}{2}(1+m) + \int_0^1 R(t)dt + o_{\mathtt{N}}(1) + O(\sqrt{u}) \quad (50)$$

where for almost every (a.e.) $u > \epsilon$, $\lim_{\mathtt{N}\to\infty} o_{\mathtt{N}}(1) = 0$. We first take the limit $\mathtt{N} \to \infty$, then $u \to \epsilon$ (along some appropriate sequence), and then $\epsilon \to 0$ to obtain a formula for the free energy where the only nonexplicit contribution is $\int_0^1 dtR(t)$. Since this is positive for all m, we obtain a lower bound on the free energy which is equivalent to $\lim_{\mathtt{K}\to\infty} C_{\mathtt{K}} \leq c_{RS}(m)$. This bound holds for any $m \in [0,1]$, which in turn implies

$$\lim_{\mathtt{K}\to\infty} C_{\mathtt{K}} \leq \min_{m\in[0,1]} c_{RS}(m).$$

To prove the equality for the capacity, we need to show that if $m = m^*$, where $m^* = \arg\min_{m\in[0,1]} c_{RS}(m)$, then $R(t) = 0$. From Theorem 7, we know that $\mathbb{E}\langle|m_1 - \mathbb{E}\langle m_1\rangle_{t,u}|\rangle_{t,u} \xrightarrow{\mathtt{N}\to\infty} 0$. Therefore, to prove the equality, it suffices to show that $\mathbb{E}\langle m_1\rangle_{t,u} \xrightarrow{\mathtt{N}\to\infty} m^*$.

IV. Concentration of Magnetization

The goal of this section is to prove Theorem 7. The proof is organized in a succession of lemmas. By the same methods used for Theorem 2, we can prove the following.

Lemma 3 (Concentration of $f_{t,u}$): There exists a strictly positive constant α (which remains positive for all t and u) such that

$$\mathbb{P}[|f_{t,u} - \mathbb{E}[f_{t,u}]| \geq \epsilon] = O(e^{-\alpha\epsilon^2\sqrt{\mathtt{K}}}).$$

The perturbation term (30) has been chosen carefully so that the following holds.

Lemma 4 (Convexity of $f_{t,u}$): When considered as a function of u, $f_{t,u}$ is convex in u.

Proof: We simply evaluate the second derivative and show it is positive. We proceed by computing

$$\frac{df_{t,u}}{du} = \langle L(\bar{x})\rangle_{t,u} - \frac{1}{\mathtt{K}2\sqrt{u}}\sum_k |h_k|$$

where we have defined

$$L(\bar{x}) = \frac{1}{\mathtt{K}}\frac{1}{2\sqrt{u}}\sum_k h_k x_k + \frac{1}{\mathtt{K}}\sum_k x_k.$$

Differentiating again, we get

$$\frac{d^2 f_{t,u}}{du^2} = \frac{1}{\mathtt{K}}\left\langle \frac{-1}{4u^{3/2}}\sum_k h_k x_k\right\rangle_{t,u} + \frac{1}{4u^{3/2}\mathtt{K}}\sum_k |h_k| + \mathtt{K}(\langle L(\bar{x})^2\rangle_{t,u} - \langle L(\bar{x})\rangle_{t,u}^2) \geq 0. \quad (51)$$

□

The quantity $L(\bar{x})$ turns out to be very useful and satisfies two concentration properties.

Lemma 5 (Self-Averaging of $L(\bar{x})$): For any $a > \epsilon > 0$ fixed

$$\int_\epsilon^a du \mathbb{E}\,\langle |L(\bar{x}) - \langle L(\bar{x})\rangle_{t,u}|\rangle_{t,u} = O\left(\frac{1}{\sqrt{\mathtt{K}}}\right).$$

Proof: From (51), we have

$$\int_\epsilon^a du \mathbb{E}\left\langle \left(L(\bar{x}) - \langle L(\bar{x})\rangle_{t,u}\right)^2\right\rangle_{t,u} \leq \int_\epsilon^a du \frac{1}{\mathtt{K}}\frac{d^2}{du^2}\mathbb{E}[f_{t,u}] \leq \frac{1}{\mathtt{K}}\left(\frac{d}{du}\mathbb{E}[f_{t,a}] - \frac{d}{du}\mathbb{E}[f_{t,\epsilon}]\right) = O\left(\frac{1}{\mathtt{K}}\right).$$

In the very last equality we use that the first derivative of $\mathbb{E}[f_{t,u}]$ is bounded for $u \geq \epsilon$. Using Cauchy–Schwartz inequality for $\int \mathbb{E}\langle - \rangle_{t,u}$, we obtain the lemma. □

Lemma 6 (Self-Averaging of $\langle L(\bar{x})\rangle_{t,u}$): For any $a > \epsilon > 0$ fixed

$$\int_\epsilon^a du \mathbb{E}\left|\langle L(\bar{x})\rangle_{t,u} - \mathbb{E}\langle L(\bar{x})\rangle_{t,u}\right| = O\left(\frac{1}{\mathtt{K}^{\frac{1}{16}}}\right).$$

Proof: From convexity of $f_{t,u}$ with respect to u (Lemma 4), we have for any $\delta > 0$

$$\frac{d}{du}f_{t,u} - \frac{d}{du}\mathbb{E}[f_{t,u}] \leq \frac{f_{t,u+\delta} - f_{t,u}}{\delta} - \frac{d}{du}\mathbb{E}[f_{t,u}] \leq \frac{f_{t,u+\delta} - \mathbb{E}[f_{t,u+\delta}]}{\delta} - \frac{f_{t,u} - \mathbb{E}[f_{t,u}]}{\delta} + \frac{d}{du}\mathbb{E}[f_{t,u+\delta}] - \frac{d}{du}\mathbb{E}[f_{t,u}].$$

A similar lower bound holds with δ replaced by $-\delta$. Now, from Lemma 3, we know that the first two terms are $O(\mathtt{K}^{\frac{1}{4}})$. Thus, from the formula for the first derivative in the proof of Lemma 4 and the fact that the fluctuations of $\frac{1}{\mathtt{K}}\sum_{k=1}^{\mathtt{K}}|h_k|$ are $O\left(\frac{1}{\sqrt{\mathtt{K}}}\right)$, we get

$$\mathbb{E}\left|\langle L(\bar{x})\rangle_{t,u} - \mathbb{E}\langle L(\bar{x})\rangle_{t,u}\right| \leq \frac{1}{\delta}O\left(\frac{1}{\sqrt{\mathtt{K}}}\right) + \frac{1}{\delta}O\left(\frac{1}{\mathtt{K}^{\frac{1}{4}}}\right) + \frac{d}{du}\mathbb{E}[f_{t,u+\delta}] - \frac{d}{du}\mathbb{E}[f_{t,u}].$$

We will choose $\delta = \frac{1}{\mathtt{K}^{\frac{1}{8}}}$. Note that we cannot assume that the difference of the two derivatives is small because the first derivative of the free energy is not uniformly continuous in $\mathtt{K}$ (as $\mathtt{K}\to\infty$, it may develop jumps at the phase transition points). The free energy itself is uniformly continuous. For this reason, if we integrate with respect to u, using (36), we get

$$\int_\epsilon^a du \mathbb{E}\left|\langle L(\bar{x})\rangle_{t,u} - \mathbb{E}\langle L(\bar{x})\rangle_{t,u}\right| \leq O\left(\frac{1}{\mathtt{K}^{\frac{1}{16}}}\right).$$ □

Using the two last lemmas, we can prove Theorem 7.

Proof of Theorem 7: Combining the concentration lemmas, we get

$$\int_\epsilon^a du \mathbb{E}\langle |L(\bar{x}) - \mathbb{E}\langle L(\bar{x})\rangle_{t,u}|\rangle_{t,u} \leq O\left(\frac{1}{\mathtt{K}^{\frac{1}{16}}}\right).$$

For any function $g(\bar{x})$ such that $|g(\bar{x})| \leq 1$, we have

$$\int_\epsilon^a du|\mathbb{E}\langle L(\bar{x})g(\bar{x})\rangle_{t,u} - \mathbb{E}\langle L(\bar{x})\rangle_{t,u}\mathbb{E}\langle g(\bar{x})\rangle_{t,u}|\rangle_{t,u} \leq \int_\epsilon^a du\mathbb{E}\langle|L(\bar{x}) - \mathbb{E}\langle L(\bar{x})\rangle_{t,u}|\rangle_{t,u}.$$

More generally, the same thing holds if one takes a function depending on many replicas such as $g(\bar{x}^{(1)}, \bar{x}^{(2)}) = q_{12}$. Using integration by parts formula with respect to H_k

$$\begin{aligned}\mathbb{E}\langle L(\bar{x})q_{12}\rangle_{t,u} =& \mathbb{E}\Big\langle \frac{1}{2\mathtt{K}\sqrt{u}}\sum_k H_k x_k q_{12}\Big\rangle_{t,u} + \mathbb{E}\langle m_1 q_{12}\rangle_{t,u}\\ =& \frac{1}{2}\mathbb{E}\langle(1+q_{12})q_{12}\rangle_{t,u} - \frac{1}{2}\mathbb{E}\langle(q_{13}+q_{14})q_{12}\rangle_{t,u}\\ &+ \mathbb{E}\langle m_1 q_{12}\rangle_{t,u}\\ =& \frac{1}{2}\mathbb{E}\langle(1+q_{12})q_{12}\rangle_{t,u}\\ =& \frac{1}{2}\mathbb{E}\langle m_1 + m_1^2\rangle_{t,u}. \end{aligned} \quad (52)$$

In the last two equalities, we used the Nishimori identity (40). By a similar calculation, we get

$$\begin{aligned}\mathbb{E}\langle L(\bar{x})\rangle_{t,u}\mathbb{E}\langle q_{12}\rangle_{t,u} &= \frac{1}{2}\mathbb{E}\langle 1 - q_{12} + 2m_1\rangle_{t,u}\mathbb{E}\langle q_{12}\rangle_{t,u}\\ &= \frac{1}{2}(\mathbb{E}\langle m_1\rangle_t + (\mathbb{E}\langle m_1\rangle_t)^2). \end{aligned} \quad (53)$$

From (52) and (53), we get

$$\int_\epsilon^a du|\mathbb{E}\langle m_1^2\rangle_{t,u} - (\mathbb{E}\langle m_1\rangle_{t,u})^2| \leq O\left(\frac{1}{\mathtt{K}^{\frac{1}{16}}}\right).$$

Now integrating with respect to t and exchanging the integrals (by Fubini's theorem), we get

$$\int_\epsilon^a du\int_0^1 dt|\mathbb{E}\langle m_1^2\rangle_{t,u} - (\mathbb{E}\langle m_1\rangle_{t,u})^2| \leq O\left(\frac{1}{\mathtt{K}^{\frac{1}{16}}}\right).$$

The limit of the left-hand side as $\mathtt{K} \to \infty$ therefore vanishes. Using Fatou–Lebesgue theorem this limit can be exchanged with the u integral and we get the desired result. [Note that one can further exchange the limit with the t-integral and obtain that the fluctuations of m_1 vanish for almost every (t, u).] □

V. Proof of Independence From Spreading Sequence Distribution: Theorem 4

We consider a communication system with spreading values r_{ik} generated from a symmetric random variable R_{ik} whose distribution belongs to Class A. We compare the capacity of this system to the Gaussian $\mathcal{N}(0,1)$ case whose spreading sequence values are denoted by s_{ik}. The comparison is done through an interpolating system with respect to the two spreading sequences

$$v_{ik}(t) = \sqrt{t}r_{ik} + \sqrt{1-t}s_{ik}, \qquad 0 \leq t \leq 1.$$

Let $\mathbf{v}(t)$ denote the matrix with entries $v_{ik}(t)$ and let $\bar{v}_i(t)$ denote the ith row of the matrix. By the fundamental theorem of calculus, the capacities are related by

$$\begin{aligned}C_\mathtt{K} - C_\mathtt{K}^g &= \mathbb{E}_\mathbf{R}[C(\mathbf{r})] - \mathbb{E}_\mathbf{S}[C(\mathbf{s})]\\ &= \int_0^1 dt\frac{d}{dt}\mathbb{E}_{\mathbf{V}(t)}[C(\mathbf{v}(t))].\end{aligned}$$

From (9), the derivative is equal to

$$\frac{d}{dt}\mathbb{E}_{\mathbf{V}(t)}[C(\mathbf{v}(t))] = -\mathbb{E}_\mathbf{S}\mathbb{E}_\mathbf{R}\frac{d}{dt}\mathbb{E}_{\bar{Y}|\mathbf{V}(t)}[f(\bar{y}, \mathbf{v}(t)].$$

As before, we can assume that the transmitted sequence is $\bar{x}^0$. It is convenient to first perform the change of variables $\bar{y} = \bar{n} + \mathtt{N}^{-1/2}\mathbf{v}(t)\bar{x}^0$ and then perform the t derivative. One finds

$$\begin{aligned}\frac{d}{dt}\mathbb{E}_{\mathbf{V}(t)}[C(\mathbf{v}(t))] &= \frac{1}{\sigma^2\mathtt{K}\sqrt{\mathtt{N}}}\mathbb{E}_{\mathbf{S},\mathbf{R},\bar{N}}\\ &\times\left\langle\left(\bar{N} + \frac{1}{\sqrt{\mathtt{N}}}\mathbf{V}(t)(\bar{x}^0 - \bar{x})\right)\cdot\mathbf{V}'(t)(\bar{x}^0 - \bar{x})\right\rangle_t \end{aligned} \quad (54)$$

where $\langle - \rangle_t$ is the average with respect to the normalized measure

$$\frac{1}{2^\mathtt{K} Z_t}\exp\left(-\frac{1}{2\sigma^2}\|\bar{n} - \mathtt{N}^{-\frac{1}{2}}\mathbf{v}(t)(\bar{x}^0 - \bar{x})\|^2\right).$$

We split (54) into two contributions $T_1 - T_2$ corresponding to

$$\mathbf{v}'(t) = \frac{1}{2\sqrt{t}}\mathbf{r} - \frac{1}{2\sqrt{1-t}}\mathbf{s}. \quad (55)$$

For T_1, we have

$$T_1 = \sum_{i,k}T_1(i,k) = \frac{1}{2\sqrt{t}}\sum_{i,k}\mathbb{E}_{\mathbf{S},\mathbf{R},\bar{N}}[R_{ik}g_{ik}] \quad (56)$$

with

$$g_{ik} = \frac{1}{\sigma^2\mathtt{K}\sqrt{\mathtt{N}}}\left\langle\left(\bar{n} + \frac{1}{\sqrt{\mathtt{N}}}\mathbf{v}(t)(\bar{x}^0 - \bar{x})\right)_i(x_k^0 - x_k)\right\rangle_t. \quad (57)$$

