Sharp Bounds for Optimal Decoding of Low Density Parity Check Codes

Shrinivas Kudekar and Nicolas Macris

Ecole Polytechnique Fédérale de Lausanne
School of Computer and Communication Sciences
LTHC, I&C, Station 14, CH-1015 Lausanne
shrinivas.kudekar@epfl.ch, nicolas.macris@epfl.ch

July 18, 2008

Abstract

Consider communication over a binary-input memoryless output-symmetric channel with low density parity check (LDPC) codes and maximum a posteriori (MAP) decoding. The replica method of spin glass theory allows to conjecture an analytic formula for the average input-output conditional entropy per bit in the infinite block length limit. Montanari proved a lower bound for this entropy, in the case of LDPC ensembles with convex check degree polynomial, which matches the replica formula. Here we extend this lower bound to any irregular LDPC ensemble. The new feature of our work is an analysis of the second derivative of the conditional input-output entropy with respect to noise. A close relation arises between this second derivative and correlation or mutual information of codebits. This allows us to extend the realm of the "interpolation method", in particular we show how channel symmetry allows to control the fluctuations of the "overlap parameters".

1 Introduction and Main Results

Linear codes based on sparse random graphs have emerged as a major chapter of coding theory [1]. While the belief propagation (BP) decoding algorithm and density evolution method have been explored in detail because of their low algorithmic complexity and good performance, much remains to be understood about the optimal (MAP) performance bounds of sparse graph codes. Recent theoretical progress on the binary erasure channel (BEC) has convincingly shown that BP and MAP decoding have intimate relationships (see [1] and in particular [4]), but understanding this relationship for other channels is still a largely open problem. In fact, the replica and/or cavity methods of statistical mechanics of dilute spin glass models allow to conjecture an analytic formula for $H_n(\underline{X}|\underline{Y})$, the entropy of the transmitted message $\underline{X} = (X_1, ..., X_n)$ conditional to the received message $\underline{Y} = (Y_1, ..., Y_n)$ in the large block length limit $n \to +\infty$. The replica formula expresses the conditional entropy as the solution of a variational problem whose critical points are given by the density evolution fixed point equation (see [2], [3]). If one is to solve the fixed point equation iteratively, the choice of initial conditions is not necessarily the one given by channel outputs (as in standard density evolution) but the one which yields the maximum conditional entropy. Note that a byproduct of the replica formula is the determination of the maximum a posteriori (MAP) noise threshold, above which reliable communication is not possible whatever the decoding algorithm.

The proof of the replica formulas is, in general, an open problem[1]. In the context of communication they have been proven for a class of low density parity check codes (LDPC) codes on the BEC [11], [12] (see also [13] for recent work going beyond the BEC) and for low density generator codes (LDGM) on a class of channels [14].

A promising approach towards a general proof of the replica formulas seems to be the use of the so-called interpolation method first developped in the context of the SK model [15], [16], [17]. Consider an LDPC(n, Λ, P) ensemble where $\Lambda(x) = \sum_d \Lambda_d x^d$, $P(x) = \sum_k P_k x^k$ are the variable and check degree distributions from the node perspective. We will always assume that the maximal degrees are finite. Montanari [7] (see also the related

[1] In a few spin glass models the replica formulas have been fully demonstrated. Remarkably Talagrand [5] has proven the Parisi formula with full symmetry breaking [6] for the Sherrington-Kirkpatrick (SK) model. In [10] it is shown that the replica symmetric formula holds for a complete p-spin model with gauge symmetry.

work of Franz-Leone [8] and Talagrand- Pachenko [9]) has developped the interpolation method for such a system and has derived a lower bound for the conditional entropy for ensembles with any polynomial $\Lambda(x)$ but $P(x)$ restricted to be convex for $-e \le x \le e$ (in particular if the check degree is constant this means it has to be even). An important fact is that these lower bounds match the replica solution, and are thus believed to be tight. Since Fano's inequality tells us that the block error probability for a code having length n and rate r is lower bounded by $\frac{1}{rn}H_n(\underline{X}|\underline{Y})$, an immediate application of the lower bound is the numerical computation of a rigorous upper bound on the MAP threshold.

In the present paper we drop the convexity requirement for $P(x)$ in the cases of the BEC, BIAWGNC with any noise level an in the case of general binary memoryless (BMS) channels in a high noise regime. In other words we prove the lower bound for any standard regular (so odd degrees are allowed) or irregular code ensemble.

Besides the main result itself, we introduce a new tool in the form of a relationship between the second derivative of the conditional entropy with respect to the noise and correlations functions of codebits. These correlation functions are shown to be intimately related to the mutual information between two codebits. The formulas are somewhat similar to those for GEXIT functions [1] which relate the first derivative of conditional entropy to soft bit estimates. By combining these relations with the interpolation method we are able to control the fluctuations of the so-called overlap parameters. This part of our analysis is crucial for proving the general lower bound on the conditional entropy and relies heavily on channel symmetry.

A preliminary summary of the present work has appeared in [20].

1.1 Variational bound on the conditional entropy

Let $p_{Y|X}(y|x)$ be the transition probability of a BMS(ϵ) channel where ϵ is the noise parameter (understood to vary in the appropriate range). We will work in terms of both the likelihood

$$l = \ln\left[\frac{p_{Y|X}(y|0)}{p_{Y|X}(y|1)}\right]$$

and difference

$$t = p_{Y|X}(y|0) - p_{Y|X}(y|1) = \tanh\frac{l}{2}$$

variables. It will be convenient to use the notation $c_L(l)$ and $c_D(t)$ for the distributions of l and t, assuming that the all zero codeword is transmitted (that is to say that $c_L(l)dl = c_D(t)dt = p_{Y|X}(y|0)dy$).

Let V be some random variable with an arbitrary density $d_V(v)$ satisfying the symmetry condition $d_V(v) = e^v d_V(-v)$. Also let

$$U = \tanh^{-1}\left[\prod_{i=1}^{k-1} \tanh V_i\right] \tag{1}$$

where V_i are i.i.d copies of V and k is the (random) degree of a check node. We denote by U_c, $c = 1, ..., d$ i.i.d copies of U where d is the (random) degree of variable nodes. Notice that in the belief propagation (BP) decoding algorithm U appears as the check to variable node message and V appears as the variable to check node message. Define the functional[2] (we view it as a functional of the probability distribution d_V)

$$\begin{aligned}
h_{RS}[d_V; \Lambda, P] =& \mathbb{E}_{l,d,U_c}\left[\ln\left(e^{\frac{l}{2}}\prod_{c=1}^{d}(1+\tanh U_c) + e^{-\frac{l}{2}}\prod_{c=1}^{d}(1-\tanh U_c)\right)\right] \\
&+ \frac{\Lambda'(1)}{P'(1)}\mathbb{E}_{k,V_i}\left[\ln(1+\prod_{i=1}^{k}\tanh V_i)\right] \\
&- \Lambda'(1)\mathbb{E}_{V,U}\left[\ln(1+\tanh V \tanh U)\right] - \frac{\Lambda'(1)}{P'(1)}\ln 2
\end{aligned}$$

Our main result is about the conditional entropy per bit, averaged over the code ensemble $\mathcal{C} = \mathrm{LDPC}(n, \Lambda, P)$.

$$\mathbb{E}_{\mathcal{C}}[h_n] = \frac{1}{n}\mathbb{E}_{\mathcal{C}}[H_n(\underline{X}|\underline{Y})]$$

Definition H. *We define the parameters (p an integer)*

$$m_0^{(2p)} = \mathbb{E}[t^{2p}], \qquad m_1^{(2p)} = \frac{d}{d\epsilon}\mathbb{E}[t^{2p}], \qquad m_2^{(2p)} = \frac{d^2}{d\epsilon^2}\mathbb{E}[t^{2p}] \tag{2}$$

[2]The subscript RS stands for "replica symmetric" because this functional has been obtained from the replica symmetric ansatz for an appropriate spin glass, see for example [3], [2]

and say that a BMS(ϵ) *channel is in the high noise regime if the following series expansions*

$$\sum_p (p+1) m_0^{(2p)} \qquad \sum_p \Big(\frac{5}{2}\Big)^{2p} |m_1^{(2p)}| \qquad \sum_p \frac{|m_2^{(2p)}|}{2p(2p-1)} \tag{3}$$

are convergent and if

$$(\sqrt{2}-1)\Big(\frac{5}{2}\Big)^2 |m_1^{(2)}| > \sum_{p\geq 2} \Big(\frac{5}{2}\Big)^{2p} |m_1^{(2p)}|$$

For example the BSC(ϵ) certainly satisfies H if the crossover noise parameter is close enough to $\frac{1}{2}$, because $\mathbb{E}[t^{2p}] = (1-2\epsilon)^{2p}$. More generaly *any channel with bounded likehood variables satisfies H for a regime of sufficiently high noise*. For channels with unbounded likehoods the condition will be satisfied if $c_L(l)$ has sufficiently good decay properties. But note that the BEC(ϵ) which has mass at $l = +\infty$ does not satisfy this condition since $\mathbb{E}[t^{2p}] = 1 - \epsilon$. However as we will see *for the BEC(ϵ) and the* BIAWGNC(ϵ) *we do not need condition H*. For these two channels our analysis can be made fully non-perturbative, and holds for all noise levels.

Theorem 1 (Variational Bound). *Assume communication using a standard irregular $\mathcal{C}$ = LDPC(n, Λ, P) code ensemble, through a BEC(ϵ) or BIAWGNC(ϵ) with any noise level or a BMS(ϵ) channel satisfying H. For all ϵ in the above ranges we have,*

$$\liminf_{n\to+\infty} \mathbb{E}_{\mathcal{C}}[h_n] \geq \sup_{d_V} h_{RS}[d_V; \Lambda, P]$$

Let us note that this theorem already appears in [18] for the special case of the BIAWGNC for a Poissonnian $\Lambda(x)$. We stress again that a formal calculation using the replica method yields

$$\lim_{n\to+\infty} \mathbb{E}_{\mathcal{C}}[h_n] = \sup_{d_V} h_{RS}[d_V; \Lambda, P]$$

For this reason it is strongly suspected that the converse inequality holds as well, but so far no progress has been made except in a limited number of situations alluded to before.

1.2 Derivatives of the conditional entropy

Our proof of the variational bound uses integral formulas for the first and second derivatives of $\mathbb{E}_{\mathcal{C}}[h_n]$ with respect to the noise parameter. The ensemble formulas follow from slightly more general ones that are valid for any fixed linear code. To give the formulation for a fixed linear code it is convenient to introduce a noise vector $\underline{\epsilon} = (\epsilon_1, ..., \epsilon_n)$ and a BMS($\underline{\epsilon}$) channel with noise level ϵ_i when bit x_i is sent. When all noise levels are set to the same value ϵ the channel is denoted BMS(ϵ). The distributions of the likelihood l_i or difference domain t_i representations of the channel outputs now depend on ϵ_i. In order to keep the notation simpler we do not explicitely indicate the ϵ_i dependence and still denote them as $c_L(l_i)$ and $c_D(t_i)$ respectively.

We introduce the soft MAP estimates of bit X_i

$$L_i = \ln\left[\frac{p_{X_i|\underline{Y}}(0|\underline{y})}{p_{X_i|\underline{Y}}(1|\underline{y})}\right], \qquad T_i = p_{X_i|\underline{Y}}(0|\underline{y}) - p_{X_i|\underline{Y}}(1|\underline{y}) = \tanh\frac{L_i}{2}$$

and the soft estimate for the modulo 2 sum $X_i \oplus X_j$,

$$L_{ij} = \ln\left[\frac{p_{X_i\oplus X_j|\underline{Y}}(0|\underline{y})}{p_{X_i\oplus X_j|\underline{Y}}(1|\underline{y})}\right], \qquad T_{ij} = p_{X_i\oplus X_j|\underline{Y}}(0|\underline{y}) - p_{X_i\oplus X_j|\underline{Y}}(1|\underline{y}) = \tanh\frac{L_{ij}}{2}$$

In the sequel the notations $\underline{v}^{\sim i}$ (resp. $\underline{v}^{\sim ij}$) means that component v_i (resp. v_i and v_j) are omitted from the vector $\underline{v}$. The following is known [1] but we state it for completeness. A derivation in the spirit of the present paper can also be found in [19].