For T_2, we have

$$T_2 = \sum_{i,k}T_2(i,k) = \frac{1}{2\sqrt{1-t}}\sum_{i,k}\mathbb{E}_{\mathbf{S},\mathbf{R},\bar{N}}[S_{ik}g_{ik}] \quad (58)$$

with the same expression for g_{ik}. For a function $g(x)$, let $g'(x) = \frac{\partial g(x)}{\partial x}$. For each contribution in the sums (56) and (58), we use integration by parts formulas. For (56), we use the formula [29, eq. 5.82] (it is an exercise to check that it is valid for any symmetric random variable)

$$\begin{aligned}&\mathbb{E}[R_{ik}g(R_{ik})]\\ &= \mathbb{E}\Big[R_{ik}^2 g'(R_{ik})\Big]\\ &\quad - \frac{1}{4}\mathbb{E}\Big[|R_{ik}|\int_{-|R_{ik}|}^{|R_{ik}|}(R_{ik}^2 - u^2)\frac{\partial^3 g(u)}{\partial u^3}du\Big]\\ &= \mathbb{E}\Big[g'(R_{ik})\Big] + \mathbb{E}\Big[(R_{ik}^2 - 1)\int_0^{R_{ik}}\frac{\partial^2 g(u)}{\partial u^2}du\Big]\\ &\quad - \frac{1}{4}\mathbb{E}\Big[|R_{ik}|\int_{-|R_{ik}|}^{|R_{ik}|}(R_{ik}^2 - u^2)\frac{\partial^3 g(u)}{\partial u^3}du\Big]. \end{aligned} \quad (59)$$

For (58), we use the standard Gaussian (unit variance) integration by parts formula

$$\mathbb{E}\big[S_{ik}g(S_{ik})\big] = \mathbb{E}\big[g'(S_{ik})\big]. \quad (60)$$

When we consider $T_1 - T_2$, the term corresponding to the expectation in (60) cancels with that of the first expectation in (59) and we get

$$T_1 - T_2 = \frac{1}{2\sqrt{t}} \sum_{i,k} \mathbb{E}\left[(R_{ik}^2 - 1) \int_0^{R_{ik}} \frac{\partial^2 g_{ik}(u)}{\partial u^2} du\right] - \frac{1}{8\sqrt{t}} \sum_{i,k} \mathbb{E}\left[|R_{ik}| \int_{-|R_{ik}|}^{|R_{ik}|} (R_{ik}^2 - u^2) \frac{\partial^3 g_{ik}(u)}{\partial u^3} du\right]. \quad (61)$$

It remains to be proved that both terms with the partial derivatives tend to zero as $\mathtt{N} \to +\infty$. This computation is rather lengthy and is deferred to Appendix II, but for the convenience of the reader, we point out the mechanism that is at work. On the expression for g_{ik}, one sees that when the $\frac{\partial^2}{\partial u^2}$ and $\frac{\partial^3}{\partial u^3}$ derivatives are performed, extra powers $\mathtt{N}^{-1}$ and $\mathtt{N}^{-3/2}$ are generated. Therefore, we get

$$\mathbb{E}\left[(R_{ik}^2 - 1) \int_0^{R_{ik}} \frac{\partial^2 g_{ik}}{\partial u_{ik}^2} du_{ik}\right] = O(\mathtt{N}^{-5/2}) \quad (62)$$

and

$$\mathbb{E}\left[|R_{ik}| \int_{-|R_{ik}|}^{|R_{ik}|} (R_{ik}^2 - u_{ik}^2) \frac{\partial^3 g_{ik}}{\partial u_{ik}^3} du_{ik}\right] = O(\mathtt{N}^{-3}). \quad (63)$$

Since one sums over $\mathtt{KN}$ terms, one finds that the final contributions are $O(\mathtt{N}^{-1/2})$ and $O(\mathtt{N}^{-1})$.

VI. Proof of Existence of Limit: Theorem 5

The relation between the free energy and the capacity in (10) implies that it is sufficient to show the existence of limit for the average free energy $\mathcal{F}_{\mathtt{K}} = \mathbb{E}[f(\bar{y}, \mathbf{s})]$. The idea is to use Fekete's lemma [30, ch. 3], which states that if a sequence $\{a_n\}$ is super additive, i.e., $a_{m+n} \geq a_m + a_n$, then $\lim_{n\to\infty} \frac{a_n}{n}$ exists (provided $\frac{a_n}{n}$ is bounded). The relevant sequence for us is $\{\mathtt{K}\mathcal{F}_{\mathtt{K}}\}$. The aim is therefore to show that $\mathtt{K}\mathcal{F}_{\mathtt{K}} \geq \mathtt{K}_1\mathcal{F}_{\mathtt{K}_1} + \mathtt{K}_2\mathcal{F}_{\mathtt{K}_2}$ for $\mathtt{K} = \mathtt{K}_1 + \mathtt{K}_2$.

As in the previous sections, working directly with this system is difficult and hence we perturb the Hamiltonian with $h_u(\bar{x})$ as defined in (30)

$$\mathtt{H}_u(\bar{x}) = -\frac{1}{2\sigma^2} \left\| \bar{n} + \frac{1}{\sqrt{\mathtt{N}}} \mathbf{s}(\bar{1} - \bar{x}) \right\|^2 + h_u(\bar{x}). \quad (64)$$

Let us define the corresponding partition function as Z_u and the free energy as $\mathcal{F}_{\mathtt{K}}(u) = \frac{1}{\mathtt{K}}\mathbb{E}[\ln Z_u]$. The original free energy is obtained by substituting $u = 0$, i.e., $\mathcal{F}_{\mathtt{K}} = \mathcal{F}_{\mathtt{K}}(0)$. From the uniform continuity of $\mathcal{F}_{\mathtt{K}}(u)$, it is sufficient to show the convergence of $\mathcal{F}_{\mathtt{K}}(u)$ for some u close to zero. Even this turns out to be difficult and what we can show is the existence of the limit $\int_{u=\epsilon}^{a} \mathcal{F}_{\mathtt{K}}(u) du$ for any $a > \epsilon > 0$. However, this is sufficient for us due to the following: from the continuity of the free energy with u(36), we have

$$\int_{\epsilon}^{2\epsilon} (\mathcal{F}_{\mathtt{K}}(u) - |O(1)|\sqrt{u}) du \leq \epsilon \mathcal{F}_{\mathtt{K}} \leq \int_{\epsilon}^{2\epsilon} (\mathcal{F}_{\mathtt{K}}(u) + |O(1)|\sqrt{u}) du.$$

Since the limit of the integral exists, we have

$$\left| \limsup_{\mathtt{K}\to\infty} \mathcal{F}_{\mathtt{K}} - \liminf_{\mathtt{K}\to\infty} \mathcal{F}_{\mathtt{K}} \right| \leq |O(1)| \sqrt{\epsilon}.$$

This ϵ can be made as small as desired and hence the theorem follows.

Let $\mathtt{K} = \mathtt{K}_1 + \mathtt{K}_2$ and let $\frac{\mathtt{K}}{\beta}, \frac{\mathtt{K}_1}{\beta}, \frac{\mathtt{K}_2}{\beta} \in \mathbb{N}$. This assumption can be removed by considering their integer parts. But we will stick to this assumption to simplify the proof. Split the $(\mathtt{N}\times\mathtt{K})$-dimensional spreading matrix $\mathbf{s}$ into two parts of dimension $\mathtt{N}_1 \times \mathtt{K}$ and $\mathtt{N}_2 \times \mathtt{K}$ and denote these matrices by $\mathbf{s}_1, \mathbf{s}_2$, respectively. Let $\mathbf{t}_1, \mathbf{t}_2$ be two spreading matrices with dimensions $\mathtt{N}_1 \times \mathtt{K}_1$ and $\mathtt{N}_2 \times \mathtt{K}_2$. All the entries of these matrices are distributed as $\mathcal{N}(0,1)$ and the noise is Gaussian with variance σ^2. Similarly, split the noise vector $\bar{n} = (\bar{n}_1, \bar{n}_2)$ where $\bar{n}_i$ is of length $\mathtt{N}_i$ and $\bar{x} = (\bar{x}_1, \bar{x}_2)$ where $\bar{x}_i$ is of length $\mathtt{K}_i$. Let us consider the following Hamiltonian:

$$\mathtt{H}_{t,u}(\bar{x}) = -\frac{1}{2\sigma^2} \left\| \bar{n}_1 + \frac{\sqrt{t}}{\sqrt{\mathtt{N}}} \mathbf{s}_1(\bar{1} - \bar{x}) + \frac{\sqrt{1-t}}{\sqrt{\mathtt{N}_1}} \mathbf{t}_1(\bar{1} - \bar{x}_1) \right\|^2 - \frac{1}{2\sigma^2} \left\| \bar{n}_2 + \frac{\sqrt{t}}{\sqrt{\mathtt{N}}} \mathbf{s}_2(\bar{1} - \bar{x}) + \frac{\sqrt{1-t}}{\sqrt{\mathtt{N}_2}} \mathbf{t}_2(\bar{1} - \bar{x}_2) \right\|^2 + h_u(\bar{x}).$$

Note that the all-one vectors $\bar{1}$ appearing above are of different dimensions (the dimension is clear from the context). For a moment neglect the $h_u(\bar{x})$ part of the Hamiltonian and consider the remaining part. At $t = 1$, we get the Hamiltonian corresponding to an $\mathtt{N} \times \mathtt{K}$ CDMA system with spreading matrix $\begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{bmatrix}$. At $t = 0$, we get the Hamiltonian corresponding to two independent CDMA systems with spreading matrices $\mathbf{t}_i$ of dimensions $\mathtt{N}_i \times \mathtt{K}_i$. As before we perturb the Hamiltonian with $h_u(\bar{x})$ so that we can use the concentration results for the magnetization.

Let $Z_{t,u}$ be the partition function with this Hamiltonian and the corresponding average free energy is given by $g_{t,u} = \frac{1}{\mathtt{K}}\mathbb{E}[\ln Z_{t,u}]$. Note that $g_{1,u} = \mathcal{F}_{\mathtt{K}}(u)$ and $g_{0,u} = \frac{\mathtt{K}_1}{\mathtt{K}}\mathcal{F}_{\mathtt{K}_1}(u) + \frac{\mathtt{K}_2}{\mathtt{K}}\mathcal{F}_{\mathtt{K}_2}(u)$. From the fundamental theorem of calculus, we have

$$g_{1,u} = g_{0,u} + \int_0^1 \frac{d}{dt} g_{t,u} dt. \quad (65)$$

Let $\bar{z}_i = \bar{1} - \bar{x}_i$, $\bar{\mathcal{Z}}_i = \bar{N}_i + \sqrt{\frac{t}{\mathtt{N}}}\mathbf{S}_i\bar{z} + \sqrt{\frac{1-t}{\mathtt{N}_i}}\mathbf{T}_i\bar{z}_i$. Using integration by parts formula with respect to the spreading sequences, the derivative can be simplified as follows:

$$\frac{d}{dt} g_{t,u} = \frac{1}{2\mathtt{K}\sigma^4} \left(\sum_{i=1,2} \mathbb{E}\left\langle \|\bar{\mathcal{Z}}_i\|^2 \left(\frac{1}{\mathtt{N}}\|\bar{z}\|^2 - \frac{1}{\mathtt{N}_i}\|\bar{z}_i\|^2 \right) \right\rangle_{t,u} - \sum_{i=1,2} \mathbb{E}\left\langle \left(\bar{\mathcal{Z}}_i^{(1)} \cdot \bar{\mathcal{Z}}_i^{(2)} \right) \left(\frac{1}{\mathtt{N}} \bar{z}^{(1)} \cdot \bar{z}^{(2)} - \frac{1}{\mathtt{N}_i} \bar{z}_i^{(1)} \cdot \bar{z}_i^{(2)} \right) \right\rangle_{t,u} \right). \quad (66)$$

The system with Hamiltonian $\mathtt{H}_{t,u}(\bar{x})$ has Nishimori symmetry and hence we can derive results similar to Theorem 7 and Lemma 1. In addition to these, we need one more Nishimori identity, given by

$$\mathbb{E}\left\langle \left(\bar{\mathcal{Z}}_i^{(1)} \cdot \bar{\mathcal{Z}}_i^{(2)} \right) \left(\frac{1}{\mathtt{N}} \bar{z}^{(1)} \cdot \bar{z}^{(2)} - \frac{1}{\mathtt{N}_i} \bar{z}_i^{(1)} \cdot \bar{z}_i^{(2)} \right) \right\rangle_{t,u} = \mathbb{E}\left\langle \left(\bar{N}_i \cdot \bar{\mathcal{Z}}_i \right) \left(\frac{1}{\mathtt{N}} \bar{1} \cdot \bar{z} - \frac{1}{\mathtt{N}_i} \bar{1} \cdot \bar{z}_i \right) \right\rangle_{t,u}.$$

Let

$$m_1 = \frac{1}{\mathtt{K}}\sum_{j=1}^{\mathtt{K}} x_j, \quad m_{11} = \frac{1}{\mathtt{K}_1}\sum_{j=1}^{\mathtt{K}_1} x_j, \quad m_{12} = \frac{1}{\mathtt{K}_2}\sum_{j=\mathtt{K}_1+1}^{\mathtt{K}} x_j.$$

Let $\epsilon > 0$ be fixed. Using Theorem 7, for a.e., $u > \epsilon$ and a.e., $t > 0$, we can express the derivative as

$$\begin{aligned}\frac{d}{dt}g_{t,u} = \frac{1}{2\mathtt{K}\sigma^4}\Bigg(&\sum_{i=1,2}\mathbb{E}\langle\|\bar{\mathcal{Z}}_i\|^2\rangle_{t,u}\mathbb{E}\left\langle\frac{1}{\mathtt{N}}\|\bar{z}\|^2 - \frac{1}{\mathtt{N}_i}\|\bar{z}_i\|^2\right\rangle_{t,u}\\ &-\sum_{i=1,2}\mathbb{E}\langle\bar{N}_i\cdot\bar{\mathcal{Z}}_i\rangle_{t,u}\\ &\times\mathbb{E}\left\langle\frac{\bar{z}^{(1)}\cdot\bar{z}^{(2)}}{\mathtt{N}} - \frac{\bar{z}_i^{(1)}\cdot\bar{z}_i^{(2)}}{\mathtt{N}_i}\right\rangle_{t,u}\Bigg) + o_{\mathtt{K}}(1).\end{aligned}$$