Proposition 1 (GEXIT formula). *For any* BMS($\underline{\epsilon}$) *channel and any fixed linear code we have*

$$\frac{\partial}{\partial\epsilon_i} H_n(\underline{X} \mid \underline{Y}) = \int_{-1}^{+1} dt_i \frac{\partial c_D(t_i)}{\partial\epsilon_i} g_1(t_i)$$

where

$$g_1(t_i) = -\mathbb{E}_{\underline{t}^{\sim i}}\left[\ln\left(\frac{1 - t_i T_i}{1 - t_i}\right)\right]$$

This formula will be used for an ensemble that is symmetric under permutation of bits and a BMS(ϵ) channel. Using

$$\frac{d}{d\epsilon} H_n(\underline{X} \mid \underline{Y}) = \sum_{i=1}^{n} \frac{\partial}{\partial\epsilon_i} H_n(\underline{X} \mid \underline{Y})\bigg|_{\epsilon_i=\epsilon}$$

and averaging over the code ensemble $\mathcal{C}$ we get for the average entropy per bit,

$$\frac{d}{d\epsilon}\mathbb{E}_{\mathcal{C}}[h_n] = \int_{-1}^{+1} dt_1 \frac{\partial c_D(t_1)}{\partial \epsilon}\mathbb{E}_{\mathcal{C}}[g_1(t_1)]$$

There are two channels where these general formulas take a simpler form. For the BEC[3]

$$\frac{\partial}{\partial \ln \epsilon_i} H_n(\underline{X} \mid \underline{Y}) = \ln 2(1 - \mathbb{E}_{\underline{t}}[T_i]) \tag{4}$$

and

$$\frac{d}{d \ln \epsilon}\mathbb{E}_{\mathcal{C}}[h_n] = \ln 2(1 - \mathbb{E}_{\mathcal{C},\underline{t}}[T_1]) \tag{5}$$

Similarly on the BIAWGNC,

$$\frac{\partial}{\partial \epsilon_i^{-2}} H_n(\underline{X} \mid \underline{Y}) = -\frac{1}{2}(1 - \mathbb{E}_{\mathcal{C},\underline{l}}[T_i]) \tag{6}$$

and

$$\frac{d}{d\epsilon^{-2}}\mathbb{E}_{\mathcal{C}}[h_n] = -\frac{1}{2}(1 - \mathbb{E}_{\mathcal{C},\underline{t}}[T_1]) \tag{7}$$

We will prove

Proposition 2 (Correlation formula). *For any* BMS($\underline{\epsilon}$) *channel and any fixed linear code we have*

$$\begin{aligned}\frac{\partial^2}{\partial \epsilon_i \partial \epsilon_j} H_n(\underline{X} \mid \underline{Y}) =& \delta_{ij} \int_{-1}^{+1} dt_i \frac{\partial^2 c_D(t_1)}{\partial \epsilon_i^2} g_1(t_i) \\ &+ (1-\delta_{ij}) \int_{-1}^{+1}\int_{-1}^{+1} dt_i dt_j \frac{\partial c_D(t_i)}{\partial \epsilon_i}\frac{\partial c_D(t_j)}{\partial \epsilon_j} g_2(t_i,t_j)\end{aligned}$$

with

$$g_2(t_i,t_j) = \mathbb{E}_{\underline{t}^{\sim ij}}\left[\ln\left(\frac{1 - t_iT_i - t_jT_j + t_it_jT_{ij}}{1 - t_iT_i - t_jT_j + t_it_jT_iT_j}\right)\right]$$

[3] In this case the ratio in the logarithm may take the ambiguous value $\frac{0}{0}$ but the formula is to be interpreted as (4). We will see in section 2 that in terms of extrinsic soft bit estimates there is an analogous expresion that is unambiguous.

Again, for the case of interest later on, we have a BMS(ϵ) channel and a linear code ensemble that is symmetric under permutations of bits, thus

$$\frac{d^2}{d\epsilon^2}\mathbb{E}_{\mathcal{C}}[h_n] = \int_{-1}^{+1} dt_1 \frac{\partial^2 c_D(t_1)}{\partial\epsilon^2}\mathbb{E}_{\mathcal{C}}[g_1(t_1)] \\ + \sum_{i\neq 1}\int_{-1}^{+1}\int_{-1}^{+1} dt_1 dt_i \frac{\partial c_D(t_1)}{\partial\epsilon}\frac{\partial c_D(t_i)}{\partial\epsilon}\mathbb{E}_{\mathcal{C}}[g_2(t_1,t_i)] \tag{8}$$

For the BEC[4] these formulas simplify

$$\frac{\partial^2}{\partial\ln\epsilon_i\partial\ln\epsilon_j}H_n(\underline{X}\mid\underline{Y}) = (1-\delta_{ij})\ln 2\mathbb{E}_{\underline{t}}\big[T_{ij}-T_iT_j\big]$$

and

$$\frac{d^2}{(d\ln\epsilon)^2}\mathbb{E}_{\mathcal{C}}[h_n] = \ln 2\sum_{i\neq 1}^{n}\mathbb{E}_{\mathcal{C},\underline{t}}\big[T_{1i}-T_1T_i\big] \tag{9}$$

For the BIAWGNC

$$\frac{\partial^2}{\partial\epsilon_i^{-2}\partial\epsilon_j^{-2}}H_n(\underline{X}\mid\underline{Y}) = \frac{1}{2}\mathbb{E}_{\underline{t}}[\big(T_{ij}-T_iT_j\big)^2], \tag{10}$$

and

$$\frac{d^2}{(d\epsilon^{-2})^2}\mathbb{E}_{\mathcal{C}}[h_n] = \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C},\underline{t}}[\big(T_{1i}-T_1T_i\big)^2] \tag{11}$$

Formulas (9) and (11) involve the "correlation" $(T_{ij}-T_iT_j)$ for bits X_i and X_j. The general formula (8) can also be recast in terms of powers of such correlations by expanding the logarithm (see section 3). Loosely speaking, in the infinite block length limit $n\to+\infty$, the second derivative will be well defined only if the correlations have sufficient decay with respect to the graph distance (the minimal length among all paths joining i and j on the Tanner graph). Thus we expect good decay properties for all noise levels except at the phase transition thresholds where, in the limit $n\to+\infty$, the first derivative generally has bounded discontinuities, and thus the second derivative cannot be uniformly bounded in n.

[4]The same remark than before applies here.

1.3 Relation to mutual information

The correlation $T_{ij} - T_iT_j$ is basicaly a measure of the independence of two codebits, thus it is natural to expect that it should be related to the mutual information $I(X_i; X_j \mid \underline{Y})$. We do not pursue this issue in all details because it is not used in the rest of the paper, but wish to briefly state the main relations which follow naturaly form the previous formulas.

The BEC($\underline{\epsilon}$). Take $i \neq j$. The chain rule implies $H_n(\underline{X} \mid \underline{Y}) = H(X_iX_j \mid \underline{Y}) + H(\underline{X}^{\sim ij} \mid X_iX_j\underline{Y})$. Also $H(\underline{X}^{\sim ij} \mid X_iX_j\underline{Y}) = H(\underline{X}^{\sim ij} \mid X_iX_j\underline{Y}^{\sim ij})$. Since $H(\underline{X}^{\sim ij} \mid X_iX_j\underline{Y}^{\sim ij})$ does not depend on ϵ_i, ϵ_j we have

$$\frac{\partial^2}{\partial\epsilon_i\partial\epsilon_j}H_n(\underline{X} \mid \underline{Y}) = \frac{\partial^2}{\partial\epsilon_i\partial\epsilon_j}H(X_iX_j \mid \underline{Y})$$

The conditional entropy on the r.h.s is explicitly $\epsilon_i\epsilon_j H(X_iX_j|\underline{Y}^{\sim ij}) + \epsilon_i(1-\epsilon_j)H(X_i|X_j\underline{Y}^{\sim ij}) + (1-\epsilon_i)\epsilon_j H(X_j|X_i\underline{Y}^{\sim ij})$. In this expression the three conditional entropies are independent of the channel parameters ϵ_i and ϵ_j. Thus

$$\begin{aligned}\frac{\partial^2}{\partial\epsilon_i\partial\epsilon_j}H_n(\underline{X} \mid \underline{Y}) &= H(X_iX_j \mid \underline{Y}^{\sim ij}) - H(X_i \mid X_j\underline{Y}^{\sim ij}) - H(X_j \mid X_i\underline{Y}^{\sim ij}) \\ &= H(X_j \mid \underline{Y}^{\sim ij}) - H(X_j \mid X_i\underline{Y}^{\sim ij}) \\ &= I(X_i; X_j \mid \underline{Y}^{\sim ij}) = \frac{1}{\epsilon_i\epsilon_j}I(X_i; X_j \mid \underline{Y})\end{aligned}$$

Summarizing, we have obtained for $i \neq j$,

$$\frac{\partial^2}{\partial\ln\epsilon_i\partial\ln\epsilon_j}H_n(\underline{X} \mid \underline{Y}) = I(X_i; X_j \mid \underline{Y}) = \mathbb{E}_{\underline{t}}[T_{ij} - T_iT_j]$$

The BIAWGNC($\underline{\epsilon}$). Take $i \neq j$. We note that

$$T_{ij} = p_{X_iX_j|\underline{Y}}(00 \mid \underline{y}) + p_{X_iX_j|\underline{Y}}(11 \mid \underline{y}) - p_{X_iX_j|\underline{Y}}(01 \mid \underline{y}) - p_{X_iX_j|\underline{Y}}(10 \mid \underline{y})$$

from which it follows

$$(T_{ij} - T_iT_j)^2 \leq 4\sum_{x_i,x_j}\left|p_{X_iX_j|\underline{Y}}(x_ix_j \mid \underline{y}) - p_{X_i|\underline{Y}}(x_i \mid \underline{y})p_{X_j|\underline{Y}}(x_j \mid \underline{y})\right|^2$$

Applying the inequality

$$\frac{1}{2}\sum_x |P(x) - Q(x)|^2 \leq D(P\|Q)$$

for the Kullback-Leibler divergence of the two distributions $P = p_{X_iX_j|\underline{y}}$ and $Q = p_{X_i|\underline{Y}}p_{X_j|\underline{y}}$, we get for $i \neq j$

$$(T_{ij} - T_iT_j)^2 \leq 8I(X_i; X_j \mid \underline{y})$$

Averaging over the outputs we get

$$\frac{\partial^2}{\partial\epsilon_i^{-2}\partial\epsilon_j^{-2}} H_n(\underline{X} \mid \underline{Y}) = \mathbb{E}_{\underline{t}}[(T_{ij} - T_iT_j)^2] \leq 8I(X_i; X_j \mid \underline{Y})$$

Highly noisy BMS channels. From the high noise expansion (see section 3 and the above remarks, we can derive an inequality like the preceding one, which holds in the high noise regime for general BMS channels. The number 8 gets replaced by some suitable factor which depends on the channel noise.

1.4 Organisation of the paper

The statistical mechanics formulation is very convenient to perform many of the necessary calculations, but also the interpolation method is best formulated in that framework. Thus we briefly recall it in section 2 as well as a few connections to the information theoretic language. Section 3 contains the derivation of the correlation formula (proposition 2) and other useful material. The interpolation method that is used to prove the variational bound (theorem 1) is presented in section 4. The main new ingredient of the proof is an estimate (see proposition 3 in section 4) on the fluctuations of overlap parameters. The proof of proposition 3 is the object of section 5. The appendices contain technical calculations involved in the proofs.

2 Statistical Mechanics Formulation

Consider a fixed code belonging to the ensemble $\mathcal{C} = LDPC(n, \Lambda, P)$. The posterior distribution $p_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y})$ used in MAP decoding can be viewed as the

Gibbs measure of a particular random spin system. For this it is convenient to use the usual mapping of bits onto spins $\sigma_i = (-1)^{x_i}$. Given any set $A \subset \{1, ..., n\}$, we use the notation $\sigma_A = \prod_{i\in} \sigma_i$. Thus $\sigma_A = (-1)^{\oplus_{i\in A} x_i}$. It will be clear from the context if the subscript is a set or a single bit. For a uniform prior over the code words and a BMS channel, Bayes rule implies $p_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = \mu(\underline{\sigma})$ with

$$\mu(\underline{\sigma}) = \frac{1}{Z} \prod_c \frac{1}{2}(1 + \sigma_{\partial c}) \prod_{i=1}^{n} e^{\frac{l_i}{2}\sigma_i}$$

where $\prod_c$ is a product over all check nodes of the given code, and $\sigma_{\partial c} = \prod_{i\in\partial c} \sigma_i$ is the product of the spins (mod 2 sum of the bits) attached to the variable nodes i that are connected to a check c. Z is the normalization factor or "partition function" and $\ln Z$ is the "pressure" associated to the Gibbs measure $\mu(\underline{\sigma})$. It is related to the conditional entropy by

$$H_n(\underline{X}|\underline{Y}) = \mathbb{E}_{\underline{l}}[\ln Z] - \sum_{i=1}^{n} \int_{-\infty}^{+\infty} dl_i c_L(l_i) \frac{l_i}{2} \tag{12}$$

Expectations with respect to $\mu(\underline{\sigma})$ for a fixed graph and a fixed channel output are denoted by the bracket $\langle - \rangle$. More precisely for any $A \subset \{1, ..., n\}$,

$$\langle \sigma_A \rangle = \sum_{\sigma^n} \sigma_A \mu(\sigma^n), \qquad \sigma_A = \prod_{i\in A} \sigma_i$$

More details on the above formalism can be found for example in [18].

The soft estimate of the bit X_i is (in the difference domain)

$$T_i = \langle \sigma_i \rangle \tag{13}$$

We will also need soft estimates for $X_i \oplus X_j$, $i \neq j$. In the statistical mechanics formalism they are simply expressed as

$$T_{ij} = \langle \sigma_i \sigma_j \rangle \tag{14}$$

In particular the correlation between bits X_i and X_j becomes $T_{ij} - T_i T_j = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle$, which is the usual notion of spin-spin correlation in statistical mechanics.

In section 3 (and appendices B, C) the algebraic manipulations are best performed in terms of "extrinsic" soft bit estimates. We will need many variants, the simplest one being the estimate of X_i when observation y_i is not available

$$T_i^{\sim i} = \tanh \frac{L_i^{\sim i}}{2} = p_{X_i|\underline{Y}^{\sim i}}(0|\underline{y}^{\sim i}) - p_{X_i|\underline{Y}^{\sim i}}(1|\underline{y}^{\sim i})$$

The second is the estimate of X_i when both y_i and y_j are not available

$$T_i^{\sim ij} = \tanh \frac{L_i^{\sim ij}}{2} = p_{X_i|\underline{Y}^{\sim ij}}(0|\underline{y}^{\sim ij}) - p_{X_i|\underline{Y}^{\sim ij}}(1|\underline{y}^{\sim ij})$$

Finally we will also need the extrinsic estimate of the mod 2 sum $X_i \oplus X_j$ when both y_i and y_j are not available,

$$T_{ij}^{\sim ij} = \tanh \frac{L_{ij}^{\sim ij}}{2} = p_{X_i \oplus X_j|\underline{Y}^{\sim ij}}(0|\underline{y}^{\sim ij}) - p_{X_i \oplus X_j|\underline{Y}^{\sim ij}}(1|\underline{y}^{\sim ij})$$

It is practical to work in terms of a modified Gibbs average $\langle \sigma_A \rangle_{\sim i}$ which means that $l_i = 0$, in other words y_i is not available. Similarly we introduce the averages $\langle \sigma_X \rangle_{\sim ij}$, in other words both y_i and y_j are unavailable. One has

$$T_i^{\sim i} = \langle \sigma_i \rangle_{\sim i}, \qquad T_i^{\sim ij} = \langle \sigma_i \rangle_{\sim ij}, \qquad T_{ij}^{\sim ij} = \langle \sigma_i \sigma_j \rangle_{\sim ij}$$

The extrinsic brackets $\langle - \rangle_{\sim i}$ and $\langle - \rangle_{\sim ij}$ are related to the usual ones $\langle - \rangle$ by the following formulas derived in appendix A,

$$\langle \sigma_i \rangle_{\sim i} = \frac{\langle \sigma_i \rangle - t_i}{1 - \langle \sigma_i \rangle t_i} \tag{15}$$

and

$$\langle \sigma_i \rangle_{\sim ij} = \frac{\langle \sigma_i \rangle - t_i - \langle \sigma_i \sigma_j \rangle t_j + t_i t_j \langle \sigma_j \rangle}{1 - \langle \sigma_i \rangle t_i - \langle \sigma_j \rangle t_j + \langle \sigma_i \sigma_j \rangle t_i t_j} \tag{16}$$

$$\langle \sigma_i \sigma_j \rangle_{\sim ij} = \frac{\langle \sigma_i \sigma_j \rangle - t_i \langle \sigma_j \rangle - \langle \sigma_i \rangle t_j + t_i t_j}{1 - \langle \sigma_i \rangle t_i - \langle \sigma_j \rangle t_j + \langle \sigma_i \sigma_j \rangle t_i t_j} \tag{17}$$

3 The Correlation Formula

A derivation of propositon 1 and of (4), (6) within the formalism outlined in section 2 can be found in [19].

3.1 Proof of proposition 2

For any BMS($\underline{\epsilon}$) channel and linear code we have from (12)

$$\frac{\partial}{\partial\epsilon_i} H_n(\underline{X} \mid \underline{Y}) = \mathbb{E}_{\underline{l}^{\sim j}}\left[\int_{-\infty}^{+\infty} dl_j \frac{\partial c_L(l_j)}{\partial\epsilon_j}(\ln Z - \frac{l_j}{2})\right]$$

The second equality follows by permutation symmetry of code bits. Differentiating once more, we get

$$\frac{\partial^2}{\partial\epsilon_i\partial\epsilon_j} H_n(\underline{X} \mid \underline{Y}) = \delta_{ij} S_1 + (1-\delta_{ij}) S_2 \tag{18}$$

where

$$S_1 = \mathbb{E}_{\underline{l}^{\sim i}}\left[\int_{-\infty}^{+\infty} dl_i \frac{\partial^2 c_L(l_i)}{\partial\epsilon_i^2}(\ln Z - \frac{l_i}{2})\right] \tag{19}$$

and

$$S_2 = \mathbb{E}_{\underline{l}^{\sim ij}}\left[\int_{-\infty}^{+\infty} dl_i dl_j \frac{\partial c_L(l_i)}{\partial\epsilon_i}\frac{\partial c_L(l_j)}{\partial\epsilon_j}(\ln Z - \frac{l_i}{2})\right]$$

First we consider S_1. Let

$$Z_{\sim i} = \sum_{\underline{\sigma}} \prod_{c\in\mathcal{C}} \frac{1}{2}(1+\sigma_{\partial c}) \prod_{k\neq i} e^{\frac{l_k}{2}\sigma_k}$$

be the partition function for the Gibbs measure $\langle - \rangle_{\sim i}$ introduced in section 2 and consider

$$\ln\frac{Z}{Z_{\sim i}} = \ln\langle e^{\frac{l_i}{2}\sigma_i}\rangle_{\sim i}$$

Using the identity

$$e^{\frac{l_i}{2}\sigma_i} = e^{\frac{l_i}{2}}\frac{1+t_i\sigma_i}{1+t_i} \tag{20}$$

we get

$$\ln Z - \frac{l_i}{2} = \ln Z_{\sim i} + \ln\left(\frac{1+t_i\langle\sigma_i\rangle_{\sim i}}{1+t_i}\right)$$

When we replace this expression in the integral (19) we see that the contribution of $\ln Z_{\sim i}$ vanishes because this later quantity is independent of l_i. Indeed

$$\int_{-\infty}^{+\infty} dl_i \frac{\partial^2 c_L(l_i)}{\partial\epsilon_i^2}\ln Z_{\sim i} = \ln Z_{\sim i}\frac{\partial^2}{\partial\epsilon_i^2}\int_{-\infty}^{+\infty} dl_1 c_L(l_i) = 0$$

since $c_L(l_i)$ is a normalized probability distribution. Then, using (15) leads to

$$S_1 = \int_{-1}^{+1} dt_i \frac{\partial^2 c_D(t_i)}{\partial \epsilon_i^2} \mathbb{E}_{\underline{t}^{\sim i}} \left[\ln\left(\frac{1 + t_i \langle \sigma_i \rangle_{\sim i}}{1 + t_i} \right) \right] \tag{21}$$

$$= -\int_{-1}^{+1} dt_i \frac{\partial^2 c_D(t_i)}{\partial \epsilon_i^2} \mathbb{E}_{\underline{t}^{\sim i}} \left[\ln\left(\frac{1 - t_i \langle \sigma_i \rangle}{1 - t_i} \right) \right] \tag{22}$$

which (because of (13)) coincides with the first term in the correlation formula.

Now we consider the term S_2. Notice that

$$\int_{-\infty}^{+\infty} dl_i dl_j \frac{\partial c_L(l_i)}{\partial \epsilon_i} \frac{\partial c_L(l_j)}{\partial \epsilon_j} \frac{l_j}{2} = \int_{-\infty}^{+\infty} dl_j \frac{\partial c_L(l_j)}{\partial \epsilon_j} \frac{l_j}{2} \frac{\partial}{\partial \epsilon_i} \int_{-\infty}^{+\infty} dl_i c_L(l_i) = 0$$

Thus we can rewrite S_2 as

$$S_2 = \mathbb{E}_{\underline{l}^{\sim ij}} \left[\int_{-\infty}^{+\infty} dl_i dl_j \frac{\partial c_L(l_i)}{\partial \epsilon_i} \frac{\partial c_L(l_j)}{\partial \epsilon_j} (\ln Z - \frac{l_i}{2} - \frac{l_j}{2}) \right]$$

Let $Z_{\sim ij} = \sum_{\underline{\sigma}} \prod_{c \in \mathcal{C}} \frac{1}{2}(1 + \sigma_{\partial c}) \prod_{k \neq i,j} e^{\frac{l_k}{2}\sigma_k}$ be the partition function for the Gibbs measure $\langle \cdot \rangle_{\sim ij}$, and consider

$$\ln \frac{Z}{Z_{\sim ij}} = \ln \langle e^{\frac{l_i}{2}\sigma_i + \frac{l_j}{2}\sigma_j} \rangle_{\sim ij}$$

Using again (20) we get

$$\ln Z - \frac{l_i}{2} - \frac{l_j}{2} = \ln Z_{\sim ij} + \ln \left(\frac{1 + t_i \langle \sigma_i \rangle_{\sim ij} + t_j \langle \sigma_j \rangle_{\sim ij} + t_i t_j \langle \sigma_i \sigma_j \rangle_{\sim ij}}{1 + t_i + t_j + t_i t_j} \right)$$

As before the contribution of $\ln Z_{\sim ij}$ vanishes because it is independent of l_i, l_j. Similarly we have

$$\begin{aligned} \int_{-\infty}^{+\infty} dl_i dl_j \frac{\partial c_L(l_i)}{\partial \epsilon_i} \frac{\partial c_L(l_j)}{\partial \epsilon_j} \ln(1 + t_i \langle \sigma_i \rangle_{\sim ij}) &= \text{same with i and j exchanged} \\ &= 0 \end{aligned}$$

$$\begin{aligned} \int_{-\infty}^{+\infty} dl_i dl_j \frac{\partial c(l_i)}{\partial \epsilon_i} \frac{\partial c(l_j)}{\partial \epsilon_j} \ln(1 + t_i) &= \text{same with i and j exchanged} \\ &= 0 \end{aligned}$$

Using these four identities then leads to

$$\begin{aligned} S_2 =& \mathbb{E}_{\underline{l}^{\sim ij}}\Big[\int_{-1}^{+1} dt_i dt_j \frac{\partial c_D(t_i)}{\partial \epsilon_i}\frac{\partial c_D(t_j)}{\partial \epsilon_j} \\ &\times \ln\Big(\frac{1 + t_i\langle\sigma_i\rangle_{\sim ij} + t_j\langle\sigma_j\rangle_{\sim ij} + t_i t_j\langle\sigma_i\sigma_j\rangle_{\sim ij}}{1 + t_i\langle\sigma_i\rangle_{\sim ij} + t_j\langle\sigma_j\rangle_{\sim ij} + t_i t_j\langle\sigma_i\rangle_{\sim ij}\langle\sigma_j\rangle_{\sim ij}}\Big)\Big] \end{aligned} \tag{23}$$

To get the formulas in terms of usual averages we use the relations (16), (17). Hence

$$S_2 = \mathbb{E}_{\underline{l}^{\sim ij}}\Big[\int_{-1}^{+1} dt_i dt_j \frac{\partial c_D(t_i)}{\partial \epsilon_i}\frac{\partial c_D(t_j)}{\partial \epsilon_j} \ln\Big(\frac{1 - t_i\langle\sigma_i\rangle - t_j\langle\sigma_j\rangle + t_i t_j\langle\sigma_i\sigma_j\rangle}{1 - t_i\langle\sigma_i\rangle - t_j\langle\sigma_j\rangle + t_i t_j\langle\sigma_i\rangle\langle\sigma_j\rangle}\Big)\Big] \tag{24}$$

Because of (13) and (14) this coincides with the second term in the correlation formula. The proposition now follows from (18), (22) and (24).

3.2 Expressions in terms of the spin-spin correlation

The BEC. From $c_D(t) = (1-\epsilon)\delta(t-1) + \epsilon\delta(t)$, the second derivative in terms of extrinsic quantities (formulas (21) and (23)) reduces to

$$\frac{\partial^2}{\partial\epsilon_i\partial\epsilon_j} H_n(\underline{X} \mid \underline{Y}) = (1-\delta_{ij})\mathbb{E}_{\underline{t}^{\sim ij}}\Big[\ln\Big(\frac{1 + \langle\sigma_i\rangle_{\sim ij} + \langle\sigma_j\rangle_{\sim ij} + \langle\sigma_i\sigma_j\rangle_{\sim ij}}{1 + \langle\sigma_i\rangle_{\sim ij} + \langle\sigma_j\rangle_{\sim ij} + \langle\sigma_i\rangle_{\sim ij}\langle\sigma_j\rangle_{\sim ij}}\Big)\Big]$$

There are various ways to see that for the BEC any Gibbs average $\langle\sigma_A\rangle$ or $\langle\sigma_A\rangle_{\sim ij}$ takes values in $\{0, 1\}$. A heuristic explanation is that bits (or their mod 2 sums) are either perfectly known or erased. A more formal explanation follows from a Nishimori identity[5] combined with the Griffith-Kelly-Sherman (GKS) correlation inequality [18]. For example, $\mathbb{E}[\langle\sigma_A\rangle^2] = \mathbb{E}[\langle\sigma_A\rangle]$ (Nishimori) and $\langle\sigma_A\rangle \geq 0$ (GKS). Thus $\langle\sigma_A\rangle(1-\langle\sigma_A\rangle)$ is a positive random variable with zero expectation and is therefore equal to 0 with probability one. These remarks imply that

$$\frac{\partial^2}{\partial\epsilon_i\partial\epsilon_j} H_n(\underline{X} \mid \underline{Y}) = \frac{1}{\epsilon_i\epsilon_j}(1-\delta_{ij})\mathbb{E}_{\underline{t}}\Big[\ln\Big(\frac{1 + \langle\sigma_i\rangle + \langle\sigma_j\rangle + \langle\sigma_i\sigma_j\rangle}{1 + \langle\sigma_i\rangle + \langle\sigma_j\rangle + \langle\sigma_i\rangle\langle\sigma_j\rangle}\Big)\Big]$$

[5] We will use various such identities. A proof of their most general form can be found in [18]. A general reference is [21].