Using $\frac{1}{\mathtt{N}_i}\mathbb{E}\langle\|\bar{\mathcal{Z}}_i\|^2\rangle_{t,u} = 1$, we get

$$\begin{aligned}\frac{d}{dt}g_{t,u} &= \frac{\beta}{2\mathtt{K}\sigma^4}\sum_{i=1,2}\mathbb{E}\langle\bar{N}_i\cdot\bar{\mathcal{Z}}_i\rangle_{t,u}\mathbb{E}\langle m_1 - m_{1i}\rangle_{t,u} + o_{\mathtt{K}}(1)\\ &= \frac{\beta}{2\mathtt{K}\sigma^4}\sum_{i=1,2}\Bigg[\mathbb{E}\left\langle\bar{N}_i\cdot\left(\sqrt{\frac{t}{\mathtt{N}}}\mathbf{S}_i\bar{z} + \sqrt{\frac{1-t}{\mathtt{N}_i}}\mathbf{T}_i\bar{z}_i\right)\right\rangle_{t,u}\\ &\qquad\times\mathbb{E}\langle m_1 - m_{1i}\rangle_{t,u}\Bigg] + o_{\mathtt{K}}(1).\end{aligned}$$

Now using integration by parts formula with respect to the spreading sequences, and doing transformations similar to Section III-D2, we get for a.e., $u > \epsilon$ and a.e., $t > 0$

$$\begin{aligned}\frac{d}{dt}g_{t,u} = &\frac{1}{2N\sigma^4}\sum_{i=1,2}\mathtt{K}_i\\ &\times\frac{\mathbb{E}\langle(1-m_1)t + (1-m_{1i})(1-t)\rangle_{t,u}\mathbb{E}\langle m_1 - m_{1i}\rangle_{t,u}}{1+\frac{\beta}{\sigma^2}\mathbb{E}\langle(1-m_1)t + (1-m_{1i})(1-t)\rangle_{t,u}}\\ &+ o_{\mathtt{K}}(1).\end{aligned} \tag{67}$$

After a few manipulations, the summation in the above equation can be expressed as

$$\frac{-1}{2\mathtt{K}\sigma^2}\sum_{i=1,2}\frac{\mathtt{K}_i\mathbb{E}\langle m_1 - m_{1i}\rangle_{t,u}}{1+\frac{\beta}{\sigma^2}\mathbb{E}\langle(1-m_1)t + (1-m_{1i})(1-t)\rangle_{t,u}}.$$

Let us denote the above function as $\eta(t)$. The function $\eta(t)$ satisfies $\eta(1) = 0$ and its derivative with respect to t is given by

$$-\frac{1}{2\mathtt{K}\sigma^4}\sum_{i=1,2}\frac{\beta\mathtt{K}_i(\mathbb{E}\langle m_1 - m_{1i}\rangle_{t,u})^2}{1+\frac{\beta}{\sigma^2}\mathbb{E}\langle(1-m_1)t + (1-m_{1i})(1-t)\rangle_{t,u}} \tag{68}$$

which is always nonpositive. Therefore, $\eta(t) \geq 0$ for all $0 \leq t \leq 1$.

Bringing the $o_{\mathtt{K}}(1)$ in (67) to the left, we get for a.e., $u > \epsilon$

$$\int_0^1 \frac{d}{dt}g_{t,u} + o_{\mathtt{K}}(1) \geq 0.$$

Therefore, for a.e., $u > \epsilon$, we get

$$g_{1,u} + o_{\mathtt{K}}(1) \geq g_{0,u}.$$

Let $a > \epsilon$ be a constant. Then

$$\int_\epsilon^a g_{1,u}du + o_{\mathtt{K}}(1) \geq \int_\epsilon^a g_{0,u}du$$

which implies

$$\int_\epsilon^a \mathcal{F}_{\mathtt{K}}(u)du + o_{\mathtt{K}}(1) \geq \frac{\mathtt{K}_1}{\mathtt{K}}\int_\epsilon^a \mathcal{F}_{\mathtt{K}_1}(u)du + \frac{\mathtt{K}_2}{\mathtt{K}}\int_\epsilon^a \mathcal{F}_{\mathtt{K}_2}(u)du.$$

This in turn implies that $\lim_{\mathtt{K}\to\infty}\int_\epsilon^a \mathcal{F}_{\mathtt{K}}(u)du$ exists.

VII. Extensions

In this section, we briefly describe three variations for which our methods extend in a straightforward manner.

A. Unequal Powers

Suppose that the users transmit with unequal powers P_k

$$y_i = \frac{1}{\sqrt{\mathtt{N}}}\sum_{k=1}^{\mathtt{K}} s_{ik}\sqrt{P}_k x_k + \sigma n_i$$

with normalized average power $\frac{1}{\mathtt{K}}\sum P_k = 1$. We assume that the empirical distribution of the P_k tends to a distribution and denote the corresponding expectation by $\mathbb{E}_P[-]$.

The interpolation method can be applied as before. We interpolate between the true communication system and a decoupled one where

$$y_k' = \sqrt{P}_k x_k + \frac{1}{\sqrt{\lambda}}w_k.$$

Let $\mathcal{P}$ denote the diagonal matrix $P_k\delta_{kk'}$. The relevant posterior measure replacing (31) is now

$$\begin{aligned}p_{t,u}(\bar{x}|\bar{n},\bar{w},\bar{h},\mathbf{s}) = &\frac{1}{Z_{t,u}}\\ &\times e^{-\frac{1}{2}\|\bar{n}-\mathtt{N}^{-\frac{1}{2}}B(t)^{\frac{1}{2}}\mathbf{s}\sqrt{\mathcal{P}}(\bar{x}^0-\bar{x})\|^2 - \frac{1}{2}\|\bar{w}-\lambda(t)^{\frac{1}{2}}\sqrt{\mathcal{P}}(\bar{x}^0-\bar{x})\|^2 + h_u(\bar{x})}\end{aligned} \tag{69}$$

where $\lambda(t)$ and $B(t)$ are related as in (25). The whole analysis can again be performed in exactly the same manner with the proviso that the correct "order parameters" are now $m_1 = \frac{1}{\mathtt{N}}\sum P_k x_k$ and $q_{12} = \frac{1}{\mathtt{N}}\sum P_k x_k^{(1)}x_k^{(2)}$. One finds in place of (50)

$$\begin{aligned}&\frac{1}{2} + \mathbb{E}[f_{1,u}]\\ &\quad= -\frac{1}{2\beta} + \mathbb{E}_P\left[\int Dz\ln(2\cosh(\sqrt{P\lambda}z + P\lambda))\right]\\ &\qquad-\frac{\lambda}{2}(1+m) - \frac{1}{2\beta}\ln(1+\beta B(1-m)) + \int_0^1 R(t)dt\end{aligned}$$

where $R(t)$ has the same form as before but with the new definition of m_1. From the positivity of $R(t)$, we deduce the upper bound (23) on the capacity with $c_{RS}(m)$ given by

$$-\mathbb{E}_P\left[\int Dz\ln(\cosh(\sqrt{P\lambda}z + P\lambda))\right] + \frac{\lambda}{2}(1+m) - \frac{1}{2\beta}\ln\lambda\sigma^2.$$

B. Colored Noise

Now consider the scenario where

$$y_i = \frac{1}{\sqrt{\mathtt{N}}}\sum_{k=1}^{\mathtt{K}} s_{ik}x_k + n_i$$

with colored noise of finite memory. More precisely, we assume that the covariance matrix $\mathbb{E}[N_iN_j] = C(i,j)$ (depends on $|i-j|$) is circulant as $N \to +\infty$ and has well-defined (real) Fourier transform (the noise spectrum) $\hat{C}(\omega)$. The covariance matrix is real symmetric and thus can be diagonalized by an orthogonal matrix: $\Gamma = OCO^T$ with $OO^T = O^TO = I$. As $N \to +\infty$, the eigenvalues are well approximated by $\gamma_n \equiv \hat{C}(2\pi\frac{n}{\mathtt{N}})$. Multiplying the received signal by $\Gamma^{-1/2}O$, the input–output relation becomes

$$y_i' = (\Gamma^{-1/2}O\bar{y})_i, \qquad n_i' = (\Gamma^{-1/2}O\bar{n})_i.$$

where

$$y_i' = (\Gamma^{-1/2}O\bar{y})_i, \qquad n_i' = (\Gamma^{-1/2}O\bar{n})_i.$$

The new noise vector $\bar{n}'$ is white with unit variance, but the spreading matrix is now correlated with

$$\mathbb{E}[T_{ik}T_{jl}] = \delta_{ij}\delta_{kl}\gamma_i^{-1}. \qquad (70)$$

One may guess that this time the interpolation is done between the true system and the decoupled channels

$$y_k' = x_k + \frac{1}{\sqrt{\lambda_{\text{col}}}}w_k$$

where this time

$$\lambda_{\text{col}} = \int_0^{2\pi}\frac{d\omega}{2\pi}\frac{B}{\hat{C}(\omega)+\beta B(1-m)}.$$

Note that $\hat{C}(\omega) = 1$ when the noise is white and we get back the λ defined in (13). The interpolating system has the same posterior as in (31) but with $\lambda_{\text{col}}(t)$ and $B(t)$ related by

$$\int_0^{2\pi}\frac{d\omega}{2\pi}\frac{B(t)}{\hat{C}(\omega)+\beta B(t)(1-m)} + \lambda_{\text{col}}(t) = \int_0^{2\pi}\frac{d\omega}{2\pi}\frac{B}{\hat{C}(\omega)+\beta B(1-m)}.$$

The only difference in the subsequent analysis is in the algebraic manipulations for the term T_2 in Section III-D2. Indeed these require integrations by parts with respect to the spreading sequence which involve (70). The analog of (47) now becomes

$$T_2 = \frac{1}{\mathtt{N}}\sum_{n=1}^{N}\frac{B'(t)\mathbb{E}\langle 1-m_1\rangle_t}{2(\gamma_n+\beta B(t)\mathbb{E}\langle 1-m_1\rangle_t)} \to \int_0^{2\pi}\frac{d\omega}{2\pi}\frac{B'(t)\mathbb{E}\langle 1-m_1\rangle_t}{2(S(\omega)+\beta B(t)\mathbb{E}\langle 1-m_1\rangle_t)}d\omega. \qquad (71)$$

This finally leads to the bound on capacity with $c_{RS}(m)$ given by

$$-\int Dz\ln(\cosh(\sqrt{\lambda_{\text{col}}}z+\lambda_{\text{col}})) + \frac{\lambda_{\text{col}}}{2}(1+m) + \frac{1}{2\beta}\int_0^{2\pi}\frac{d\omega}{2\pi}\ln\frac{\hat{C}(\omega)}{\hat{C}(\omega)+\beta(1-m)}.$$

C. Gaussian Input

The interpolation method also works for nonbinary inputs. Here we consider the simplest case of Gaussian inputs with distribution (17) (which achieves the maximum of the mutual information for any symmetric s_{ik}). Here we outline the necessary changes in the analysis.

The interpolation is done as explained in Section II-D except that (27) is multiplied by the Gaussian distribution (17). In (28), we also have to include this Gaussian factor and the sum over $\bar{x}^0$ is replaced by an integral. Then, as in Section III-A, we do the change of variables $\bar{y} \to B(t)^{-1/2}\bar{n} + \mathtt{N}^{-1/2}\mathbf{s}\bar{x}^0$ and $\bar{y}' \to \lambda(t)^{-1/2}\bar{w} + \bar{x}^0$. The posterior measure used for the interpolation is given by

$$p_{t,u}(\bar{x}|\bar{n},\bar{w},\bar{h},\mathbf{s}) = \frac{1}{Z_{t,u}(2\pi)^{\frac{\mathtt{N}}{2}}}e^{-\frac{\|\bar{x}\|^2}{2}} \times e^{-\frac{1}{2}\|\bar{n}-\mathtt{N}^{-\frac{1}{2}}B(t)^{\frac{1}{2}}\mathbf{s}(\bar{x}^0-\bar{x})\|^2-\frac{1}{2}\|\bar{w}-\lambda(t)^{\frac{1}{2}}(\bar{x}^0-\bar{x})\|^2+h_u(\bar{x})}. \qquad (72)$$

The quantity we have to compute is given by

$$\lim_{\mathtt{K}\to+\infty}\lim_{u\to 0}\mathbb{E}[f_{1,u}(\bar{N},\bar{W},\bar{H},\mathbf{S},\bar{x}^0)].$$

The main difference is that now the expectation $\mathbb{E}$ is with respect to the Gaussian vector $\bar{x}^0$. The algebra is done as in Section III except that x_k^0 is not set to one, z_k is replaced by $x_k^0 - x_k$, and the correct order parameters are $m_1 = \frac{1}{\mathtt{K}}\sum x_k^0x_k$ and $q_{12} = \frac{1}{\mathtt{K}}\sum x_k^{(1)}x_k^{(2)}$.

The interpolation method then yields in place of (50)

$$\frac{1}{2} + \mathbb{E}[f_{1,u}] = -\frac{1}{2\beta} - \frac{1}{2}\ln(1+\lambda) - \frac{1}{2\beta}\ln(1+\beta B(1-m)) + \frac{\lambda}{2}(1-m) + \int_0^1 R(t)dt + O(\sqrt{u})$$

where $R(t)$ is the same function as before but with new definition of m_1. Again the positivity of $R(t)$ implies that the replica solution is an upper bound to the capacity.

VIII. Concluding Remarks

In this contribution, we have shown that the capacity of binary input CDMA system with random spreading is upper bounded by the formula conjectured by Tanaka using replica method. The approach we follow is by developing an interpolation method for this system. This idea has its origins in statistical mechanics and has been applied to Gaussian energy models. The current system is very much different from those models and the proof we develop is also significantly different. In fact, this model is closer to the Hopfield model for neural networks, for which the interpolation method is still an open problem.

We also show that the capacity and the free energy functions concentrate around their average in the large-system limit. In addition, we prove a weak concentration for the magnetization for a system which is slightly perturbed using a Gaussian field. It might be interesting to show a similar result for the CDMA system itself which has some implications towards proving the concentration of the BER. We also show the independence of the capacity from the spreading sequence distributions in the large-system limit.

We expect that the powerful probabilistic tools used here have applications for other similar situations in communication systems. We have shown some of the extensions here but there are many other cases like constellations other than binary, CDMA with LDPC coded communication to name a few, to which this method can be applied. In all these cases, we can prove an upper bound on the capacity. The most interesting and also important open problem is to prove the lower bound. This seems to be a difficult problem and again the standard techniques fail. Other important problems are proving the conjectures related to the BER of various decoders.