Note that in deriving the last expression we used the fact that $l_i = \infty$ ($l_j = \infty$) implies that $\sigma_i = +1$ ($\sigma_j = +1$) which makes the logarithm term equal to zero. From the previous remarks we also have

$$\begin{aligned}\ln\big(1 + \langle\sigma_i\rangle + \langle\sigma_j\rangle + \langle\sigma_i\sigma_j\rangle\big) =& (\ln 2)\big(\langle\sigma_i\rangle + \langle\sigma_j\rangle + \langle\sigma_i\sigma_j\rangle\big)\\ &+ (\ln 3 - 2\ln 2)\big(\langle\sigma_i\rangle\langle\sigma_j\rangle + \langle\sigma_i\rangle\langle\sigma_i\sigma_j\rangle + \langle\sigma_j\rangle\langle\sigma_i\sigma_j\rangle\big)\\ &+ (5\ln 2 - 3\ln 3)\langle\sigma_i\rangle\langle\sigma_j\rangle\langle\sigma_i\sigma_j\rangle\end{aligned}$$

and

$$\ln\big(1 + \langle\sigma_i\rangle + \langle\sigma_j\rangle + \langle\sigma_i\rangle\langle\sigma_j\rangle\big) = (\ln 2)\big(\langle\sigma_i\rangle + \langle\sigma_j\rangle\big)$$

The difference of the two logarithms is simplified using the following four Nishimori identities,

$$\mathbb{E}_{\underline{t}}[\langle\sigma_i\rangle\langle\sigma_j\rangle] = \mathbb{E}_{\underline{t}}[\langle\sigma_i\rangle\langle\sigma_i\sigma_j\rangle] = \mathbb{E}_{\underline{t}}[\langle\sigma_j\rangle\langle\sigma_i\sigma_j\rangle] = \mathbb{E}_{\underline{t}}[\langle\sigma_i\sigma_j\rangle\langle\sigma_i\rangle\langle\sigma_j\rangle]$$

Finaly we obtain the simple expression

$$\begin{aligned}\frac{\partial^2}{\partial\epsilon_i\partial\epsilon_j}H_n(\underline{X} \mid \underline{Y}) &= \frac{\ln 2}{\epsilon_i\epsilon_j}(1 - \delta_{ij})\mathbb{E}_{\underline{t}}\big[\langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle\big]\\ &= \frac{\ln 2}{\epsilon_i\epsilon_j}(1 - \delta_{ij})\mathbb{E}_{\underline{t}}\big[T_{ij} - T_iT_j\big]\end{aligned}$$

Let us point out that the second GKS inequality (for the BEC) implies that $\langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle \geq 0$, thus the correlation takes values in $\{0, 1\}$ and we have $\mathbb{E}_{\underline{t}}\big[T_{ij} - T_iT_j\big] = \mathbb{E}_{\underline{t}}\big[(T_{ij} - T_iT_j)^2\big]$.

The BIAWGNC. From the explicit form

$$c_L(l) = \frac{1}{\sqrt{2\pi\epsilon^{-2}}}\exp\left(-\frac{(l - \epsilon^{-2})^2}{2\epsilon^{-2}}\right)$$

one can show that the correlation formula reduces to

$$\begin{aligned}\frac{\partial^2}{\partial\epsilon_i^{-2}\partial\epsilon_j^{-2}}H_n(\underline{X} \mid \underline{Y}) &= \mathbb{E}_{\underline{t}}[(\langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle)^2]\\ &= \mathbb{E}_{\underline{t}}\big[\big(T_{ij} - T_iT_j\big)^2\big]\end{aligned}$$

Otherwise diffrentiating (12) thanks to

$$\frac{d^2 c_L(l)}{(d\epsilon^{-2})^2} = \left(-\frac{\partial}{\partial l} + \frac{\partial^2}{\partial l^2}\right)^2 c_L(l)$$

and using integration by parts also leads to this simpler form. This route is much simpler and the details can be found in [18].

Highly noisy BMS channels. We use the extrinsic form of the correlation formula given by (21) and (23). First we expand the logarithms in S_1 and S_2 in powers of t_i and t_j and then use various Nishimori identities. After some tedious algebra (see Appendices B and C) we can organize the expansion in powers of the channel parameters (2). In the high noise regime this expansion is absolutely convergent. To lowest order we have

$$\begin{aligned}
&\frac{\partial^2}{\partial\epsilon_i\partial\epsilon_j} H_n(\underline{X} \mid \underline{Y}) = \delta_{ij} S_1 + (1-\delta_{ij}) S_2 \\
&\approx \frac{1}{2}\delta_{ij} {m_2}^{(2)} (\mathbb{E}_{\underline{l}}[\langle\sigma_i\rangle^2] - 1) + \frac{1}{2}(1-\delta_{ij})[{m_1}^{(2)}]^2 \mathbb{E}_{\underline{l}}\Big[\big(\langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle\big)^2\Big] + \dots \\
&= \frac{1}{2}\delta_{ij} {m_2}^{(2)} (\mathbb{E}_{\underline{t}}[T_i^2] - 1) + \frac{1}{2}(1-\delta_{ij})[{m_1}^{(2)}]^2 \mathbb{E}_{\underline{t}}\Big[\big(T_{ij} - T_i T_j\big)^2\Big] + \dots \qquad (25)
\end{aligned}$$

The second derivative of the conditional entropy is directly related to the average square of the code-bit or spin-spin correlation.

4 The Interpolation Method

We use the interpolation method in the form developed by Montanari. As explained in [7] it is difficult to establish directly the bounds for the standard ensembles. Rather, one introduces a "multi-Poisson" ensemble which approximates the standard ensemble. Once the bounds are derived for the multi-Poisson ensemble they are extended to the standard ensemble by a limiting procedure. The interpolation construction is fairly complicated so that it helpful to briefly review the simpler pure Poisson case.

4.1 Poisson ensemble

We introduce the ensemble Poisson-LDPC$(n, 1-r, P) = \mathcal{P}$ where n is the block length, r the rate and $P(x) = \sum_k P_k x^k$ the check degree distribution.

A bipartite graph from the Poisson ensemble is constructed as follows. The graph has n variable nodes. For any k choose a Poisson number m_k of check nodes with mean $n(1-r)P_k$. Thus graph has a total of $m = \sum_k m_k$ check nodes which is also a Poisson variable with mean $n(1-r)$. For each check node c of degree k, choose k variable nodes uniformly at random and connect them to c. One can show that the left degree distribution concentrates around a Poisson distribution $\Lambda_{\mathcal{P}}(x) = e^{P'(1)(1-r)(x-1)}$. In other words the fraction Λ_l of variable nodes with degree l is Poisson with mean $P'(1)(1-r)$.

The main idea behind the interpolation technique is to recursively remove the check node constraints and compensate them with extra observations U distributed as (1) where d_V is a trial distribution to be optimized in the final inequality. One can interpret these extra observations as coming from a repetition code whose rate is tuned in a such a way that the total design rate r remains fixed. More precisely let $s \in [0,1]$ be an interpolating parameter. At "time" s we have a Poisson-LDPC$(n,(1-r)s,P) = \mathcal{P}_s$ code. Besides the usual channel outputs l_i, each node i receives e_i extra i.i.d observations U_a^i, $a = 1,...,e_i$, where e_i is Poisson with mean $n(1-r)(1-s)$ (so the total effective rate is fixed to r). The interpolating Gibbs measure is

$$\mu_s(\underline{\sigma}) = \frac{1}{Z_s}\prod_c \frac{1}{2}(1+\sigma_{\partial c})\prod_{i=1}^n e^{(\frac{l_i}{2}+\sum_{a=1}^{e_i} U_a^i)\sigma_i} \tag{26}$$

Here $\prod_c$ is a product over checks of a given graph in the ensemble $\mathcal{P}_s$. At $s=1$ one recovers the original measure while at $s=0$ (no checks) we have a simple product measure (corresponding to a repetition code) which is tailored to yield the replica symmetric entropy $h_{RS}[d_V;\Lambda_{\mathcal{P}},P]$ (up to an extra constant).

The central result of [7] is the sum rule

$$\mathbb{E}_{\mathcal{P}}[h_n] = h_{RS}[d_V;\Lambda_{\mathcal{P}},P] + \int_0^1 R_n(s)ds \tag{27}$$

Let us explain the notation. The first term on the right hand side $h_{RS,\mathcal{P}}[d_V;\Lambda_{\mathcal{P}},P]$ is the replica symmetric functional of section 1 evaluated for the Poisson ensemble. The remainder term $R_n(s)$ is

$$R_n(s) = \sum_{p=1}^{\infty}\frac{1}{2p(2p-1)}\mathbb{E}_s\Big[\big\langle P(Q_{2p}) - P'(q_{2p})(Q_{2p}-q_{2p}) - P(q_{2p})\big\rangle_{2p,s}\Big]$$

with $q_{2p} = \mathbb{E}_V[(\tanh V)^{2p}]$ and Q_{2p} the overlap parameters

$$Q_{2p} = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^{(1)}\sigma_i^{(2)}\cdots\sigma_i^{(2p)} \tag{28}$$

Here $\sigma_i^{(\alpha)}, \alpha = 1, 2, \ldots, 2p$ are $2p$ independent copies (replicas) of the spin σ_i and $\langle - \rangle_{2p,s}$ is the Gibbs bracket associated to the product measure (replica measure)

$$\prod_{\alpha=1}^{2p} \mu_s(\underline{\sigma}^{(\alpha)})$$

4.2 Multi-Poisson ensemble

The multi-Poisson-LDPC$(n, \Lambda, P, \gamma) = \mathcal{MP}$ ensemble, is a more elaborate construction which allows to approximate a target LDPC(n, Λ, P) ensemble. Its parameters are the block length n, the target variable and check node degree distributions $\Lambda(x)$ and $P(x)$ and the real number γ which controls the closeness to the standard ensemble. We recall that variable and check node degrees have finite maximum degrees. The construction of a bipartite graph from the multi-Poisson ensemble proceeds via rounds: the process starts with a high rate code and at each round one adds a very small number of check nodes till one ends up with a code

with almost the desired rate and degree distribution. A graph process $\mathcal{G}_t$ is defined for discrete times $t = 0, ..., t_{max}$, $t_{max} = \lfloor \Lambda'(1)/\gamma \rfloor - 1$ as follows. For $t = 0$, $\mathcal{G}_0$ has no check nodes and has n variable nodes. The set of variable nodes is partitioned into the subsets $\mathcal{V}_l$ of cardinality $n\Lambda_l$ for every l and every node $i \in \mathcal{V}_l$ is decorated with l free sockets. The number $d_i(t)$ keeps track of the number of free sockets on node i once round t is completed. So for $t = 0$, $\mathcal{G}_0$ has no check nodes and each variable node $i \in \mathcal{V}_l$ has $d_i(0) = l$ free sockets. At round t, $\mathcal{G}_t$ is constructed from $\mathcal{G}_{t-1}$ as follows. For all k, choose a Poisson number m_k^t of check nodes with mean $n\gamma P_k/P'(1)$. Connect each outgoing edge of these new degree k check nodes (added at time t) to variable node i according to the probability $w_i(t) = \frac{d_i(t-1)}{\sum_i d_i(t-1)}$. This is the fraction of free sockets at node i after round $t-1$ was completed. Once all new check nodes are connected, update the number of free sockets for each variable node $d_i(t) = d_i(t-1) - \Delta_i(t)$. where $\Delta_i(t)$ is the number of times the variable node i was chosen during the round t. For $n \to \infty$

this construction yields graphs with variable degree distributions $\Lambda_\gamma(x)$ (the check degree distribution remains $P(x)$). The variational distance between $\Lambda_\gamma(x)$ and $P(x)$ tends to zero as $\gamma \to 0$.

The interpolating ensemble now uses two parameters (t_*, s) where $t_* \in \{0, ..., t_{max}\}$ and $0 \leq s \leq \gamma$. For rounds $0, ..., t_* - 1$ one proceeds exactly as before to obtain a graph $\mathcal{G}_{t_*-1}$. At the next round t_*, one proceeds as before but with γ replaced by s. The rate loss is compensated by adding e_i extra observations for each node i, where e_i is a Poisson integer with mean $n(\gamma - s)w_i(t_*)$. The round is ended by updating the number of free sockets $d_i(t_*) = d_i(t_* - 1) - \Delta_i(t_*) - e_i(t_*)$. Finally, for rounds after $t_* + 1, ..., t_{max}$ no new check node is added but for each variable node i, e_i external observations are added, where e_i is a Poisson integer with mean $n\gamma w_i(t_*)$. Moreover the free socket counter is updated as $d_i(t) = d_i(t-1) - e_i(t)$. Recall that the external observations are i.i.d copies of the random variable U (see (1)).

The interpolating Gibbs measure $\mu_{t_*,s}(\underline{\sigma})$ has the same form than (26) with the appropriate products over checks and extra observations. Let $h_{n,\gamma}$ the conditional entropy of the multi-Poisson ensemble $\mathcal{MP}$ (corresponding to $t_* = t_{max}$ and $s = \gamma$). Again, the central result of [7] is the sum rule

$$\mathbb{E}_{\mathcal{MP}}[h_{n,\gamma}] = h_{RS}[d_V; \Lambda_\gamma, P] + \sum_{t_*=0}^{t_{max}-1} \int_0^\gamma R_n(t_*, s) ds + o_n(1) \tag{29}$$

Explanations on the notation are in order. The first term $h_{RS,\gamma}[d_V; \Lambda_\gamma, P]$ is the replica symmetric functional of 1 evaluated for the multi-Poisson ensemble. The remainder term $R_n(t_*, s)$ is given by

$$R_n(t_*, s) = \sum_{p=1}^{\infty} \frac{1}{(2p)(2p-1)} \mathbb{E}_s \Big[\big\langle P(Q_{2p}) - P'(q_{2p})(Q_{2p} - q_{2p}) - P(q_{2p}) \big\rangle_{2p,t_*,s} \Big] \tag{30}$$

where $q_{2p} = \mathbb{E}_V[(\tanh V)^{2p}]$ as before and Q_{2p} are modified overlap parameters

$$Q_{2p} = \sum_{i=1}^{n} w_i(t_*) X_i(t_*) \sigma_i^{(1)} \sigma_i^{(2)} \cdots \sigma_i^{(2p)} \tag{31}$$

Here as before $\sigma_i^{(\alpha)}, \alpha = 1, 2, \ldots, 2p$ are $2p$ independent copies (replicas) of the spin σ_i and $\langle - \rangle_{2p,t_*,s}$ is the Gibbs bracket associated to the product measure

$$\prod_{\alpha=1}^{2p} \mu_{t_*,s}(\underline{\sigma}^{(\alpha)})$$

The overlap parameter is now more complicated than in the Poisson case because of the (positive) terms $w_i(t_*)$ and $X_i(t_*)$. Here $X_i(t_*)$ are new i.i.d random variables whose precise description is quite technical and can be found in [7]. The reader may think of the terms $w_i(t_*)X_i(t_*)$ as behaving like the $\frac{1}{n}$ factor of the pure Poisson ensemble overlap parameter (28). More precisely the only properties (see Appendix E in [7]) that we need are

$$\sum_{i=1}^{n} w_i(t_*) = 1, \qquad \mathbb{P}\big[w_i(t_*) \leq \frac{A}{n}\big] \geq 1 - e^{-Bn} \tag{32}$$

and

$$0 \leq X_i(t_*) \leq x, \qquad \mathbb{E}[x^k] \leq A_k \tag{33}$$

for any finite k and finite positive constants A, B, A_k independent of n (they may depend on some of the other parameters but this turns out to be unimportant). Finaly we use the shorthand $\mathbb{E}_s[-]$ for the expectation with respect to all random variables involved in the interpolation measure. The subscript s is here to remind us that this expectation depends on s, afact that is important to keep in mind because the remainder involves an intgral over s. When we use $\mathbb{E}$ (without the subscript s; as in (33) for example) it means that the quantity does not depend on s. In the sequel the replcated Gibbs bracket $\langle - \rangle_{2p,t_*,s}$ is simply denoted by $\langle - \rangle_s$. There will be no risk of confusion because the only property that we us is its linearity.

In [7] it is shown that

$$\mathbb{E}_{\mathcal{C}}[h_n] = \mathbb{E}_{\mathcal{MP}}[h_{n,\gamma}] + O(\gamma^b) + o_n(1) \tag{34}$$

where $O(\gamma^b)$ is uniform in n ($b > 0$ a numerical constant) and $o_n(1)$ (depends on γ) tends to 0 as $n \to +\infty$.

In the next paragraph we prove the variational bound on the conditional entropy of the multi-Poisson ensemble, namely

$$\liminf_{n\to+\infty} \mathbb{E}_{\mathcal{MP}}[h_{n,\gamma}] \geq h_{RS}[d_V; \Lambda_\gamma, P] \tag{35}$$

Note that here $o_n(1)$ again depends on γ. By combining this bound with (34) and taking limits

$$\liminf_{n\to+\infty} \mathbb{E}_{\mathcal{C}}[h_n] = \lim_{\gamma\to 0}\liminf_{n\to+\infty} \mathbb{E}_{\mathcal{MP}}[h_{n,\gamma}] \geq \lim_{\gamma\to 0} h_{RS}[d_V;\Lambda_\gamma, P] = h_{RS}[d_V;\Lambda, P] \tag{36}$$

The main theorem 1 then follows by maximizing the right hand side over d_V.

4.3 Proof of the Variational Bound (35)

In view of the sum rule (29) it is sufficient to prove that $\liminf_{n\to+\infty} R_n(t_*, s) \geq 0$. In the case of a convex P considered in [7] this is immediate because convexity is equivalent to

$$P(Q_{2p}) - P(q_{2p}) \geq P'(q_{2p})(Q_{2p} - q_{2p})$$

Note that $P(x) = \sum_k P_k x^k$ is anyway convex for $x \geq 0$ since all $P_k \geq 0$. So if do not assume convexity of the check node degree distribution we have to circumvent the fact that Q_{2p} can be negative. But note

$$\begin{aligned}\langle Q_{2p}\rangle &= \sum_{i=1}^n w_i(t_*)X_i(t_*)\langle\sigma_i^{(1)}\sigma_i^{(2)}\cdots\sigma_i^{(2p)}\rangle \\ &= \sum_{i=1}^n w_i(t_*)X_i(t_*)\langle\sigma_i^{(1)}\rangle\langle\sigma_i^{(2)}\rangle\cdots\langle\sigma_i^{(2p)}\rangle \\ &= \sum_{i=1}^n w_i(t_*)X_i(t_*)\langle\sigma_i\rangle^{2p} \geq 0\end{aligned}$$

Therefore we are assured that for any P (i.e not necessarily convex for $x \in \mathbb{R}$) we have

$$P(\langle Q_{2p}\rangle) - P(q_{2p}) \geq P'(q_{2p})(\langle Q_{2p}\rangle - q_{2p}) \tag{37}$$

and the proof will follow if we can show that with high probability

$$P(Q_{2p}) \approx P(\langle Q_{2p}\rangle)$$

The following concentration estimate will suffice and is proven in section 5.

Proposition 3. *Fix any $\delta < \frac{1}{4}$. On the BEC(ϵ) and BIAWGNC(ϵ) for a.e ϵ, or on general BMS(ϵ) satisfying H, we have for a.e ϵ,*

$$\lim_{n\to\infty}\int_0^\gamma ds\mathbb{P}_s\left[|P(Q_{2p}) - P(\langle Q_{2p}\rangle_s)| > \frac{2p}{n^\delta}\right] = 0 \tag{38}$$

Here $\mathbb{P}_s(X)$ *is the probability distribution* $\mathbb{E}_s\langle \mathbb{I}_X\rangle_s$*.*

This proposition can presumably be strengthened in two directions. First we conjecture that hypothesis H is not needed (this is indeed the case for the BEC and BIAWGNC). Secondly the statement should hold for all ϵ except at a finite set of threshold values of ϵ where the conditional entropy is not differentiable, and its first derivative is expected to have jumps (except for cycle codes where higher order derivatives are singular). Since we are unable to control the locations of theses jumps our proof only works for Lebesgue almost every ϵ.

We are now ready to complete the proof of the variational bound (35).

End of Proof of (35). From (31) and (33)

$$|Q_{2p}| \leq \sum_{i=1}^{n} w_i(t_*)X_i(t_*) \leq x \tag{39}$$

and

$$\mathbb{E}_s[\langle Q_{2p}^k\rangle_s] \leq A_k \tag{40}$$

Combined with $q_{2p} \leq 1$, this implies (since the maximal degree of P is finite) that

$$\mathbb{E}_s[\langle P(Q_{2p}) - P'(q_{2p})(Q_{2p} - q_{2p}) - P(q_{2p})\rangle_s] \leq C_1 \tag{41}$$

for some positive constant C_1. The only crucial feature here is that this constant does not depend on n and on the number of replicas $2p$ (a more detailed analysis shows that it depends only on the degree of $P(x)$).

Now we split the sum (30) into terms with $1 \leq p \leq n^\delta$ (call this contribution R_A) and terms with $p \geq n^\delta$ (call this contribution R_B), where $\delta > 0$ is the constant of proposition 3. For the second contribution (41) implies

$$R_B \leq C_1 \sum_{p\geq n^\delta} \frac{1}{2p(2p-1)} = O(n^{-\delta}) \tag{42}$$

For the first contribution we write

$$\begin{aligned} R_A = &\sum_{p\leq n^\delta} \frac{1}{2p(2p-1)}\mathbb{E}_s[\langle P(Q_{2p})\rangle_s - P(\langle Q_{2p}\rangle_s)] \\ &+ \sum_{p\leq n^\delta} \frac{1}{2p(2p-1)}\mathbb{E}[P(\langle Q_{2p}\rangle_s) - P'(q_{2p})(\langle Q_{2p}\rangle_s - q_{2p}) - P(q_{2p})] \end{aligned}$$

In this equation, the second sum is positive due to (37). Thus we find

$$
\begin{aligned}
R_n(t_*, s) &= R_A + R_B \\
&\geq \sum_{p \leq n^\delta} \frac{1}{2p(2p-1)} \mathbb{E}[\langle P(Q_{2p})\rangle_s - P(\langle Q_{2p}\rangle_s)] - O(n^{-\delta})
\end{aligned}
$$

Below we use proposition 3 to show that for almost every ϵ in the appropriate range

$$
\lim_{n\to+\infty} \int_0^\gamma ds \sum_{p \leq n^\delta} \frac{1}{2p(2p-1)} \mathbb{E}_s[\langle P(Q_{2p})\rangle_s - P(\langle Q_{2p}\rangle_s)] = 0 \tag{43}
$$

which implies by Fatou's lemma

$$
\liminf_{n\to+\infty} \sum_{t_*=0}^{t_{max}-1} \int_0^\gamma R_n(t_*, s) ds \geq 0
$$

and thus proves (35) for almost every ϵ in the appropriate range. A general convexity argument allows to extend this result to all ϵ in the same range. Indeed convexity arguments imply that both sides of the inequality (35) are continuous functions of ϵ[6]. To show continuity of the left hand side we use inequality (56) in Appendix C: it implies that there exists a positive number ρ (independent of ϵ and n) such that $\frac{d^2}{d\epsilon^2}\mathbb{E}_s[h_{n,\gamma}] \geq -\rho$. Therefore $\mathbb{E}_s[h_{n,\gamma}] + \frac{\rho}{2}\epsilon^2$ is convex in ϵ; so the $\liminf_{n\to+\infty}$ is also convex and thus continuous on any open ϵ set. To show continuity of the right hand side we first note that for each d_V, h_{RS} is a linear functional of the channel distribution $c_L(l)$; thus the $\sup_{d_V}$ is a convex functional of $c_L(l)$; thus it is continuous in any open ϵ where $c_L(l)$ varies smoothly in ϵ (this last point can be made more precise using tools from functional analysis).