Appendix I
Concentration Proofs

A. Probabilistic Tools

Our proofs rely on a general concentration theorem for suitable Lipschitz functions of many Gaussian random variables [10], [31, Th. 2.2.4] and this is why we need Gaussian signature sequences. In the version that we use here we need functions that are Lipschitz with respect to the Euclidean distance. More precisely, we say that a function $f : \mathbb{R}^M \to \mathbb{R}$ is a Lipschitz function with constant L_M if for all $(\bar{u}, \bar{v}) \in \mathbb{R}^M \times \mathbb{R}^M$

$$|f(\bar{u}) - f(\bar{v})| \leq L_M \|\bar{u} - \bar{v}\|.$$

When another distance is used the function will still be Lipschitz but one has to carefully keep track of the possibly qualitatively different M dependence.

Theorem 8 (Concentration of Lipschitz Function of Gaussian Random Variables [10]): Let $\bar{U} = (U_1, \dots, U_M)$ be a vector of M i.i.d. random variables distributed as $\mathcal{N}(0, v^2)$. Let $f : \mathbb{R}^M \to \mathbb{R}$ be a Lipschitz function with respect to the Euclidean distance, with Lipschitz constant L_M. Then, f satisfies

$$\mathbb{P}[|f(\bar{u}) - \mathbb{E}[f(\bar{u})]| \geq t] \leq 2e^{-\frac{t^2}{4v^2 L_M^2}}.$$

In our application, it will not be possible to apply directly this theorem because the relevant functions (capacity and free energy) are Lipschitz only on a subset $G \subset \mathbb{R}^M$. It turns out that the measure of the complement G^c is negligible as $M \to +\infty$. For the "good part" of the function supported on G, we will use the following result of McShane and Whitney.

Theorem 9 (Lipschitz Extension [32]): Let $f : G \to \mathbb{R}$ be Lipschitz over $G \subset \mathbb{R}^M$ with constant L_M. Then, there exists an extension $g : \mathbb{R}^M \to \mathbb{R}$ such that $g|_G = f$ which is Lipschitz with the same constant over the whole of $\mathbb{R}^M$.

From these two theorems, we can prove the following.

Lemma 7 (Concentration of Almost Lipschitz Function): Let f and g be as in Theorem 9. Assume $0 \in G$ and $\mathbb{E}[f(\bar{u})^2] \leq C^2$, $f(0)^2 \leq C^2$ for some positive number C. Then, for

$$\frac{t}{2} \geq 3\left(C + v\sqrt{ML_M}\right)\sqrt{\mathbb{P}(G^c)}$$

we have

$$\mathbb{P}[|f(\bar{u}) - \mathbb{E}[f(\bar{u})]| \geq t] \leq 2e^{-\frac{t^2}{16v^2 L_M^2}} + \mathbb{P}[G^c].$$

Proof: We drop the $\bar{u}$ dependence to lighten the notation. Notice that $0 \in G$ implies $f(0) = g(0)$. Thus, $g(0)^2 \leq C^2$. Also, since g is Lipschitz on the whole of $\mathbb{R}^M$

$$\begin{aligned}\mathbb{E}[g^2] &\leq 2(g(0)^2 + \mathbb{E}[(g - g(0))^2]) \\ &\leq 2(C^2 + L_M \mathbb{E}[\|\bar{u}^2\|]) \\ &= 2(C^2 + Mv^2 L_M).\end{aligned}$$

Furthermore, on G, we have $g = f$, so by the Cauchy–Schwartz inequality

$$\begin{aligned}|\mathbb{E}[g - f]| &= |\mathbb{E}[(g - f)1_{G^c}]| \\ &\leq (\mathbb{E}[g^2]^{1/2} + \mathbb{E}[f^2]^{1/2})\sqrt{\mathbb{P}[G^c]} \\ &\leq (C + \sqrt{2}(C^2 + Mv^2 L_M)^{1/2})\sqrt{\mathbb{P}[G^c]} \\ &\leq 3(C + v\sqrt{ML_M})\sqrt{\mathbb{P}[G^c]} \leq \frac{t}{2}.\end{aligned}$$

Moreover

$$\begin{aligned}&\mathbb{P}[|f - \mathbb{E}f| \geq t] \\ &\quad = \mathbb{P}[|g - \mathbb{E}f| \geq t \mid \bar{U} \in G]\mathbb{P}[G] \\ &\qquad + \mathbb{P}[|f - \mathbb{E}f| \geq t \mid \bar{U} \in G^c]\mathbb{P}[G^c] \\ &\quad \leq \mathbb{P}[|g - \mathbb{E}g| \geq t - |\mathbb{E}g - \mathbb{E}f|] + \mathbb{P}[G^c].\end{aligned}$$

The claim follows by combining

$$\mathbb{P}[|g - \mathbb{E}g| \geq t - |\mathbb{E}g - \mathbb{E}f|] \leq \mathbb{P}\left[|g - \mathbb{E}g| \geq \frac{t}{2}\right]$$

with Theorem 8. □

In order to prove Theorems 1 and 2, it will be sufficient to find suitable sets G with measure nearly equal to one (as $M \to +\infty$), on which the capacity and free energy have a Lipschitz constant $L_M \to 0$.

B. Proofs of Theorems 1 and 2

For the proofs, it is convenient to reformulate the statements of the theorems as follows. Let $\bar{1}$ be the $\mathtt{K}$-dimensional vector $(1, \dots, 1)$, $\mathbf{s}^0$ be the $\mathtt{K} \times \mathtt{N}$ matrix with elements $s_{ik} x_k^0$. We set $p_{\bar{X}}^0(\bar{x}) = \prod_{k=1}^{\mathtt{K}} p_X(x_k x_k^0)$ and consider the partition function

$$Z'(\bar{n}, \mathbf{s}^0) = \sum_{\bar{x}} p_{\bar{X}}^0(\bar{x}) e^{-\frac{1}{2\sigma^2}\|\mathtt{N}^{-1/2}\mathbf{s}^0(\bar{x} - \bar{1}) - \sigma\bar{n}\|^2} \quad (73)$$

where we recall that $\bar{n} = (n_1, \dots, n_{\mathtt{N}})$ are independent Gaussian variables $\mathcal{N}(0, 1)$. Notice that due to the invariance of the distribution of S_{ik} under the transformation $S_{ik} \to x_k^0 S_{ik}$

$$\mathbb{E}_{\bar{N}, \mathbf{S}}[\ln Z'(\bar{N}, \mathbf{S}^0)] = \mathbb{E}_{\bar{N}, \mathbf{S}}[\ln Z'(\bar{N}, \mathbf{S})].$$

The statements of Theorems 1 and 2 are equivalent to

$$\mathbb{P}\left[\left|\sum_{\bar{x}^0} p_{\bar{X}}^0(\bar{x}^0)\mathbb{E}_{\bar{N}}[\ln Z'(\bar{N}, \mathbf{S}^0)] - \mathbb{E}_{\bar{N}, \mathbf{S}}[\ln Z'(\bar{N}, \mathbf{S})]\right| \geq \epsilon \mathtt{K}\right] \leq 3e^{-\alpha_1 \mathtt{K}} \quad (74)$$

and

$$\mathbb{P}\left[\left|\sum_{\bar{x}^0} p^0_{\bar{X}}(\bar{x}^0)\ln Z'(\bar{N},\mathbf{S}^0) - \mathbb{E}_{\bar{N},\mathbf{S}}[\ln Z'(\bar{N},\mathbf{S})]\right| \geq \epsilon\mathtt{K}\right] \leq 3e^{-\alpha_2\sqrt{\mathtt{K}}}. \quad (75)$$

To see this use the change of variable $\bar{y} = \mathtt{N}^{-1/2}\mathbf{s}\bar{x}^0 + \sigma\bar{n}$ followed by $x_k \to x_k x^0_k$ in the partition function summation (6).

C. Proof of (74)

Let B be a positive constant to be chosen later and define

$$G = \{\mathbf{s} : \text{for all } \bar{x}, \bar{x}^0, ||\mathbf{s}^0(\bar{x} - \bar{1})||^2 \leq B\mathtt{N}\}.$$

Lemma 8: We have the following estimate for the measure of G^c:

$$\mathbb{P}(G^c) \leq 3^{\mathtt{K}}2^{\frac{\mathtt{N}}{2}}e^{-\frac{B}{16\beta}}.$$

Proof: First, notice that for any given $\bar{x}$

$$\frac{1}{\sqrt{\mathtt{K}}}\sum_{k=1}^{\mathtt{K}} S^0_{ik}(x_k - 1), \qquad i = 1, \ldots \mathtt{N}$$

are independent Gaussian random variables with zero mean and variance (say a^2) smaller than 4. Thus, the identity

$$\int dx \frac{e^{-\frac{x^2}{2a^2}}}{\sqrt{2\pi a^2}}e^{\frac{x^2}{16}} = \sqrt{\frac{8}{8-a^2}}$$

implies (because $a^2 \leq 4$)

$$\mathbb{E}[e^{\frac{1}{16\mathtt{K}}||\mathbf{S}^0(\bar{x}-\bar{1})||^2}] \leq 2^{\frac{\mathtt{N}}{2}}.$$

Then, from the Markov inequality, for any $\bar{x}$

$$\mathbb{P}(||\mathbf{S}^0(\bar{x} - \bar{1})||^2 \geq B\mathtt{N}) \leq 2^{\frac{\mathtt{N}}{2}}e^{-\frac{B\mathtt{N}}{16\mathtt{K}}} = 2^{\frac{\mathtt{N}}{2}}e^{-\frac{B}{16\beta}}.$$

The result of the lemma then follows from the union bound over $3^{\mathtt{K}}$ possible $\bar{x}^0 - \bar{x}$ vectors. □

We will apply Lemma 8 to

$$f(\mathbf{s}) = \frac{1}{\mathtt{K}}\sum_{\bar{x}^0} p^0_{\bar{X}}(\bar{x}^0)\mathbb{E}_{\bar{N}}[\ln Z'(\bar{N},\mathbf{s}^0)]$$

for a suitable choice of B. In the following, the matrix $\mathbf{s}$ is to be thought as a vector with $\mathtt{KN}$ components and norm

$$||\mathbf{s}|| = \left(\sum_{i=1}^{\mathtt{N}}\sum_{k=1}^{\mathtt{K}} s^2_{ik}\right)^{\frac{1}{2}}.$$

Clearly $0 \in G$ and $f(0)^2 = (\frac{1}{\mathtt{K}}\mathbb{E}_{\bar{N}}[\frac{1}{2}||\bar{N}||^2])^2 = 1/4\beta^2$. Also, it is evident that $\ln Z'(\bar{n},\mathbf{s}^0) \leq 0$. On the other hand, restricting the sum in the partition function to $\bar{x} = \bar{1}$, we have

$$\frac{1}{\mathtt{K}}\sum_{\bar{x}^0} p^0_{\bar{X}}(\bar{x}^0)\mathbb{E}_{\bar{N}}[\ln Z'(\bar{N},\mathbf{s}^0)]$$
$$\geq -\frac{1}{2\sigma^2\mathtt{K}}\mathbb{E}_{\bar{N}}[\sigma^2||\bar{N}||^2] - \frac{1}{\mathtt{K}}H(\bar{X})$$
$$\geq -\frac{\mathtt{N}}{2\mathtt{K}} - \ln 2.$$

Therefore, we have

$$\mathbb{E}_{\mathbf{S}}[f(\mathbf{S})^2] \leq \left(\frac{1}{2\beta} + \ln 2\right)^2.$$

Let us now compute the Lipschitz constant.

Lemma 9: The quantity

$$\mathtt{K}^{-1}\mathbb{E}_{\bar{N}}[\sum_{\bar{x}^0} p^0_{\bar{X}}(\bar{x}^0)\ln Z'(\bar{N},\mathbf{s}^0)]$$

is Lipschitz on G, with Lipschitz constant

$$L_{\mathtt{K}} = \sigma^{-2}2\sqrt{\beta}\mathtt{K}^{-1}\left(\sqrt{B} + \sigma\sqrt{\mathtt{N}}\right).$$

Proof: The exponent of the partition function is

$$\mathtt{H}(\bar{n},\mathbf{s}^0,\bar{x}) = -\frac{1}{2\sigma^2}||\mathtt{N}^{-1/2}\mathbf{s}^0(\bar{x} - \bar{1}) - \sigma\bar{n}||^2. \quad (76)$$

In Appendix I-F, we show that for $(\mathbf{s},\mathbf{t}) \in G \times G$

$$|\mathtt{H}(\bar{n},\mathbf{s}^0,\bar{x}) - \mathtt{H}(\bar{n},\mathbf{t}^0,\bar{x})| \leq \sigma^{-2}2\sqrt{\beta}\left(\sqrt{B} + \sigma||\bar{n}||\right)||\mathbf{s} - \mathbf{t}||. \quad (77)$$

For any $(\mathbf{s},\mathbf{t}) \in G \times G$, using (77) together with

$$\mathtt{H}(\bar{n},\mathbf{s}^0,\bar{x}) \leq \mathtt{H}(\bar{n},\mathbf{t}^0,\bar{x}) + |\mathtt{H}(\bar{n},\mathbf{s}^0,\bar{x}) - \mathtt{H}(\bar{n},\mathbf{t}^0,\bar{x})|$$

we get

$$\ln\frac{\sum_{\bar{x}} p^0_{\bar{X}}(\bar{x})\exp(\mathtt{H}(\bar{n},\mathbf{s}^0,\bar{x}))}{\sum_{\bar{x}} p^0_{\bar{X}}(\bar{x})\exp(\mathtt{H}(\bar{n},\mathbf{t}^0,\bar{x}))}$$
$$\leq \ln\frac{\sum_{\bar{x}} p^0_{\bar{X}}(\bar{x})\exp(|\mathtt{H}(\bar{n},\mathbf{s}^0,\bar{x}) - \mathtt{H}(\bar{n},\mathbf{t}^0,\bar{x})| + \mathtt{H}(\bar{n},\mathbf{t}^0,\bar{x}))}{\sum_{\bar{x}} p^0_{\bar{X}}(\bar{x})\exp(\mathtt{H}(\bar{n},\mathbf{t}^0,\bar{x}))}$$
$$\leq \sigma^{-2}2\sqrt{\beta}\left(\sqrt{B} + \sigma||\bar{n}||\right)||\mathbf{s} - \mathbf{t}||.$$