Let us now prove (43). First we set

$$
F_{2p} = \left| \langle P(Q_{2p})\rangle_s - P(\langle Q_{2p}\rangle_s) \right|
$$

[6]At this point one could use arguments involving physical degradation if BMS(ϵ) is degraded as a function of ϵ. But we take a more direct route that does not assume physical degraddation as a function of ϵ

and use Cauchy-Schwarz and then (40) to obtain

$$\begin{aligned}\mathbb{E}[F_{2p}] &= \mathbb{E}_s\big[F_{2p}\mathbb{I}_{F_{2p}\le\frac{2p}{n^\delta}}\big] + \mathbb{E}_s\big[F_{2p}\mathbb{I}_{F_{2p}\ge\frac{2p}{n^\delta}}\big] \\ &\le \frac{2p}{n^\delta} + \mathbb{E}_s\big[F_{2p}^2\big]^{1/2}\mathbb{P}_s\big[F_{2p}\ge\frac{2p}{n^\delta}\big]^{1/2} \\ &\le \frac{2p}{n^\delta} + C_2\mathbb{P}_s\big[F_{2p}\ge\frac{2p}{n^\delta}\big]^{1/2}\end{aligned}$$

for some positive constant C_2 independent of n and p (depending only on the degree of $P(x)$). Thus

$$\begin{aligned}&\int_0^\gamma ds\sum_{p\le n^\delta}\frac{1}{2p(2p-1)}\mathbb{E}[F_{2p}] \\ &\quad\le \frac{1}{n^\delta}\sum_{p\le n^\delta}\frac{1}{2p-1} + C_2\int_0^\gamma ds\sum_{p\le n^\delta}\frac{1}{2p(2p-1)}\mathbb{P}_s\big[F_{2p}\ge\frac{2p}{n^\delta}\big]^{1/2} \\ &\quad\le O\big(\frac{\ln n^\delta}{n^\delta}\big) + C_2\sum_{p\le n^\delta}\frac{\sqrt{\gamma}}{2p(2p-1)}\bigg(\int_0^\gamma ds\mathbb{P}_s\big[F_{2p}\ge\frac{2p}{n^\delta}\big]\bigg)^{1/2}\end{aligned}$$

In the second inequality we have permuted the integral with a finite sum and used Cauchy-Schwarz. Finaly we can apply proposition 3 and Lebesgue's dominated convergence theorem to the last sum over p, to conclude that (43) holds.

5 Fluctuations of overlap parameters

In this section we prove proposition 3. The proofs are done directly for the multi-Poisson ensemble. We start by a relation between the overlap fluctuation and the spin-spin correlation.

Lemma 1. *For any BMS(ϵ) channel there exists a finite constant C_3 independent of n and p (dpending only on the maximal check degree) such that*

$$\mathbb{P}_s\bigg[\big|P(Q_{2p})-P(\langle Q_{2p}\rangle_s)\big|\ge\frac{2p}{n^\delta}\bigg] \le \frac{C_3}{p^2n^{2\delta-\frac{1}{2}}}\bigg(\sum_{i=1}^n\mathbb{E}_s\big[(\langle\sigma_1\sigma_i\rangle_s-\langle\sigma_1\rangle_s\langle\sigma_i\rangle_s)^2\big]\bigg)^{1/2} \tag{44}$$

Proof. Using the identity

$$Q_{2p}^k - \langle Q_{2p}\rangle_s^k = (Q_{2p} - \langle Q_{2p}\rangle_s)\sum_{l=0}^{k-1} Q_{2p}^{k-l-1}\langle Q_{2p}\rangle_s^l \tag{45}$$

and (33) we get

$$\begin{aligned}
|P(Q_{2p}) - \langle P(Q_{2p}\rangle_s| &= |Q_{2p} - \langle Q_{2p}\rangle_s||\sum_k P_k \sum_{l=0}^{k-1} Q_{2p}^{k-l-1}\langle Q_{2p}\rangle_s^l| \\
&\leq |Q_{2p} - \langle Q_{2p}\rangle_s|\sum_k kP_k x^{k-1} \\
&\leq P'(x)|Q_{2p} - \langle Q_{2p}\rangle|
\end{aligned}$$

Here x is the bound in (39). Therefore applying the Chebycheff inequality

$$\mathbb{P}_s\left[|P(Q_{2p}) - P(\langle Q_{2p}\rangle_s)| \geq \frac{2p}{n^\delta}\right] \leq \frac{n^{2\delta}}{4p^2}\mathbb{E}_s\left[P'(x)^2\big(\langle Q_{2p}^2\rangle_s - \langle Q_{2p}\rangle_s^2\big)\right] \tag{46}$$

From the definition of the overlap parameters it follows that

$$\begin{aligned}
\langle Q_{2p}^2\rangle_s - \langle Q_{2p}\rangle_s^2 &= \sum_{i,j=1}^n w_i(t_*)w_j(t_*)X_i(t_*)X_j(t_*)\big(\langle\sigma_i\sigma_j\rangle_s^{2p} - \langle\sigma_i\rangle_s^{2p}\langle\sigma_j\rangle_s^{2p}\big) \\
&\leq 2p\sum_{i,j=1}^n x^2 w_i(t_*)w_j(t_*)\big(\langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle\big)
\end{aligned}$$

Substituting in (46) and applying Cauchy-Schwarz to $\sum_{i,j}\mathbb{E}_s[-]$ we get

$$\begin{aligned}
\mathbb{P}_s\big[|P(Q_{2p}) - P(\langle Q_{2p}\rangle_s)| \geq \frac{2p}{n^\delta}\big] \leq & \frac{n^{2\delta}}{2p}\left(\sum_{i,j=1}^n \mathbb{E}_s[x^4 P'(x)^4 w_i(t_*)^2 w_j(t_*)^2]\right)^{1/2} \\
& \times\left(\sum_{i,j=1}^n \mathbb{E}_s[(\langle\sigma_i\sigma_j\rangle_s - \langle\sigma_i\rangle_s\langle\sigma_j\rangle_s)^2]\right)^{1/2}
\end{aligned}$$

From (32), (33) it is easy to see that for any i, j

$$\mathbb{E}_s[x^4 P'(x)^4 w_i(t_*)^2 w_j(t_*)^2] \leq \frac{C_3^2}{n^4}$$

where C_3 is independent of n. It follows that

$$\mathbb{P}_s\left[|P(Q_{2p}) - P(\langle Q_{2p}\rangle_s)| \geq \frac{2p}{n^\delta}\right]$$

$$\leq \frac{n^{2\delta-1}}{2p} C_3 \left(\sum_{i,j=1}^{n} \mathbb{E}_s[(\langle\sigma_i\sigma_j\rangle_s - \langle\sigma_i\rangle_s\langle\sigma_j\rangle_s)^2]\right)^{1/2}$$

$$= \frac{n^{2\delta-\frac{1}{2}}}{2p} C_3 \left(\sum_{i=1}^{n} \mathbb{E}_s[(\langle\sigma_i\sigma_1\rangle_s - \langle\sigma_i\rangle_s\langle\sigma_1\rangle_s)^2]\right)^{1/2} \tag{47}$$

In the last equality we have used the symmetry of the ensemble with respect to variable node permutations. □

Denote by $h_{n,\gamma}(t_*, s)$ the entropy of the $\mu_{t_*,s}$ interpolating measure. Note that this should not be confused with the multi-Poisson ensemble entropy $h_{n,\gamma}$ (which corresponds to $t_* = t_{\max}$ and $s = \gamma$).

Lemma 2. *For the BEC and BIAWGNC with any noise value and for general BMS(ϵ) channels satisfying H we have*

$$\sum_{i=1}^{n} \mathbb{E}_s[(\langle\sigma_1\sigma_i\rangle_s - \langle\sigma_1\rangle_s\langle\sigma_i\rangle_s)^2] \leq F(\epsilon) + G(\epsilon)\frac{d^2}{d\epsilon^2}\mathbb{E}_s[h_{n,\gamma}(t_*, s)] \tag{48}$$

where $F(\epsilon)$ and $G(\epsilon)$ are two finite constants depending only on the channel parameter.

The proof of lemma 2 is based on the correlation formula of section 1. These are true for any linear code ensemble so they are in particular true for the interpolating (t_*, s) ensemble[7]. For the BEC and BIAWGNC we have already shown the two equalities (9) and (11): thus the inequality (48) is in fact an equality for appropriate values of F and G. The case of general (but highly noisy) BMS channels is presented in appendix C. A converse inequality can also be proven by the methods of appendices B and C.

Proof of proposition 3. Note that for all points of the parameter space (ϵ, s) such that the second derivative of the average conditional entropy is bounded uniformly in n the proof immediately follows from (47), (48) (and the last

[7]in fact one has to check that the addition of $\sum_{a=1}^{e_i} U_a^i$ to l_i does not change the derivation and the final formulas. For this it suffices to follow the calculation of section 3

inequality before that one) by choosing $\delta < \frac{1}{4}$. However, in the large block length limit $n \to +\infty$, genericaly the first derivative of the average conditional entropy has jumps for some threshold values of ϵ (these values depend on the interpolation parameter s). This means that for these threshold values the second derivative cannot be bounded uniformly in n. Since we cannot control these locations we introduce a test function $\psi(\epsilon)$: non negative, infinitely differentiable and with small enough bounded support included in the range of ϵ satisfying H. We consider the averaged quantity

$$\mathcal{Q} = \int d\epsilon \psi(\epsilon) \int_0^\gamma ds \mathbb{P}_s \Big[|P(Q_{2p}) - P(\langle Q_{2p} \rangle)| \geq \frac{2p}{n^\delta} \Big] \tag{49}$$

Writing $\psi(\epsilon) = \sqrt{\psi(\epsilon)}\sqrt{\psi(\epsilon)}$ Cauchy-Schwarz implies

$$\mathcal{Q} \leq \int_0^\gamma ds \Big(\int d\epsilon \psi(\epsilon) \mathbb{P}_s \Big[|P(Q_{2p}) - P(\langle Q_{2p} \rangle)| \geq \frac{2p}{n^\delta} \Big]^2 \Big)^{1/2}$$

Combining this inequality with (47) and (48) we get

$$\begin{aligned}
\mathcal{Q} &\leq \frac{n^{2\delta - \frac{1}{2}}}{2p} C_3 \int_0^\gamma ds \Big(\int d\epsilon \psi(\epsilon) \big(F(\epsilon) + G(\epsilon) \frac{d^2}{d\epsilon^2} \mathbb{E}_s[h_{n,\gamma}(t_*, s)] \big) \Big)^{1/2} \\
&= \frac{n^{2\delta - \frac{1}{2}}}{2p} C_3 \int_0^\gamma ds \Big(\int d\epsilon \psi(\epsilon) F(\epsilon) - \int d\epsilon \frac{d}{d\epsilon} \big(\psi(\epsilon) G(\epsilon) \big) \frac{d}{d\epsilon} \mathbb{E}_s[h_{n,\gamma}(t_*, s)] \Big)^{1/2}
\end{aligned}$$

Note that from the bounds in appendix C $F(\epsilon)$, $G(\epsilon)$ and $G'(\epsilon)$ are integrable except possibly at the edge of the ϵ range defined by H. This is not a problem because we can take the support of $\psi(\epsilon)$ away from such points or alternatively take a $\psi(\epsilon)$ which vanishes sufficiently fast at these points. Moreover the first derivative of the average conditional entropy is bounded uniformly in n and s (see appendix D) by a constant $k(\epsilon)$ that has at most a power singularity at $\epsilon = 0$, and again this is not a problem. Thus by choosing $0 < \delta < \frac{1}{4}$ we obtain

$$\lim_{n \to +\infty} \mathcal{Q} = 0$$

Applying Lebesgue's dominated convergence theorem to convergent subsequences (of the integrand of $\int d\epsilon \psi(\epsilon)$ in (49)) we deduce that

$$\int d\epsilon \psi(\epsilon) \lim_{n_k \to +\infty} \int_0^\gamma ds \mathbb{P}_s[|P(Q_{2p}) - P(\langle Q_{2p} \rangle_s)| \geq \frac{2p}{n_k^\delta}] = 0$$

which implies that along any convergent subsequences, for almost all ϵ

$$\lim_{n_k\to+\infty}\int_0^\gamma ds\mathbb{P}_s\left[|P(Q_{2p})-P(\langle Q_{2p}\rangle_s)|\geq\frac{2p}{n_k^\delta}\right]=0 \tag{50}$$

as long as $\delta \leq \frac{1}{4}$. Now we apply this last statement to two subsequences that attain the lim inf and the lim sup (on the intersection of the two measure one ϵ sets). This proves that the $\lim_{n\to+\infty}$ exists and vanishes. □

6 Conclusion

The main new tool introduced in this paper are relationships between the second derivative of the conditional entropy and correlation functions or mutual information between code bits. This allowed us to estimate the overlap fluctuations in order to get a better handle on the remainder. Some aspects of our analysis bear some similarity with techniques introduced by Talagrand [17] but is independent. One difference is that we use specific symmetry properties of the communications problem.

We expect that the technique developped here can be extended to remove the restriction to high noise (condition H). Indeed the only place in the analysis where we need this restriction is lemma 2. For the BEC and BI-AWGNC the lemma is trivialy satisfied for any noise level (with appropriate constants). Another issue that would be worthwhile investigating is whether the related inequalities of paragraph 1.3 and the converse of lemma 2 can be derived irrespective of the noise level.

The next obvious problem is to prove the converse of the variational bound (theorem 1).

For this one should show that the remainder vanishes when d_V is replaced by the maximizing distribution of $h_{RS}[d_V;\Lambda,P]$. This program has been carried out explicitely in the case of the BEC and the Poisson ensemble [12]. It would be desirable to extend this to more general ensembles and channels but the problem becomes quite hard. However a similar program has been succesfuly carried out for a p-spin model with gauge symmetry[8] (see [10]). A solution of these problems would allow for a rigorous determination of MAP thresholds and would extend our understanding of the intimate relationship between BP and MAP decoding.