Therefore, taking the expectation over the noise, we get

$$\left|\sum_{\bar{x}^0} p^0_{\bar{X}}(\bar{x}^0)\mathbb{E}_{\bar{N}}[\ln Z'(\bar{N},\mathbf{s}^0)] - \sum_{\bar{x}^0} p^0_{\bar{X}}(\bar{x}^0)\mathbb{E}_{\bar{N}}[\ln Z'(\bar{N},\mathbf{t}^0)]\right|$$
$$\leq \sigma^{-2}2\sqrt{\beta}(\sqrt{B} + \sigma\mathbb{E}[||\bar{N}||])||\mathbf{s} - \mathbf{t}||$$
$$\leq \sigma^{-2}2\sqrt{\beta}(\sqrt{B} + \sigma\mathbb{E}[||\bar{N}||^2]^{1/2})||\mathbf{s} - \mathbf{t}||$$

which yields the Lipschitz constant of the lemma. □

To prove (74) choose $B = 32\beta(2\mathtt{K} + \mathtt{N})$. From Lemma 8, we get

$$\mathbb{P}(G^c) \leq 3^{\mathtt{K}}2^{\frac{\mathtt{N}}{2}}e^{-2(2\mathtt{K}+\mathtt{N})} \leq e^{-(2\mathtt{K}+\mathtt{N})}.$$

Note that Lemma 7 is valid only for

$$\epsilon \geq 6\left(C + \mathtt{N}\sqrt{L_{\mathtt{K}}}\right)\sqrt{\mathbb{P}(G^c)} = O(e^{-\mathtt{K}})$$

which is equivalent to saying that $\mathtt{K} \geq O(|\ln\epsilon|)$. Therefore, for $\mathtt{K} \geq O(|\ln\epsilon|)$, Lemma 7 implies that

$$\mathbb{P}\left[\frac{1}{\mathtt{K}}\Big|\sum_{\bar{x}^0} p^0_{\bar{X}}(\bar{x}^0)\mathbb{E}_{\bar{N}}[\ln Z'(\bar{N},\mathbf{S}^0)] - \mathbb{E}_{\bar{N},\mathbf{S}}[\ln Z'(\bar{N},\mathbf{S})]\Big| \geq \epsilon\right] \leq 2e^{-\frac{\epsilon^2}{16L_{\mathtt{K}}^2}} + e^{-(2\mathtt{K}+\mathtt{N})} \leq \begin{cases} 3e^{-\frac{\epsilon^2}{16L_{\mathtt{K}}^2}}, & \text{if } \epsilon \leq 4L_{\mathtt{K}}\sqrt{2\mathtt{K}+\mathtt{N}} \\ 3e^{-(2\mathtt{K}+\mathtt{N})}, & \text{otherwise.}\end{cases}$$

For $B = 32\beta(2\mathtt{K}+\mathtt{N})$, we obtain

$$\frac{1}{16L_{\mathtt{K}}^2} \geq \frac{\sigma^4\mathtt{K}}{128(64\beta^2+32\beta+\sigma^2)}$$

and

$$4L_{\mathtt{K}}\sqrt{2\mathtt{K}+\mathtt{N}} = \frac{8\sqrt{\beta}}{\sigma^2\mathtt{K}}\left(\sqrt{32\beta(2\mathtt{K}+\mathtt{N})}+\sigma\sqrt{\mathtt{N}}\right)\sqrt{2\mathtt{K}+\mathtt{N}} = \frac{8\sqrt{2\beta+1}}{\sigma^2\sqrt{\beta}}\left(\sqrt{32\beta(2\beta+1)}+\sigma\right).$$

To summarize, we have shown that, for $\mathtt{K} \geq O(|\ln\epsilon|)$

$$\mathbb{P}[|\imath(\bar{X};\bar{Y}|\mathbf{S}) - \mathbb{E}_{\mathbf{S}}[\imath(\bar{X};\bar{Y}|\mathbf{S})]| \geq \epsilon\mathtt{K}] \leq 3e^{-\alpha_1\mathtt{K}}$$

where α_1 is defined as

$$\begin{cases} \frac{\sigma^4\epsilon^2}{128(64\beta^2+32\beta+\sigma^2)}, & \text{if } \epsilon \leq \frac{8\sqrt{2\beta+1}}{\sigma^2\sqrt{\beta}}\left(\sqrt{32\beta(2\beta+1)}+\sigma\right) \\ 2+\frac{1}{\beta}, & \text{otherwise.}\end{cases}$$

D. *Proof of (75)*

This case is more cumbersome but the ideas are the same. We choose the set G as

$$G = \left\{(\bar{n},\mathbf{s}) : \max_i |n_i| \leq \sqrt{A},\ ||\mathbf{s}^0(\bar{x}-\bar{1})||^2 \leq B\mathtt{N}\ \forall\bar{x}\right\}$$

where A and B will be chosen appropriately later on. For Gaussian noise $\mathbb{P}[|n_i| \geq \sqrt{A}] \leq 4e^{-\frac{A}{4}}$, therefore, from the union bound $\mathbb{P}(\max_i |n_i| \geq \sqrt{A}) \leq 4\mathtt{N}e^{-\frac{A}{4}}$. Using Lemma 8, we obtain an estimate for the measure of G^c

$$\mathbb{P}[G^c] \leq 4Ne^{-\frac{A}{4}} + 3^{\mathtt{K}}2^{\frac{\mathtt{N}}{2}}e^{-\frac{B}{16\beta}}.$$

The goal is to apply Lemma 7 to $f(\bar{n},\mathbf{s}) = \ln Z'(\bar{n},\mathbf{s}^0)$ defined on $\mathbb{R}^{\mathtt{K}} \times \mathbb{R}^{\mathtt{NK}}$.

Clearly, $(0,0) \in G$, $f(0,0) = \ln 2$ and by the same argument as before, we have $\mathbb{E}[f(\bar{N},\mathbf{S})^2] \leq (\frac{1}{2\beta}+\ln 2)^2 = C^2$. It remains to compute the Lipschitz constant.

Lemma 10: The free energy $\mathtt{K}^{-1}\ln Z'(\bar{n},\mathbf{s}^0)$ is Lipschitz on G with Lipschitz constant

$$L_{\mathtt{K}} = \sigma^{-2}\left(2\sqrt{\beta}+\sigma\right)\mathtt{K}^{-1}\left(\sigma\sqrt{\mathtt{N}A}+\sqrt{B}\right).$$

Proof: For the Hamiltonian (76), we show in Appendix I-E

$$|\mathtt{H}\left(\bar{n},\mathbf{s}^0,\bar{x}\right) - \mathtt{H}\left(\bar{n},\mathbf{t}^0,\bar{x}\right)| \leq \sigma^{-2}\left(2\sqrt{\beta}+\sigma\right) \times \left(\sigma\sqrt{\mathtt{N}A}+\sqrt{B}\right)||(\bar{n},\mathbf{s})-(\bar{m},\mathbf{t})||. \quad (78)$$

Then, proceeding in the same way as in the proof of Lemma 9, we get

$$|\ln Z'\left(\bar{n},\mathbf{s}^0\right) - \ln Z'\left(\bar{m},\mathbf{t}^0\right)| \leq \sigma^{-2}\left(2\sqrt{\beta}+\sigma\right) \times \left(\sigma\sqrt{\mathtt{N}A}+\sqrt{B}\right)||(\bar{n},\mathbf{s})-(\bar{m},\mathbf{t})||.$$

□

To complete the proof of (75), let us choose $A = \sqrt{\mathtt{N}}/\sigma^2$ and $B = 32\beta(2\mathtt{K}+\mathtt{N})$. This implies that $\mathbb{P}(G^c) = O(e^{-c(\beta,\sigma)\sqrt{\mathtt{K}}})$ for some $c_1(\beta,\sigma) > 0$. Therefore, Lemma 7 implies that for $\mathtt{K} \geq O(|\ln\epsilon|)$

$$\mathbb{P}[|f(\bar{Y},\mathbf{S}) - \mathbb{E}_{\bar{Y},\mathbf{S}}[f(\bar{Y},\mathbf{S})]| \geq \epsilon] \leq 2e^{-\frac{\epsilon^2}{16L_{\mathtt{K}}^2}} + e^{-c_1(\beta,\sigma)\sqrt{\mathtt{K}}}.$$

Note that $\frac{1}{16L_{\mathtt{K}}^2} = O(\sqrt{\mathtt{K}})$. This implies that there exists a constant $c_2(\beta,\sigma) > 0$ such that $\frac{1}{L_{\mathtt{K}}^2} \geq c_2(\beta,\sigma)\sqrt{\mathtt{K}}$. Therefore

$$\mathbb{P}[|f(\bar{Y},\mathbf{S}) - \mathbb{E}_{\bar{Y},\mathbf{S}}[f(\bar{Y},\mathbf{S})]| \geq \epsilon] \leq 3e^{-\alpha_2\sqrt{\mathtt{K}}}$$

where α_2 is defined as

$$\begin{cases} \epsilon^2 c_2(\beta,\sigma), & \text{if } \epsilon^2 \leq \frac{c_1(\beta,\sigma)}{c_2(\beta,\sigma)} \\ c_1(\beta,\sigma), & \text{otherwise.}\end{cases}$$

The constants derived are sufficient for our purpose, but they are by no means optimal.

E. *Proof of (78)*

Let $\bar{n}$ and $\bar{m}$ be two noise realizations and $\mathbf{s}$ and $\mathbf{t}$ two spreading sequences all belonging to the appropriate set G. Let $\bar{y} = \bar{x} - \bar{1}$. First, we expand the Euclidean norms

$$\begin{aligned}&||\mathtt{N}^{-\frac{1}{2}}\mathbf{s}^0\bar{y} - \sigma\bar{n}||^2 - ||\mathtt{N}^{-\frac{1}{2}}\mathbf{t}^0\bar{y} - \sigma\bar{m}||^2 \\ &= \sigma^2||\bar{n}||^2 - \sigma^2||\bar{m}||^2 + \mathtt{N}^{-1}(||\mathbf{s}^0\bar{y}||^2 - ||\mathbf{t}^0\bar{y}||^2) \\ &\quad -2\sigma\mathtt{N}^{-\frac{1}{2}}(\bar{n}\cdot\mathbf{s}^0\bar{y} - \bar{m}\cdot\mathbf{t}^0\bar{y}) \\ &= \sigma^2(\bar{n}-\bar{m})\cdot(\bar{n}+\bar{m}) + \mathtt{N}^{-1}(\mathbf{s}^0\bar{y} - \mathbf{t}^0\bar{y}) \\ &\quad \cdot(\mathbf{s}^0\bar{y} + \mathbf{t}^0\bar{y}) - 2\sigma\mathtt{N}^{-\frac{1}{2}}(\bar{n}-\bar{m}) \\ &\quad \cdot\mathbf{s}^0\bar{y} - 2\sigma\mathtt{N}^{-\frac{1}{2}}\bar{m}\cdot(\mathbf{s}^0\bar{y} - \mathbf{t}^0\bar{y}).\end{aligned}$$

We estimate each of the four terms on the right-hand side of the last equality. By Cauchy–Schwartz, the first term is bounded by

$$\begin{aligned}&\sigma^2||\bar{n}-\bar{m}||||\bar{n}+\bar{m}|| \\ &\leq \sigma^2\sqrt{\mathtt{N}}\max_i(|n_i|+|m_i|)||\bar{n}-\bar{m}|| \\ &\leq \sigma^2 2\sqrt{\mathtt{N}A}||\bar{n}-\bar{m}||.\end{aligned}$$

Using Cauchy–Schwartz and $\|(\mathbf{s}^0-\mathbf{t}^0)\bar{y}\| \le \|\mathbf{s}^0-\mathbf{t}^0\|\|\bar{y}\|$, where $\|\mathbf{s}^0-\mathbf{t}^0\| = \|\mathbf{s}-\mathbf{t}\|$ is the (Hilbert–Schmidt) norm

$$\|\mathbf{s}-\mathbf{t}\| = \left(\sum_{i=1}^{\mathtt{N}}\sum_{l=1}^{\mathtt{K}}(s_{il}-t_{il})^2\right)^{1/2}$$

we can upper bound the second term as

$$\begin{aligned}&\mathtt{N}^{-1}\|\mathbf{s}-\mathbf{t}\|\|\bar{y}\|\left(\|\mathbf{s}^0\bar{y}\|+\|\mathbf{t}^0\bar{y}\|\right)\\ &\le \mathtt{N}^{-1}\|\mathbf{s}-\mathbf{t}\|2\sqrt{\mathtt{K}}2\sqrt{B\mathtt{N}}\\ &=4\sqrt{\beta B}\|\mathbf{s}-\mathbf{t}\|.\end{aligned}$$

Similarly, the third term is bounded by

$$\begin{aligned}2\sigma\mathtt{N}^{-\frac{1}{2}}\|\bar{n}-\bar{m}\|\|\mathbf{s}^0\bar{y}\| &\le 2\sigma\mathtt{N}^{-\frac{1}{2}}\|\bar{n}-\bar{m}\|\sqrt{B\mathtt{N}}\\ &=2\sigma\sqrt{B}\|\bar{n}-\bar{m}\|\end{aligned}$$

and the fourth one by

$$\begin{aligned}2\sigma\mathtt{N}^{-\frac{1}{2}}\|\bar{m}\|\|\mathbf{s}-\mathbf{t}\|\|\bar{y}\| &\le 2\sigma\mathtt{N}^{-\frac{1}{2}}\sqrt{\mathtt{N}A}\|\mathbf{s}-\mathbf{t}\|2\sqrt{\mathtt{K}}\\ &=4\sigma\sqrt{\beta\mathtt{N}A}\|\mathbf{s}-\mathbf{t}\|.\end{aligned}$$

Collecting all four estimates, we obtain

$$\begin{aligned}&\|\mathtt{N}^{-\frac{1}{2}}\mathbf{s}^0(\bar{x}-\bar{1})-\sigma\bar{n}\|^2-\|\mathtt{N}^{-\frac{1}{2}}\mathbf{t}^0(\bar{x}-\bar{1})-\sigma\bar{m}\|^2\\ &\quad\le 2\sigma(\sigma\sqrt{\mathtt{N}A}+\sqrt{B})\|\bar{n}-\bar{m}\|\\ &\qquad+4\sqrt{\beta}(\sigma\sqrt{\mathtt{N}A}+\sqrt{B})\|\mathbf{s}-\mathbf{t}\|\\ &\quad\le 2(2\sqrt{\beta}+\sigma)(\sigma\sqrt{\mathtt{N}A}+\sqrt{B})\|(\bar{n},\mathbf{s})-(\bar{m},\mathbf{t})\|\end{aligned}$$

where the norm in the last term is the Euclidean norm in $\mathbb{R}^{\mathtt{N}}\times\mathbb{R}^{\mathtt{NK}}$.