[8]In the present context gauge symmetry and channel symmetry are equivalent

A Appendix A

We prove the identities (15), (16) , (17). By definition

$$\langle \sigma_i e^{-\frac{l_i}{2}\sigma_i} \rangle = \frac{1}{Z} \sum_{\underline{\sigma}} \sigma_i \prod_c \frac{1}{2}(1 + \sigma_{\partial c}) \prod_{j \neq i} e^{\frac{l_i}{2}\sigma_i}$$

and

$$\langle e^{-\frac{l_i}{2}\sigma_i} \rangle = \frac{1}{Z} \sum_{\underline{\sigma}} \prod_c \frac{1}{2}(1 + \sigma_{\partial c}) \prod_{j \neq i} e^{\frac{l_i}{2}\sigma_i}$$

Thus

$$\langle \sigma_i \rangle_{\sim i} = \frac{\langle \sigma_i e^{-\frac{l_i}{2}\sigma_i} \rangle}{\langle e^{-\frac{l_i}{2}\sigma_i} \rangle}$$

and plugging the identity

$$e^{-\frac{l_i}{2}\sigma_i} = e^{-\frac{l_i}{2}} \frac{1 - \sigma_i t_i}{1 - t_i}$$

in the brackets immediately leads to (15). For the second and third identities we proceed similarly. Namely,

$$\langle \sigma_i \rangle_{\sim ij} = \frac{\langle \sigma_i e^{-\frac{l_i}{2}\sigma_i} e^{-\frac{l_j}{2}\sigma_j} \rangle}{\langle e^{-\frac{l_i}{2}\sigma_i} e^{-\frac{l_j}{2}\sigma_j} \rangle}$$

and

$$\langle \sigma_i \sigma_j \rangle_{\sim ij} = \frac{\langle \sigma_i \sigma_j e^{-\frac{l_i}{2}\sigma_i} e^{-\frac{l_j}{2}\sigma_j} \rangle}{\langle e^{-\frac{l_i}{2}\sigma_i} e^{-\frac{l_j}{2}\sigma_j} \rangle}$$

Plugging

$$e^{-\frac{l_i}{2}\sigma_i} e^{-\frac{l_j}{2}\sigma_j} = e^{\frac{l_i + l_j}{2}} \frac{1 - \sigma_i t_i - \sigma_j t_j + \sigma_i \sigma_j t_i t_j}{1 - t_i - t_j + t_i t_j}$$

in the brackets, leads immediately to (16) and (17).

B Appendix B

We indicate the main steps of the derivation of the full high noise expansion for

$$\frac{\partial^2}{\partial\epsilon_i\partial\epsilon_j}H_n(\underline{X}\mid\underline{Y})=\delta_{ij}S_1+(1-\delta_{ij})S_2$$

The expansion for S_1 is given by (51) and that for S_2 by (54). They are derived in a form that is suitable to prove lemma 2 of section 5 (see appendix C). For this later proof we need to extract a square correlation at each order as in (54). This is achieved here through the use of appropriate remarquable Nishimori identities, and in order to use these we take the extrinsic forms (21) and (23) of S_1 and S_2.

Let us start with S_1 which is simple. Using the power series expansion of $\ln(1+x)$ we have

$$\ln\left(\frac{1+t_i\langle\sigma_i\rangle_{\sim i}}{1+t_i}\right)=\sum_{p=1}^{+\infty}\frac{(-1)^{p+1}}{p}t_1^p(\langle\sigma_i\rangle_{\sim i}^p-1)$$

This yields an infinite series for S_1 which we will now simplify. Because of the Nishimori identities

$$\mathbb{E}[t_i^{2p-1}]=\mathbb{E}[t_i^{2p}],\qquad \mathbb{E}_{\underline{t}^{\sim i}}[\langle\sigma_i\rangle_{\sim i}^{2p-1}]=\mathbb{E}_{\underline{t}^{\sim i}}[\langle\sigma_{\sim i}\rangle_1^{2p}]$$

we can combine odd and even terms and get

$$S_1=\sum_{p=1}^{+\infty}\frac{m_2^{(2p)}}{2p(2p-1)}\left(\mathbb{E}_{\underline{t}^{\sim i}}[\langle\sigma_i\rangle_{\sim i}^{2p}]-1\right)\tag{51}$$

This series is absolutely convergent as long as

$$\sum_{p=1}^{+\infty}\frac{m_2^{(2p)}}{2p(2p-1)}<+\infty$$

which is true for channels satisfying H.

In the rest of the appendix we deal with S_2 which is considerably more complicated. However the general idea is the same as above. First we use the expansion of $\ln(1+x)$ to get

$$\ln\left(\frac{1+\langle\sigma_i\rangle_{\sim ij}t_i+\langle\sigma_j\rangle_{\sim ij}t_j+\langle\sigma_i\sigma_j\rangle_{\sim ij}t_it_j}{1+\langle\sigma_i\rangle_{\sim ij}t_i+\langle\sigma_j\rangle_{\sim ij}t_j+\langle\sigma_i\rangle_{\sim ij}\langle\sigma_j\rangle_{\sim ij}t_it_j}\right)=\mathrm{I}-\mathrm{II}-\mathrm{III}\tag{52}$$

where

$$\mathrm{I} = \sum_{p=1}^{\infty} \frac{(-1)^{p+1}}{p} \Big(\langle \sigma_i \rangle_{\sim ij} t_i + \langle \sigma_j \rangle_{\sim ij} t_j + \langle \sigma_i \sigma_j \rangle_{\sim ij} t_i t_j \Big)^p$$

$$\mathrm{II} = \sum_{p=1}^{\infty} \frac{(-1)^{p+1}}{p} t_i^p \langle \sigma_i \rangle_{\sim ij}^p, \qquad \mathrm{III} = \sum_{p=1}^{\infty} \frac{(-1)^{p+1}}{p} t_j^p \langle \sigma_j \rangle_{\sim ij}^p$$

We expand the multinomial in I

$$\sum_{\substack{k_a, k_b, k_c \\ k_a + k_b + k_c = p}} \frac{p!}{k_a! k_b! k_c!} t_i^{k_a + k_c} t_j^{k_b + k_c} \langle \sigma_i \rangle_{\sim ij}^{k_a} \langle \sigma_j \rangle_{\sim ij}^{k_b} \langle \sigma_i \sigma_j \rangle_{\sim ij}^{k_c}$$

and subtract the terms II and III. Then only terms that have powers of the form $t_i^k t_j^l$ with $k, l \geq 1$ will survive in (52). Moreover because of the identities $\mathbb{E}[t_i^{2k-1}] = \mathbb{E}[t_i^{2k}]$ and $\mathbb{E}[t_j^{2l-1}] = \mathbb{E}[t_j^{2l}]$ we find for S_2

$$\begin{aligned} S_2 = & \sum_{k \geq l \geq 1}^{+\infty} m_1^{(2k)} m_1^{(2l)} \big(T_{00} + T_{01} + T_{10} + T_{11} \big) \\ & + \sum_{l > k \geq 1}^{+\infty} m_1^{(2k)} m_1^{(2l)} \big(T'_{00} + T'_{01} + T'_{10} + T'_{11} \big) \end{aligned} \tag{53}$$

with (we abuse notation by not indicating the (kl) and (ij) dependence in the T and T' factors)

$$\begin{aligned} T_{\kappa\lambda} = & \sum_{p = 2k - \kappa}^{2k - \kappa + 2l - \lambda} \frac{(-1)^{p+1}}{p} \frac{p!}{(p - (2l - \lambda))!(p - (2k - \kappa))!(2k - \kappa + 2l - \lambda - p)!} \\ & \times \mathbb{E}_{\underline{t}^{\sim ij}} \Big[\langle \sigma_i \rangle_{\sim ij}^{p - (2l - \lambda)} \langle \sigma_j \rangle_{\sim ij}^{p - (2k - \kappa)} \langle \sigma_i \sigma_j \rangle_{\sim ij}^{2k - \kappa + 2l - \lambda - p} \Big] \end{aligned}$$

and

$$T'_{\kappa\lambda} = \text{exchange k, l and } \kappa, \lambda \text{ and i, j}$$

The next simplification step occurs by using the Nishimori identity for the expectation in the above formula

$$\mathbb{E}_{\underline{t}^{\sim ij}} \Big[\langle \sigma_i \rangle_{\sim ij}^{m_1} \langle \sigma_j \rangle_{\sim ij}^{m_2} \langle \sigma_i \sigma_j \rangle_{\sim ij}^{m_3} \Big] = \mathbb{E}_{\underline{t}^{\sim ij}} \Big[\langle \sigma_i^{m_1} \sigma_j^{m_2} (\sigma_i \sigma_j)^{m_3} \rangle_{\sim ij} \langle \sigma_i \rangle_{\sim}^{m_1} \langle \sigma_j \rangle_{\sim ij}^{m_2} \langle \sigma_i \sigma_j \rangle_{\sim ij}^{m_3} \Big]$$

and using $\sigma_i \in \{\pm 1\}$, to "linearize" the terms $(\sigma_i\sigma_j)^{m_1}\sigma_i^{m_2}\sigma_j^{m_3}$. Tedious but straighforward algebra then yields

$$\sum_{\kappa,\lambda} T_{\kappa,\lambda} = \sum_{p=2k-1}^{2k+2l-1} \frac{(-1)^{p+1}}{p(p+1)} \frac{(p+1)!}{(p+1-2k)!(p+1-2l)!(2k+2l-p-1)!} \times \mathbb{E}_{\underline{t}\sim ij}\left[\langle\sigma_i\rangle_{\sim}^{p-2l+1}\langle\sigma_j\rangle_{\sim ij}^{p-2k+1}(\langle\sigma_i\sigma_j\rangle_{\sim ij}^{2k+2l-1-p}\right]$$

A similar formula obtained by exchanging k, l and i, j holds for $\sum_{\kappa\lambda} T'_{\kappa\lambda}$. Replacing these sums in (53) yields a high noise expansion for S_2.

However this is not yet pratical for us because we need to extract a general square correlation factor $\big(\langle\sigma_i\sigma_j\rangle_{\sim ij} - \langle\sigma_i\rangle_{\sim ij}\langle\sigma_j\rangle_{\sim ij}\big)^2$. The fact that this is possible is a "miracle" that comes out of the Nishimori identities that were used. Setting

$$X = \langle\sigma_i\rangle_{\sim ij}\langle\sigma_j\rangle_{\sim ij}, \qquad Y = \langle\sigma_i\sigma_j\rangle_{\sim ij}$$

and using the change of variables $m = p - 2k + 1$ the last expression becomes $(k \geq l)$

$$\mathbb{E}_{\underline{t}\sim ij}\left[\frac{\langle\sigma_i\rangle_{\sim ij}^{2k-2l}}{(2l)!}\sum_{m=0}^{2l}(-1)^m\binom{2l}{m}X^mY^{2l-m}(m+2k-2)\cdots(m+2k-(2l-1))\right]$$

One can check that this is equal to

$$\frac{\langle\sigma_i\rangle_{\sim ij}^{2k-2l}}{(2l)!}X^{2l-2k}\frac{\partial^{2l-2}}{\partial X^{2l-2}}\Big(X^{2k-2}(X-Y)^{2l}\Big)$$

The latter can be checked by first expanding $(X-Y)^{2l}$ and then differentiating. On the other hand one can use the Leibnitz rule

$$\frac{\partial^{2l-2}}{\partial X^{2l-2}}\Big(X^{2k-2}(X-Y)^{2l}\Big) = \sum_{r=0}^{2l-2}\binom{2l-2}{r}\frac{\partial^r}{\partial X^r}X^{2k-2}\frac{\partial^{2l-2-r}}{\partial X^{2k-2-r}}(X-Y)^{2l}$$

to find that the last expectation above is equal to

$$\mathbb{E}_{\underline{t}\sim ij}\left[(X-Y)^2\langle\sigma_i\rangle_{\sim ij}^{2k-2l}\sum_{r=0}^{2l-2}A_{rlk}X^r(X-Y)^{2l-r-2}\right]$$

where

$$A_{rlk} = \frac{1}{(2l)!}\binom{2l-2}{r}[2l]_r[2k-2]_{2l-2-r}, \qquad [m]_r = m(m-1)\cdots(m-r+1)$$

We define $A_{011} = \frac{1}{2}$. We proceed similarly for the terms with $k < l$. Finaly one finds

$$\begin{aligned} S_2 = \sum_{k\geq l\geq 1} & m_1^{(2k)} m_1^{(2l)} \mathbb{E}_{\underline{t}\sim ij}\Bigg[\Big(\langle\sigma_i\sigma_j\rangle_{\sim ij} - \langle\sigma_i\rangle_{\sim ij}\langle\sigma_i\rangle_{\sim ij}\Big)^2 \langle\sigma_i\rangle_{\sim ij}^{2k-2l} \\ & \times \sum_{r=0}^{2l-2} A_{rlk}\langle\sigma_i\rangle_{\sim ij}^r\langle\sigma_j\rangle_{\sim ij}^r\Big(\langle\sigma_i\rangle_{\sim ij}\langle\sigma_j\rangle_{\sim ij} - \langle\sigma_i\sigma_j\rangle_{\sim ij}\Big)^{2l-2-r}\Bigg] \\ & + \sum_{l>k\geq 1} \text{idem with k, l and i, j exchanged} \end{aligned} \tag{54}$$

Let us now briefly justify that the series is absolutely convergent for channels satisfying H. We Note the following facts: $A_{rlk} \leq \binom{2l-2}{r}2^{2k-3}$ and $2^{2k-2}3^{2l-2} \leq (\frac{5}{2})^{2k+2l-4}$ for $k \geq l$ together with the version with k, l exchanged. It easily follows that

$$|S_2| \leq \frac{8}{625}\mathbb{E}_{\underline{t}\sim ij}\Bigg[\Big(\langle\sigma_i\sigma_j\rangle_{\sim ij} - \langle\sigma_i\rangle_{\sim ij}\langle\sigma_i\rangle_{\sim ij}\Big)^2\Bigg] \sum_{k,l\geq 1}\Big(\frac{5}{2}\Big)^{2k+2l}|m_1^{(2k)}m_1^{(2l)}| \tag{55}$$

Thus the series for S_2 is absolutely convergent as long as

$$\sum_{p=1}^{+\infty}\Big(\frac{5}{2}\Big)^{2p}|m_1^{(2p)}| < +\infty$$

Note that we have not attempted to optimize the above estimates.