F. Proof of (77)

Let $\mathbf{s}$ and $\mathbf{t}$ be two spreading sequences both belonging to the appropriate G. Let $\bar{y}=\bar{x}-\bar{1}$. Following similar steps as in the previous paragraph with $\bar{n}=\bar{m}$ the result can be read off

$$\begin{aligned}&\|\mathtt{N}^{-\frac{1}{2}}\mathbf{s}^0\bar{y}-\sigma\bar{n}\|^2-\|\mathtt{N}^{-\frac{1}{2}}\mathbf{t}^0\bar{y}-\sigma\bar{n}\|^2\\ &\qquad\le 4\sqrt{\beta}\left(\sqrt{B}+\sigma\|\bar{n}\|\right)\|\mathbf{s}-\mathbf{t}\|.\end{aligned}$$

G. Proof of Theorem 3

The idea of this proof is based on [27] and [28].

Proof: Here, for simplicity of notation and without loss of generality, we assume the noise variance to be 1 and the fourth moment of spreading sequences to be less than 1. For $l\le\mathtt{K}$, let ϕ_l be the sigma algebra generated by $\{S_{ik}:1\le i\le\mathtt{N},1\le k\le l\}$, and set

$$f_l=\frac{1}{\mathtt{K}}\mathbb{E}\left[\imath(\bar{X};\bar{Y}|\mathbf{S})|\phi_l\right]\quad\text{and}\quad\psi_l=f_l-f_{l-1}.$$

Then

$$\frac{1}{\mathtt{K}^2}\mathbb{E}(\imath(\bar{X};\bar{Y}|\mathbf{S})-I(\bar{X};\bar{Y}\mid\mathbf{S}))^2=\sum_{l=1}^{\mathtt{K}}\mathbb{E}[\psi_l^2].$$

The goal is to bound each term in this sum by $O\left(\frac{1}{\mathtt{K}^2}\right)$. Let us recall that the relation between mutual information and free energy is given by (9)

$$\begin{aligned}\frac{1}{\mathtt{K}}\imath(\bar{X};\bar{Y}|\mathbf{s}) &= -\frac{1}{2\beta}\\ &\quad-\frac{1}{\mathtt{K}}\mathbb{E}_{\bar{N}}\left[\sum_{\bar{x}^0}p_{\bar{X}}(\bar{x}^0)\ln\sum_{\bar{x}}p_{\bar{X}}(\bar{x})e^{\mathtt{H}(\bar{x}^0,\bar{x})}\right]\end{aligned}$$

where

$$\begin{aligned}&\mathtt{H}(\bar{x}^0,\bar{x})\\ &\quad=-\frac{1}{2}\sum_i\left(n_i+\frac{1}{\sqrt{\mathtt{N}}}\sum_k s_{ik}(x_k^0-x_k)\right)^2\\ &\quad=-\frac{1}{2}\sum_i n_i^2-\frac{1}{\sqrt{\mathtt{N}}}\sum_{i,k}n_is_{ik}x_k^0\\ &\qquad-\frac{1}{2\mathtt{N}}\sum_i\left(\sum_k s_{ik}(x_k^0-x_k)\right)^2\\ &\qquad+\frac{1}{\sqrt{\mathtt{N}}}\sum_{ik}n_is_{ik}x_k.\end{aligned}$$

In the above expanded form, the first two terms do not involve $\bar{x}$ and hence the concentration of these terms follows very easily. Therefore, in the rest of the proof, we consider the Hamiltonian with only the remaining two terms. From now on in the notation, we do not explicitly show the dependency of $\mathtt{H}$ on $\bar{x}^0$ and $\bar{x}$. To this end, we define the following three Hamiltonians:

$$\begin{aligned}\mathtt{H}_l &= \frac{-1}{2\mathtt{N}}\sum_i\sum_{k_1\neq l,k_2\neq l}s_{ik_1}s_{ik_2}(x_{k_1}^0-x_{k_1})\\ &\qquad\times(x_{k_2}^0-x_{k_2})+\frac{1}{\sqrt{\mathtt{N}}}\sum_i\sum_{k\neq l}n_is_{ik}x_k\\ \mathtt{R}_l &= -\frac{1}{\mathtt{N}}\sum_{i,k}s_{ik}s_{il}(x_l^0-x_l)(x_k^0-x_k)\\ &\qquad+\frac{1}{2\mathtt{N}}\sum_i s_{il}^2(x_l^0-x_l)^2+\frac{1}{\sqrt{\mathtt{N}}}\sum_i n_is_{il}x_l\\ \tilde{\mathtt{H}}_l(t) &= \mathtt{H}_l+t\mathtt{R}_l\end{aligned}$$

where $t\in[0,1]$ will play the role of an interpolating parameter. We also introduce the difference of free energies associated to the Hamiltonian $\tilde{\mathtt{H}}_l(t)$ and $\mathtt{H}_l$

$$\tilde{f}_l(t)=\sum_{\bar{x}^0}p_{\bar{X}}(\bar{x}^0)(\ln Z(\tilde{\mathtt{H}}_l(t))-\ln Z(\tilde{\mathtt{H}}_l(0))).$$

In the last definition, the partition function is defined by the usual summation over all configurations $\bar{x}$.

With these definitions, we have the representation

$$\psi_l=\frac{1}{\mathtt{K}}\mathbb{E}_{\ge l+1}\tilde{f}_l(1)-\frac{1}{\mathtt{K}}\mathbb{E}_{\ge l}\tilde{f}_l(1)$$

where $\mathbb{E}_{\ge l}$ denotes expectation with respect to $\{s_{ik}\ \forall\ k\ge l\}$. Using convexity in the form of $\mathbb{E}_{\ge l+1}[\tilde{f}_l(1)]^2\le\mathbb{E}_{\ge l+1}[\tilde{f}_l(1)^2]$, it follows that

$$\begin{aligned}\mathbb{E}[\psi_l^2] &\le \frac{1}{\mathtt{K}^2}\mathbb{E}\mathbb{E}_{\ge l+1}\tilde{f}_l(1)^2+\frac{1}{\mathtt{K}^2}\mathbb{E}\mathbb{E}_{\ge l}\tilde{f}_l(1)^2\\ &\quad-\frac{2}{\mathtt{K}^2}\mathbb{E}[(\mathbb{E}_{\ge l+1}\tilde{f}_l(1)|\phi_{l-1})(\mathbb{E}_{\ge l}\tilde{f}_l(1))]\end{aligned}$$

$$= \frac{2}{\mathtt{K}^2}\mathbb{E}\tilde{f}_l(1)^2 - \frac{2}{\mathtt{K}^2}\mathbb{E}[(\mathbb{E}_{\geq l}\tilde{f}_l(1))^2] \leq \frac{2}{\mathtt{K}^2}\mathbb{E}\tilde{f}_l(1)^2.$$

Notice that $\frac{d^2}{dt^2}\tilde{f}_l(t) \geq 0$. Therefore

$$\tilde{f}_l'(0) \leq \frac{\tilde{f}_l(1) - \tilde{f}_l(0)}{1-0} \leq \tilde{f}_l'(1).$$

Since $\tilde{f}_l(0) = 0$, the above inequality implies

$$\tilde{f}_l'(0) \leq \tilde{f}_l(1) \leq \tilde{f}_l'(1).$$

We can bound $\tilde{f}_l(1)$ as

$$\mathbb{E}[\tilde{f}_l(1)^2] \leq \mathbb{E}[\tilde{f}_l'(0)^2] + \mathbb{E}[\tilde{f}_l'(1)^2].$$

Note that the bound $\tilde{f}_l(1)^2 \leq \tilde{f}_l'(1)^2$ is valid only if $\tilde{f}_l(1) \geq 0$, which need not be true. This shows that our task is reduced to a proof of $\mathbb{E}[\tilde{f}_l'(0)^2] = O(1)$, $\mathbb{E}[\tilde{f}_l'(1)^2] = O(1)$. This is a technical calculation which is proved in the next lemma. □

Lemma 11: Let $\tilde{f}_l'(t)$ be as defined in the previous lemma. Then, $\mathbb{E}[(\tilde{f}_l'(0))^2] = O(1)$, and $\mathbb{E}[(\tilde{f}_l'(1))^2] = O(1)$.

Proof: From convexity, we get

$$(\tilde{f}'(t))^2 \leq \sum_{\bar{x}^0} p_{\bar{X}}(\bar{x}^0)\langle R_l^2\rangle_{\tilde{\mathtt{H}}_l(t)} \leq 3\sum_{\bar{x}^0} p_{\bar{X}}(\bar{x}^0) \times \Bigg[\Big\langle\Big(-\frac{1}{2\mathtt{N}}\sum_i (s_{il})^2(x_l^0 - x_l)^2\Big)^2\Big\rangle_{\tilde{\mathtt{H}}_l(t)} + \Big\langle\Big(\sum_{k\neq l}\frac{1}{\mathtt{N}}\sum_i s_{ik}s_{il}(x_l^0 - x_l)(x_k^0 - x_k)\Big)^2\Big\rangle_{\tilde{\mathtt{H}}_l(t)} + \Big\langle\Big(\sum_i n_i\frac{1}{\sqrt{\mathtt{N}}}s_{il}x_l\Big)^2\Big\rangle_{\tilde{\mathtt{H}}_l(t)}\Bigg].$$

We will find a uniform bound over $\bar{x}^0$ for each term in the above sum. Let us consider a particular term in the above sum and set $x_k^0 - x_k = z_{0k}$. We use the simple bound of $z_{0k}^2 \leq 4$ in the following and hence we remove the average over $\bar{x}^0$

$$\mathbb{E}[(\tilde{f}_l'(0))^2] \leq 12 + 3\mathbb{E}\Big\langle\sum_{k_1,k_2\neq l}\frac{1}{\mathtt{N}^2}\sum_{i_1,i_2} S_{i_1k_1}S_{i_1l}S_{i_2k_2} \times S_{i_2l}z_{0k_1}z_{0k_2}z_{0l}^2\Big\rangle_{\tilde{\mathtt{H}}_l(0)} + 3\mathbb{E}\Big\langle\frac{1}{\mathtt{N}}\sum_{i_1,i_2} N_{i_1}N_{i_2}S_{i_1l}S_{i_2l}\Big\rangle_{\tilde{H}_l(0)}.$$

Since $\tilde{H}(0)$ does not depend on s_{il} and since they are symmetric random variables, in the above sums, only those terms remain where s_{il} are repeated even number of times. Therefore

$$\mathbb{E}\Big\langle\sum_{k_1,k_2\neq l}\frac{1}{\mathtt{N}^2}\sum_{i_1,i_2} S_{i_1k_1}S_{i_1l}S_{i_2k_2}S_{i_2l}z_{0k_1}z_{0k_2}z_{0l}^2\Big\rangle_{\tilde{\mathtt{H}}_l(0)} = \mathbb{E}\Big\langle\sum_{k_1,k_2\neq l}\frac{1}{\mathtt{N}^2}\sum_i S_{ik_1}S_{ik_2}S_{il}^2 z_{0k_1}z_{0k_2}z_{0l}^2\Big\rangle_{\tilde{\mathtt{H}}_l(0)}$$

$$\stackrel{(a)}{\leq} \mathbb{E}\Big\langle\sum_{k_1,k_2\neq l}\frac{1}{\mathtt{N}^2}\sum_i S_{ik_1}S_{ik_2}z_{0k_1}z_{0k_2}z_{0l}^2\Big\rangle_{\tilde{\mathtt{H}}_l(0)}$$

where (a) follows from the fact that $\mathbb{E}[S_{il}^2] \leq \sqrt{\mathbb{E}[S_{ik}^4]} \leq 1$. We claim that we can bound the last term as

$$\mathbb{E}\Big\langle\sum_{k_1,k_2\neq l}\frac{1}{\mathtt{N}^2}\sum_i S_{ik_1}S_{ik_2}z_{0k_1}z_{0k_2}z_{0l}^2\Big\rangle_{\tilde{\mathtt{H}}_l(0)} \leq \mathbb{E}\Big\langle\sum_{k_1,k_2}\frac{1}{\mathtt{N}^2}\sum_i S_{ik_1}S_{ik_2}z_{0k_1}z_{0k_2}z_{0l}^2\Big\rangle_{\tilde{\mathtt{H}}_l(0)}.$$

This follows from the fact that on the right-hand side, among the additional terms only the term corresponding to $k_1 = k_2 = l$ is nonzero which is also nonnegative. The terms corresponding to $k_1 = l, k_2 \neq l$ and $k_1 \neq l, k_2 = l$ vanish because $\mathbb{E}[S_{ik}] = 0$.

Let J be a $\mathtt{K} \times \mathtt{K}$ matrix with entries given by $J_{k_1k_2} = \frac{1}{\mathtt{N}}\sum_i S_{ik_1}S_{ik_2}$ and let $||J||$ denote its largest singular value. From [33], we have $||J|| \to (1+\sqrt{\beta})^2\ a.s.$. From [34, ch. 4], we even have exponential concentration of $||J||$ around $(1+\sqrt{\beta})^2$. Therefore, $\mathbb{E}||J|| \xrightarrow{\mathtt{N}\to\infty}(1+\sqrt{\beta})^2$. Therefore

$$\mathbb{E}[(\tilde{f}_l(0)')^2] \leq 12 + 3\mathbb{E}\Big\langle\sum_{k_1k_2}\frac{\beta}{\mathtt{K}}J_{k_1,k_2}z_{0k_1}z_{0k_2}z_{0l}^2\Big\rangle_{\tilde{\mathtt{H}}_l(0)} + 3\mathbb{E}\Bigg[\Big(\frac{1}{\sqrt{\mathtt{N}}}\sum_i N_iS_{il}\Big)^2\Bigg] \leq 12 + 3\times 2^4\beta\mathbb{E}||J|| + 3\mathbb{E}\Bigg[\frac{1}{\mathtt{N}}\sum_i S_{il}^2\Bigg] = O(1).$$

For bounding $\mathbb{E}[(\tilde{f}_l(1)')^2]$, we use symmetry of the indices and take the sum over l and divide by $\mathtt{K}$. Let A be an $\mathtt{N}\times\mathtt{N}$ matrix with entries given by $A_{ij} = \frac{1}{\mathtt{K}}\sum_l S_{il}S_{jl}$. For this case, from [33] and [34], we get $\mathbb{E}||A|| \xrightarrow{\mathtt{N}\to\infty}(1+\frac{1}{\sqrt{\beta}})^2$. Therefore

$$\mathbb{E}[(\tilde{f}_l'(1))^2] \leq 12 + 3\mathbb{E}\Big\langle\frac{1}{\mathtt{K}}\sum_l\sum_{k_1,k_2} J_{lk_1}J_{lk_2}z_{0k_1}z_{0k_2}z_{0l}^2\Big\rangle_{\tilde{\mathtt{H}}_l(1)} + 3\mathbb{E}\Big\langle\frac{1}{\mathtt{K}}\frac{1}{\mathtt{N}}\sum_{i_1,i_2}N_{i_1}N_{i_2}\sum_l S_{i_1l}S_{i_2l}\Big\rangle_{\tilde{\mathtt{H}}_l(1)} \leq 12 + 6\times 2^4\mathbb{E}||J||^2 + 3\mathbb{E}\Bigg[||A||\frac{1}{\mathtt{N}}\sum_i (N_i)^2\Bigg] = 12 + 96\mathbb{E}||J||^2 + 3\mathbb{E}||A|| = O(1).$$ □

APPENDIX II
ESTIMATES (62) AND (63)

Let $z_k^{(\alpha)} = x_k^0 - x_k^{(\alpha)}$ and $\bar{z}^{(\alpha)}$ denote the vector $(z_1^{(\alpha)}, \ldots, z_{\mathtt{K}}^{(\alpha)})$. Let us split the contribution from $T_1 - T_2$ in to $T_{11}+T_{12}$ corresponding to the two terms appearing in (61). For $T_{11}(i,k)$, we get

$$T_{11}(i,k)=\frac{1}{2\sqrt{t}}\mathbb{E}_{R_{ik}}\left[(R_{ik}^2-1)\int_0^{R_{ik}}\mathbb{E}\left[\frac{\partial^2 g_{ik}(u)}{\partial u^2}\right]du\right] \tag{79}$$

where $g_{ik}(u)$ denotes the function in (57) with r_{ik} replaced by u. The expectation $\mathbb{E}_{R_{ik}}[\cdot]$ denotes expectation with respect to R_{ik}. Let $\langle\cdot\rangle_{t,i,k}$ denote the Gibbs measure with $r_{ik}=u$. Let $\bar{v}_i^k(t)$ denote the vector $\bar{v}_i(t)$ with r_{ik} replaced by u. We now show that the term inside the integral decays with $\mathtt{N}$. We, therefore, start by computing the derivatives. The first derivative is given by

$$\frac{\partial g_{ik}(u)}{\partial u}=\frac{\sqrt{t}}{\sigma^4\mathtt{KN}}\left\langle\sigma^2 z_k^2-(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i(t)\cdot\bar{z})^2 z_k^2 + \prod_{a=1,2}(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i(t)\cdot\bar{z}^{(a)})(z_k^{(a)})\right\rangle_{t,i,k}. \tag{80}$$

Differentiating once more, we get

$$\begin{aligned}\frac{\partial^2 g_{ik}(u)}{\partial u^2}=&\frac{t}{\sigma^6\mathtt{KN}^{\frac{3}{2}}}\\ &\times\Big\langle -3\sigma^2(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i^k(t)\cdot\bar{z})z_k^3\\ &+3\sigma^2(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i^k(t)\cdot\bar{z}^{(2)})(z_k^{(1)})^2 z_k^{(2)}\\ &+(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i^k(t)\cdot\bar{z})^3 z_k^3\\ &-3\left(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i^k(t)\cdot\bar{z}^{(1)}\right)^2\\ &\times\left(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i^k(t)\cdot\bar{z}^{(2)}\right)(z_k^{(1)})^2 z_k^{(2)}\\ &+2\Pi_{a=1,2,3}\left(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i^k(t)\cdot\bar{z}^{(a)}\right)\\ &\times\left(z_k^{(a)}\right)\Big\rangle_{t,i,k}.\end{aligned} \tag{81}$$

The Hamiltonians corresponding to $\langle\cdot\rangle_t$ and $\langle\cdot\rangle_{t,i,k}$ are given by

$$\begin{aligned}\mathtt{H}(\bar{z})&=-\frac{1}{2\sigma^2}\|\bar{n}+\mathtt{N}^{-\frac{1}{2}}\mathbf{v}(t)\bar{z}\|^2\\ \mathtt{H}_{i,k}(\bar{z})&=-\frac{1}{2\sigma^2}\|\bar{n}+\mathtt{N}^{-\frac{1}{2}}\mathbf{v}_{i,k}(t)\bar{z}\|^2\end{aligned}$$

where $\mathbf{v}_{i,k}(t)$ differs from $\mathbf{v}(t)$ only in the (i,k)th entry with u replacing r_{ik}. Expanding $\mathtt{H}_{i,k}$

$$\begin{aligned}\mathtt{H}_{i,k}(\bar{z})=\frac{1}{2\sigma^2}\Bigg[&-\sum_{j\neq i}\left(n_j+\mathtt{N}^{-\frac{1}{2}}\bar{v}_j\cdot\bar{z}\right)^2\\ &-\left(n_i+\mathtt{N}^{-\frac{1}{2}}\sum_{l\neq k}v_{il}z_l+\mathtt{N}^{-\frac{1}{2}}\sqrt{1-t}s_{ik}z_k\right)^2\\ &-\frac{u^2tz_k^2}{\mathtt{N}}+\frac{u\sqrt{t}z_k}{\sqrt{\mathtt{N}}}\\ &\times\left(n_i+\mathtt{N}^{-\frac{1}{2}}\sum_{l\neq k}v_{il}z_l+\mathtt{N}^{-\frac{1}{2}}\sqrt{1-t}s_{ik}z_k\right)\Bigg].\end{aligned}$$

Let the sum of the first two terms be denoted as $H'_{ik}(\bar{z})$ and the terms involving u be $H''_{ik}(\bar{z})$. Consider the set G defined as

$$G=\left\{(\bar{n},\mathbf{r},\mathbf{s}):\forall i\ \frac{1}{\sqrt{\mathtt{N}}}|n_i|+\frac{1}{\mathtt{N}}\sum_k 2(|r_{ik}|+|s_{ik}|)\le C\right\}.$$

For sufficiently large C, we have $P(G^c)=O(e^{-\alpha\mathtt{N}})$ for some constant $\alpha>0$. If $(\bar{n},\mathbf{s},\mathbf{r})\in G$, then for all $\bar{z}\in\{0,2\}^{\mathtt{K}}$

$$|H''_{i,k}(\bar{z})|\le\frac{4|u|^2}{\mathtt{N}}+2|u|C\equiv C'(u). \tag{82}$$

Therefore, for the first term in (81)

$$\begin{aligned}&\left|\mathbb{E}\left\langle(N_i+\mathtt{N}^{-\frac{1}{2}}\bar{V}_i^k\cdot\bar{z})\right\rangle_{t,i,k}\right|\\ &\le\mathbb{E}\left[\frac{\sum_{\bar{z}}e^{H'_{ik}(\bar{z})}e^{\frac{C'(u)}{2\sigma^2}}|N_i+\mathtt{N}^{-\frac{1}{2}}\bar{V}_i^k\cdot\bar{z}-u\sqrt{t}\mathtt{N}^{-\frac{1}{2}}z_k|}{\sum_{\bar{z}}e^{H'_{ik}(\bar{z})}e^{-\frac{C'(u)}{2\sigma^2}}}\mathbb{1}_G\right]\\ &\quad+O\left(\frac{|u|}{\sqrt{\mathtt{N}}}\right)+\mathbb{E}\left\langle|N_i+\mathtt{N}^{-\frac{1}{2}}\bar{V}_i\cdot\bar{z}|\mathbb{1}_{G^c}\right\rangle_{t,i,k}.\end{aligned}$$

The expectation over G^c can be bounded as $O(e^{-\alpha\mathtt{N}})(1+O(|u|))$. Therefore, the last two terms contribute $O\left(\frac{|u|}{\sqrt{\mathtt{N}}}\right)+O(e^{-\alpha\mathtt{N}})$. For the first term after we have removed the terms with u dependence, the Hamiltonian H'_{ik} satisfies Nishimori symmetry. Therefore, we get the first term to be equal to

$$\begin{aligned}&\mathbb{E}\int\frac{e^{2\frac{C'(u)}{2\sigma^2}}}{2^{\mathtt{K}}}\sum_{\bar{z}}e^{H'_{ik}(\bar{z})}|n_i+\mathtt{N}^{-\frac{1}{2}}\bar{V}_i^k\cdot\bar{z}-u\sqrt{t}\mathtt{N}^{-\frac{1}{2}}z_k|\,d\bar{n}\\ &\overset{(a)}{=}\mathbb{E}\int e^{2\frac{C'(u)}{2\sigma^2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{m_i^2}{2}}|m_i|\,dm_i\\ &=\sqrt{\frac{\sigma^2}{2\pi}}e^{\frac{C'(u)}{\sigma^2}}.\end{aligned}$$

The step (a) follows by doing the change of variables $m_i=n_i+\mathtt{N}^{-\frac{1}{2}}\bar{V}_i^k\cdot\bar{z}-u\sqrt{t}\mathtt{N}^{-\frac{1}{2}}z_k$. Using similar arguments, we can show that

$$\begin{aligned}&\mathbb{E}\left[\frac{\partial^2 g_{ik}(u)}{\partial u^2}\right]\\ &\le\frac{1}{\mathtt{N}^{\frac{5}{2}}}\left(O(1)e^{\frac{3C'(u)}{\sigma^2}}+O\left(\frac{|u|^3}{\sqrt{\mathtt{N}}}\right)+O(e^{-\alpha\mathtt{N}})\right).\end{aligned} \tag{83}$$

The exponent 3 is due the occurrence of three replicas in the (81) and the factor $\frac{1}{\mathtt{N}^{\frac{5}{2}}}$ is due to the factor $\frac{1}{\mathtt{K}\mathtt{N}^{\frac{3}{2}}}$ in (81). Therefore

$$\begin{aligned}
&\mathbb{E}_{R_{ik}}\left[(R_{ik}^2-1)\int_0^{R_{ik}}\mathbb{E}\left[\frac{\partial^2 g_{ik}(u)}{\partial u^2}\right]du\right]\\
&\leq \mathbb{E}_{R_{ik}}\left[R_{ik}^2\int_0^{R_{ik}}\frac{1}{\mathtt{N}^{\frac{5}{2}}}\right.\\
&\qquad\left.\times\left(O(1)e^{3\frac{C'(u)}{\sigma^2}}+O\left(\mathtt{N}^{-\frac{1}{2}}|u|^3\right)+O\left(e^{-\alpha\mathtt{N}}\right)\right)du\right]\\
&\leq O\left(\mathtt{N}^{-\frac{5}{2}}\right) \qquad (84)
\end{aligned}$$

where we have assumed that R_{ik} belongs to Class A (Section II-B). Now summing this over all i,k, we get

$$|T_{11}|\leq O\left(\mathtt{N}^{-\frac{1}{2}}\right). \qquad (85)$$

Now consider the term T_{12}. For this, we have to evaluate $\frac{\partial^3 g_{ik}(u)}{\partial u^3}$. After some manipulations, the derivative can be expressed as

$$\begin{aligned}
&\frac{t^{\frac{3}{2}}}{\sigma^8\mathtt{K}\mathtt{N}^2}\Big\langle -3\sigma^4 z_k^4+6\sigma^2(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i(t)\cdot\bar{z})^2 z_k^4\\
&\quad-12\sigma^2\Pi_{a=1,2}(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i(t)\cdot\bar{z}^{(a)})(z_k^{(1)})^3 z_k^{(2)}\\
&\quad+3\sigma^4(z_k^{(1)})^2(z_k^{(2)})^2-(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i(t)\cdot\bar{z})^4 z_k^4\\
&\quad-6\sigma^2(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i(t)\cdot\bar{z}^{(2)})^2(z_k^{(1)})^2(z_k^{(2)})^2\\
&\quad+9\sigma^2\Pi_{a=2,3}(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i(t)\cdot\bar{z}^{(a)})(z_k^{(1)})^2 z_k^{(2)} z_k^{(3)}\\
&\quad+4(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i(t)\cdot\bar{z}^{(1)})^3(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i(t)\cdot\bar{z}^{(2)})\\
&\quad\times(z_k^{(1)})^3 z_k^{(2)}+3\Pi_{a=1,2}(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i(t)\cdot\bar{z}^{(a)})^2\\
&\quad\times(z_k^{(1)})^2(z_k^{(2)})^2-12(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i(t)\cdot\bar{z}^{(1)})^2\\
&\quad\times(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i(t)\cdot\bar{z}^{(2)})(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i(t)\cdot\bar{z}^{(3)})\\
&\quad\times(z_k^{(1)})^2 z_k^{(2)} z_k^{(3)}\\
&\quad+6\Pi_{a=1,2,3,4}(n_i+\mathtt{N}^{-\frac{1}{2}}\bar{v}_i(t)\cdot\bar{z}^{(a)})(z_k^{(a)})\Big\rangle_{t,i,k}.
\end{aligned}$$

We can prove along similar lines that $|T_{12}|\leq O(\mathtt{N}^{-1})$.

Appendix III
Proofs of Nishimori Identities

The proofs are based on a standard method used frequently in the Gauge theory of spin glasses [23, ch. 4].

Proof of Lemma 1: The proof is similar to the proof of Lemma 2 and hence we only give a brief sketch here. We write explicitly the expression for $\mathbb{P}_{m_1}^t(x)$ and perform the gauge transformation $x_k\to x_k^0 x_k$, $S_{ik}\to x_k^0 S_{ik}$ where $\bar{x}^0$ is an arbitrary binary sequence. Since $\mathbb{P}_{m_1}^t(x)$ does not depend on $\bar{x}^0$, we sum over all such $2^{\mathtt{K}}$ sequences and obtain a lengthy expression. Exactly the same procedure is applied to $\mathbb{P}_{q_{12}}^t(x)$ and one gets another lengthy expression. Then, one can recognize that these two expressions are the same. □

Proof of Lemma 2:

Proof of (41): We will prove it for $t=1$ and for general t it is similar. Let the transmitted sequence be the all-one sequence, and the received vector be $\bar{y}=\frac{1}{\sqrt{\mathtt{N}}}\mathbf{s}\bar{1}+\sigma\bar{n}$. The proof follows by using gauge transformation. Let $\bar{u}$ denote the $\mathtt{K}$ dimensional vector $(u,\ldots,u)$. Then

$$\begin{aligned}
&\mathbb{E}[\langle\|\bar{\mathcal{Z}}\|^2\rangle_{1,u}]\\
&=\mathbb{E}_{\mathbf{S}}\left[\int\frac{e^{-\frac{\|\bar{h}-\bar{u}\|^2}{2u}}e^{-\frac{1}{2\sigma^2}\|\bar{y}-\mathtt{N}^{-\frac{1}{2}}\mathbf{S}\|^2}}{(2\pi u)^{\frac{\mathtt{K}}{2}}(2\pi\sigma^2)^{\frac{\mathtt{N}}{2}}}\langle\|\bar{\mathcal{Z}}\|^2\rangle_{1,u}d\bar{y}\,d\bar{h}\right]\\
&=\mathbb{E}_{\mathbf{S}}\Bigg[\int\frac{e^{-\frac{\|\bar{h}\|^2}{2u}+\bar{h}\cdot\bar{1}-\frac{\mathtt{K}u}{2}}e^{-\frac{1}{2\sigma^2}\|\bar{y}-\mathtt{N}^{-\frac{1}{2}}\mathbf{S}\|^2}}{(2\pi u)^{\frac{\mathtt{K}}{2}}(2\pi\sigma^2)^{\frac{\mathtt{N}}{2}}}\\
&\qquad\times\frac{\sum_{\bar{x}}e^{-\frac{1}{2\sigma^2}\|\bar{y}-\mathtt{N}^{-\frac{1}{2}}\mathbf{S}\bar{x}\|^2+\bar{h}\cdot\bar{x}}\|\bar{\mathcal{Z}}\|^2}{\sum_{\bar{x}}e^{-\frac{1}{2\sigma^2}\|\bar{y}-\mathtt{N}^{-\frac{1}{2}}\mathbf{S}\bar{x}\|^2+\bar{h}\cdot\bar{x}}}d\bar{y}\,d\bar{h}\Bigg]\\
&\overset{(a)}{=}\frac{1}{2^{\mathtt{K}}}\mathbb{E}_{\mathbf{S}}\Bigg[\int\frac{e^{-\frac{\|\bar{h}\|^2}{2u}-\frac{\mathtt{K}u}{2}}}{(2\pi u)^{\frac{\mathtt{K}}{2}}(2\pi\sigma^2)^{\frac{\mathtt{N}}{2}}}\sum_{\bar{x}^0}e^{-\frac{1}{2\sigma^2}\|\bar{y}-\mathtt{N}^{-\frac{1}{2}}\mathbf{S}\bar{x}^0\|^2+\bar{h}\cdot\bar{x}^0}\\
&\qquad\times\frac{\sum_{\bar{x}}e^{-\frac{1}{2\sigma^2}\|\bar{y}-\mathtt{N}^{-\frac{1}{2}}\mathbf{S}\bar{x}\|^2+\bar{h}\cdot\bar{x}}\|\bar{\mathcal{Z}}\|^2}{\sum_{\bar{x}}e^{-\frac{1}{2\sigma^2}\|\bar{y}-\mathtt{N}^{-\frac{1}{2}}\mathbf{S}\bar{x}\|^2+\bar{h}\cdot\bar{x}}}d\bar{y}\,d\bar{h}\Bigg]\\
&=\mathtt{N}.
\end{aligned}$$

The equality (a) is obtained by performing the gauge transformation $x_k\to x_k x_k^0$, $S_{ik}\to S_{ik}x_k^0$ and $h_k\to h_k x_k^0$ and summing over all the $2^{\mathtt{K}}$ possibilities of $\bar{x}^0$. Now canceling the summation over $\bar{x}^0$ with the denominator and then integrating, we get it to be equal to $\mathtt{N}$.

Proof of (42): The proof is complete if we show $\mathbb{E}[\langle(\bar{N}\cdot\bar{\mathcal{Z}}^{(2)})(\bar{x}^{(1)}\cdot\bar{z}^{(2)})\rangle_{t,u}]=0$. We will prove this for $t=1$ and it is similar for other t

$$\begin{aligned}
&\mathbb{E}[\langle(\bar{N}\cdot\bar{\mathcal{Z}}^{(2)})(\bar{x}^{(1)}\cdot\bar{z}^{(2)})\rangle_{t,u}]\\
&=\sum_{i,k}\mathbb{E}\Bigg[\Bigg\langle\left(Y_i-\mathtt{N}^{-\frac{1}{2}}\sum_l S_{il}\right)\left(Y_i-\mathtt{N}^{-\frac{1}{2}}\sum_l S_{il}x_l^{(2)}\right)\\
&\qquad\times\left(x_k^{(1)}-x_k^{(1)}x_k^{(2)}\right)\Bigg\rangle_{t,u}\Bigg].
\end{aligned}$$

Now performing the gauge transformation $x_k^{(1)}\to x_k^{(1)}x_k^0$, $x_k^{(2)}\to x_k^{(2)}x_k^0$, $S_{ik}\to S_{ik}x_k^0$ and $h_k\to h_k x_k^0$, we get

$$\begin{aligned}
&\sum_{i,k}\mathbb{E}\Bigg[\Bigg\langle\left(Y_i-\mathtt{N}^{-\frac{1}{2}}\sum_l S_{il}x_l^0\right)\left(Y_i-\mathtt{N}^{-\frac{1}{2}}\sum_l S_{il}x_l^{(2)}\right)\\
&\qquad\times\left(x_k^{(1)}x_k^0-x_k^{(1)}x_k^{(2)}\right)\Bigg\rangle_{t,u}\Bigg].
\end{aligned}$$

This quantity can be shown to be equal to 0 by noticing that the $\bar{x}^0$ and $\bar{x}^{(2)}$ play symmetric roles.

APPENDIX IV
PROOF OF INEQUALITY (43)

For a given configuration of $\bar{z}$, $\frac{1}{\sqrt{\mathtt{N}}}\sum_l S_{il}z_l \equiv T_i$ is a Gaussian random variable with mean 0 and variance smaller than 4. Thus, for $N_i \sim \mathcal{N}(0,1)$ and independent of T_i

$$\mathbb{E}\left[e^{\frac{N_i T_i}{\alpha}}\right] = \mathbb{E}\left[e^{-\frac{N_i T_i}{\alpha}}\right] \le \sqrt{\frac{\alpha^2}{\alpha^2-4}}.$$

If $\alpha > 2$, we have both the expectations to be less than some constant $C > 1$. Therefore, for any $\bar{z}$

$$\mathbb{E}\left[e^{\frac{1}{\alpha}\frac{1}{\sqrt{\mathtt{N}}}\sum_{i,l}N_i S_{il} z_l}\right] = \mathbb{E}\left[e^{-\frac{1}{\alpha}\frac{1}{\sqrt{\mathtt{N}}}\sum_{i,l}N_i S_{il} z_l}\right] \le C^{\mathtt{N}}.$$

Using the Markov inequality, we get

$$\mathbb{P}\left(\left|\frac{1}{\alpha}\sum_i N_i \frac{1}{\sqrt{\mathtt{N}}}\sum_k S_{ik}z_k\right| > A\mathtt{N}\right) \le 2C^{\mathtt{N}}e^{-A\mathtt{N}}. \quad (86)$$

Let G denote the set

$$G = \left\{(\bar{n},\mathbf{s}) : \forall \bar{z} \in \{0,2\}^{\mathtt{K}}, \left|\frac{1}{\mathtt{N}^{3/2}}\sum_{i,k} n_i s_{ik} z_k\right| \le \alpha A\right\}.$$

Using the union bound over $\bar{z}$ for the bound (86), we get

$$\mathbb{P}(G^c) \le 22^{\mathtt{K}} C^{\mathtt{N}} e^{-A\mathtt{N}}.$$

For A large enough there exists a constant $\gamma > 0$ such that $\mathbb{P}(G^c) \le 2^{-\gamma \mathtt{N}}$. Splitting the expectation into two parts corresponding to G and G^c and using Cauchy–Schwartz inequality, we get

$$\begin{aligned}
&\mathbb{E}\left\langle\left(\frac{1}{\mathtt{N}^{3/2}}\sum_{i,k}N_i S_{ik}z_k\right)^2\right\rangle_t \\
&\le \alpha^2 A^2 + \sqrt{\mathbb{P}(G^c)}\left(\mathbb{E}\left\langle\left(\frac{1}{\mathtt{N}^{3/2}}\sum_{i,l}N_i S_{ik}z_k\right)^4\right\rangle_t\right)^{1/2} \\
&\le \alpha^2 A^2 + O\left(2^{-\frac{\gamma}{2}\mathtt{N}}\right).
\end{aligned}$$

ACKNOWLEDGMENT

The authors would like to thank S. Kudekar, O. Lévêque, A. Montanari, and R. Urbanke for useful discussions.

REFERENCES

[1] S. Verdú, "Capacity region of Gaussian CDMA channels: The symbol synchronous case," in *Proc. Allerton Conf. Commun. Control Comput.*, Monticello, IL, Oct. 1986, pp. 1025–1034.

[2] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[3] A. J. Grant and P. D. Alexander, "Randomly selected spreading sequences for coded CDMA," in *Proc. 4th Int. Spread Spectrum Tech. Appl.*, Mainz, Germany, Sep. 1996, pp. 54–57.

[4] A. J. Grant and P. D. Alexander, "Random sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2832–2836, Nov. 1998.

[5] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 622–640, Mar. 1999.

[6] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 641–657, Mar. 1999.

[7] D. N. C. Tse and S. Verdú, "Optimum asymptotic multiuser efficiency of randomly spread CDMA," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2718–2722, Nov. 2000.

[8] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.

[9] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, Jun. 2005.

[10] M. Talagrand, *Spin Glasses: A Challenge for Mathematicians: Cavity and Mean Field Models*. New York: Springer-Verlag, 2003.

[11] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka's formula)," in *Proc. IEEE Inf. Theory Workshop*, Punta del Este, Uruguay, Mar. 13–17, 2006, pp. 160–164.

[12] A. Montanari, "The glassy phase of Gallager codes," *Eur. Phys. J. B*, vol. 23, pp. 121–136, 2001.

[13] Y. Kabashima and T. Hosaka, "Statistical mechanics of source coding with a fidelity criterion," *Progr. Theor. Phys.*, no. 157, pp. 197–204, 2005.

[14] K. Nakamura, Y. Kabashima, R. Morelos-Zaragoza, and D. Saad, "Statistical mechanics of broadcast channels using low-density parity-check codes," *Phys. Rev. E*, vol. 67, 2003, 036703 (1-9).

[15] F. Guerra, "Sum rules for the free energy in the mean field spin glass model," *Fields Inst. Commun.*, vol. 30, pp. 161–170, 2001.

[16] F. Guerra and F. L. Toninelli, "Quadratic replica coupling in the Sherrington-Kirkpatrick mean field spin glass model," *J. Math. Phys.*, vol. 43, pp. 3704–3716, 2002.

[17] F. Guerra and F. L. Toninelli, "The thermodynamic limit in mean field spin glass models," *Commun. Math. Phys.*, vol. 230, pp. 71–79, 2002.

[18] F. Guerra and F. L. Toninelli, "The infinite volume limit in generalized mean filed disordered models," *Markov Proc. Rel. Fields*, vol. 49, no. 2, pp. 195–207, 2003.

[19] A. Montanari, "Tight bounds for LDPC and LDGM codes under MAP decoding," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3221–3246, Sep. 2005.

[20] S. Kudekar and N. Macris, "Sharp bounds for MAP decoding of general irregular LDPC codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, WA, Sep. 2006, pp. 2259–2263.

[21] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.

[22] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, 1998.

[23] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction*. Oxford, U.K.: Oxford Science, 2001.

[24] H. Nishimori, "Comment on statistical mechanics of CDMA multiuser demodulation by T. Tanaka," *Europhys. Lett.*, vol. 57, no. 2, pp. 302–303, 2002.

[25] S. B. Korada and N. Macris, "On the concentration of the capacity for a code division multiple access system," in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, France, Jun. 24–29, 2007, pp. 2801–2805.

[26] S. B. Korada and N. Macris, "On the capacity of a code division multiple access system," in *Proc. Allerton Conf. Commun. Control Comput.*, Monticello, IL, Sep. 26–28, 2007, pp. 959–966.

[27] L. A. Pastur and M. Shcherbina, "Absence of self-averaging of the order parameter in the Sherrington-Kirkpatrick model," *J. Stat. Phys.*, vol. 62, no. 1-2, pp. 1–19, Jan. 1991.

[28] M. Shcherbina and B. Tirozzi, "The free energy of a class of Hopfield models," *J. Stat. Phys.*, vol. 72, pp. 113–125, 1993.

[29] F. L. Toninelli, "Rigorous results for mean field spin glasses: Thermodynamic limit and sum rules for the free energy," Ph.D. dissertation, Scuola Normale Superiore, Pisa, Italy, 2002.

[30] G. Pólya and G. Szegö, *Problems and Theorems in Analysis I*. New York: Springer-Verlag, 1978.

[31] M. Talagrand, "A new look at independence," *Ann. Probab.*, vol. 24, no. 1, pp. 1–34, 1996.

[32] J. Heinonen, "Lectures on Lipschitz analysis," Dept. Math. Stat., Univ. Jyvaskyla, Jyvaskyla, Finland, 2005.

[33] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah, "On the limit of the largest eigenvalue of the large dimensional sample covariance matrix," *Probab. Theory Rel. Fields*, vol. 78, no. 4, pp. 509–521, 1988.

[34] A. Bovier and V. Gayrard, "Hopfield models as generalized random mean field models," in *Mathematical Aspects of Spin Glasses and Neural Networks*, A. Bovier and P. Picco, Eds. Boston, MA: Birkhauser, 1998, vol. 41, pp. 1–89.

Satish Babu Korada received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi, India, in 2004 and the Ph.D. degree in communications and information theory from EPFL, Lausanne, Switzerland, in 2009.

Since September 2009, he has been a Postdoctoral Researcher at Stanford University, Stanford, CA. His research interests include coding theory, communications, low-complexity algorithms, and applications of statistical mechanics.

Dr. Korada received the Best Student Paper awards at the IEEE International Symposium on Information Theory in 2008 and 2009.

Nicolas Macris received the Ph.D. degree in theoretical physics from EPFL, Lausanne, Switzerland.

He then pursued his scientific activity at the Mathematics Department, Rutgers—The State University of New Jersey, Piscataway. Then, he joined the School of Basic Sciences, EPFL, where he worked in the field of quantum statistical mechanics and mathematical aspects of the quantum Hall effect. In 2005, he joined the Communication Theories Laboratory in the School of Communication and Computer Science and is currently working at the interface between the theory of error correcting codes, statistical mechanics, and information theory.