C Appendix C

We prove lemma 2 for highly noisy general BMS channels. For this we use the high noise expansion derived in appendix B. There it was derived for a general linear code ensemble, and this is also the framework of the proof below. Of course the result applies to the interpolating ensemble of lemma

2. Note that the the final constants $F(\epsilon)$ and $G(\epsilon)$ do not depend on the code ensemble but only on the channel.

Consider equation (8) for $\frac{d^2}{d\epsilon^2}\mathbb{E}_{\mathcal{C},\underline{t}}[h_n]$. By the same estimates than those for S_1 in appendix B, the first term on the right hand side is certainly greater than

$$-\sum_{p=1}^{+\infty}\frac{|m_2^{(2p)}|}{2p(2p-1)}=-A$$

To get a lower bound for the second term we consider the series expansion given by that for S_2 in (54). In that series we keep the first term corresponding to $k=l=1$, namely

$$\frac{1}{2}(m_1^{(2)})^2\sum_{j\neq 1}\mathbb{E}_{\mathcal{C},\underline{t}^{\sim 1j}}\Big[\Big(\langle\sigma_1\sigma_j\rangle_{\sim 1j}-\langle\sigma_1\rangle_{\sim 1j}\langle\sigma_j\rangle_{\sim 1j}\Big)^2\Big]=B$$

and lower bound the rest of the series $(k,l)\neq(1,1)$ by using estimates of appendix B. More precisely this part is lower bounded by

$$-\left(\frac{8}{625}\Big(\sum_{p=1}^{+\infty}\big(\frac{5}{2}\big)^{2p}|m_1^{(2p)}|\Big)^2-\frac{1}{2}(m_1^{(2)})^2\right)$$
$$\times\sum_{j\neq 1}\mathbb{E}_{\mathcal{C},\underline{t}^{\sim 1j}}\Big[\Big(\langle\sigma_1\sigma_j\rangle_{\sim 1j}-\langle\sigma_1\rangle_{\sim 1j}\langle\sigma_j\rangle_{\sim 1j}\Big)^2\Big]=-C$$

Putting these three estimates together we get

$$\frac{d^2}{d\epsilon^2}\mathbb{E}_{\mathcal{C},\underline{t}}[h_n]\geq -A+B-C \tag{56}$$

As long as the noise level is high enough so that (see H)

$$\sum_{p=2}^{+\infty}\big(\frac{5}{2}\big)^{2p}|m_1^{(2p)}|<(\sqrt{2}-1)\big(\frac{5}{2}\big)^2|m_1^{(2)}|$$

the inequality (56) implies

$$\sum_{j\neq 1}\mathbb{E}_{\mathcal{C},\underline{t}^{\sim 1j}}\Big[\Big(\langle\sigma_1\sigma_j\rangle_{\sim 1j}-\langle\sigma_1\rangle_{\sim 1j}\langle\sigma_j\rangle_{\sim 1j}\Big)^2\Big]\leq\tilde{F}(\epsilon)+\tilde{G}(\epsilon)\frac{d^2}{d\epsilon^2}\mathbb{E}_{\mathcal{C},\underline{t}}[h_n] \tag{57}$$

for two noise dependent positive finite constants $\tilde{F}(\epsilon)$, $\tilde{G}(\epsilon)$.

The final step of the proof consists in passing from the extrinsic average $\langle - \rangle_{\sim 1j}$ in the correlation to the ordinary one $\langle - \rangle_{1j}$. This is achieved as follows. From the formulas (16) and (17) we deduce that

$$\langle \sigma_j \sigma_i \rangle - \langle \sigma_j \rangle \langle \sigma_i \rangle = \big(\langle \sigma_j \sigma_i \rangle_{\sim ij} - \langle \sigma_j \rangle_{\sim ij} \langle \sigma_i \rangle_{\sim ij} \big) R_{ij}$$

with

$$R_{ij} = \frac{\big(1 - \langle \sigma_i \rangle t_i - \langle \sigma_j \rangle t_j + \langle \sigma_i \sigma_j \rangle t_i t_j\big)^2}{(1-t_i^2)(1-t_j^2)} \leq \frac{4}{(1-t_i^2)(1-t_j^2)}$$

a function that depends on all log-likelihood variables.

Thus we have

$$\begin{aligned}
(\langle \sigma_j \sigma_i \rangle - \langle \sigma_j \rangle \langle \sigma_i \rangle)^2 &= \big(\langle \sigma_j \sigma_i \rangle_{\sim ij} - \langle \sigma_j \rangle_{\sim ij} \langle \sigma_i \rangle_{\sim ij} \big)^2 R_{ij}^2 \\
&\leq \big(\langle \sigma_j \sigma_i \rangle_{\sim ij} - \langle \sigma_j \rangle_{\sim ij} \langle \sigma_i \rangle_{\sim ij} \big)^2 \frac{16}{(1-t_i^2)^2(1-t_j^2)^2}
\end{aligned}$$

Taking now the expectation $\mathbb{E}_{\mathcal{C},\underline{t}}$ we get

$$\begin{aligned}
\mathbb{E}_{\mathcal{C},\underline{t}}\left[\Big(\langle \sigma_j \sigma_i \rangle - \langle \sigma_j \rangle \langle \sigma_i \rangle\Big)^2\right] &\leq \mathbb{E}_{\mathcal{C},\underline{t}^{\sim ij}}\left[\Big(\langle \sigma_j \sigma_i \rangle_{\sim ij} - \langle \sigma_j \rangle_{\sim ij} \langle \sigma_i \rangle_{\sim ij} \Big)^2\right] \\
&\quad \times \mathbb{E}_{t_i,t_j}\left[\frac{16}{(1-t_i^2)^2(1-t_j^2)^2}\right]
\end{aligned}$$

Since t_i, t_j are independent we get

$$\begin{aligned}
\mathbb{E}_{t_i,t_j}\left[\frac{16}{(1-t_i^2)^2(1-t_j^2)^2}\right] &= 16\bigg(\mathbb{E}\left[\frac{1}{(1-t^2)^2}\right]\bigg)^2 = 16\bigg(\mathbb{E}\Big[\sum_{p\geq 0}(p+1)t^{2p}\Big]\bigg)^2 \\
&= 16\bigg(\Big[\sum_{p\geq 0}(p+1)m_0^{(2p)}\Big]\bigg)^2 \qquad (58)
\end{aligned}$$

which converges for highly noisy channels satisfying H. The result of the lemma follows by combining (57) and (58). The constants $F(\epsilon)$ and $G(\epsilon)$ are equal to $\tilde{F}(\epsilon)$ and $\tilde{G}(\epsilon)$ divided by the expression on the right hand side of the last inequality.

D Appendix D

We prove the boundedness and positivity of $\frac{d}{d\epsilon}\mathbb{E}_s[h_{n,\gamma}(t_*, s)]$ which is needed in the proof of lemma 2.

Lemma 3. *For the BEC and BIAWGNC with any noise level, and any BMS satisfying H, there exists a constant $k(\epsilon)$ independent of n, γ, t_* and s such that*

$$0 \leq \frac{d}{d\epsilon}\mathbb{E}_s[h_{n,\gamma}(t_*, s)] \leq k(\epsilon) \tag{59}$$

For the BEC we can take $k(\epsilon) = \frac{\ln 2}{\epsilon}$ and for the BIAWGNC $k(\epsilon) = \frac{2}{\epsilon^{-3}}$. For general BMS channels satisfying H the constant remains bounded as a function of ϵ (i.e. in the high noise regime).

Here we have stated the lemma for the multi-Poisson interpolating ensemble which is our specific need. However as the proof below shows it is independent of the specific code ensemble and the bound depends only on the channel.

Proof. We will use the GEXIT formula of lemma 1. Since the proposition applies for any linear code it also applies for the interpolating ensemble of interest here. In the case of the BEC and BIAWGNC we have (see (5), (7)

$$\frac{d}{d\epsilon}\mathbb{E}_s[h_{n,\gamma}(t_*, s)] = \frac{\ln 2}{\epsilon}(1 - \mathbb{E}_s[\langle\sigma_1\rangle_s]$$

and

$$\frac{d}{d\epsilon}\mathbb{E}_s[h_{n,\gamma}(t_*, s)] = \frac{2}{\epsilon^3}(1 - \mathbb{E}_s[\langle\sigma_1\rangle_s]$$

The bounds of the lemma follow immediately since $-1 \leq \sigma_1 \leq 1$.

For highly noisy BMS channels we proceed by expansions. For this reason we have to use the "extrinsic form" of the GEXIT formula (analogous to (21))

$$\frac{d}{d\epsilon}\mathbb{E}_s[h_{n,\gamma}(t_*, s)] = \int_{-1}^{+1} dt_1 \frac{\partial c_D(t_1)}{\partial\epsilon}\mathbb{E}_{s,\sim t_1}\left[\ln\left(\frac{1 + t_1\langle\sigma_1\rangle_{s,\sim 1}}{1 + t_1}\right)\right]$$

Expanding the logarithm and using Nishimori identities (as in the expansion of S_1 in appendix B we obtain

$$\frac{d}{d\epsilon}\mathbb{E}_s[h_{n,\gamma}(t_*, s)] = \sum_{p=1}^{\infty}\frac{m_1^{(2p)}}{2p(2p-1)}\mathbb{E}_{s,\sim 1}[\langle\sigma_1\rangle_{s,\sim 1}^{2p} - 1]$$

The positivity follows from $m_1^{(2p)} \leq 0$ [1] and $-1 \leq \sigma_1 \leq 1$. The upper bound (and absolute convergence) follow from condition H. In particular we get

$$k(\epsilon) = 2 \sum_{k=1}^{+\infty} \frac{|m_1^{(2p)}|}{2p(2p-1)}$$

which is independent of n, γ, t_* and s. $\square$

Acknowledgment

The work of Shrinivas Kudekar has been supported by a grant of the Swiss National Foundation 200021-105604. The authors acknowledge various discussions with S. Korada, O. Leveque and R. Urbanke.

References

[1] T. Richardson, R. Urbanke; "Modern Coding Theory" *Cambridge University Press (2008).*

[2] Y. Kabashima, T. Murayama, D. Saad; "Typical performance of Gallager-Type Error-Correcting Codes" *Phys. Rev. Lett. vol 84, (2000) 1355 - 1358*

[3] A. Montanari; "The glassy phase of Gallager codes" *Eur. Phys. J. B. vol 23, (2001) 121 - 136*

[4] C. Measson, A. Montanari, R. Urbanke; "Maxwell's construction: the hidden bridge between maximum-likelihood and iterative decoding" *Int. Symp. Inf. Theory, Chicago (2004) 225*; see also preprint arXiv:cs/0506083.

[5] M. Talagrand; "The Parisi formula" *Ann. of Math. vol 163, no 1 (2006), 221 - 263.*

[6] M. Mezard, G. Parisi, M. A. Virasoro; "Spin glass theory and beyond" *World Scientific (1987)*

[7] A. Montanari, "Tight Bounds for LDPC and LDGM Codes Under MAP Decoding" *IEEE Trans. Inf. Theory, vol 5, no. 9 (2005) 3221–3246.*

[8] S. Franz, M. Leone; "Replica bounds for optimization and diluted spin glasses" *J. Stat. Phys. vol 111 (2003) 535 - 564*

[9] D. Pachenko, M. Talagrand; "Bounds for diluted mean-fields spin glass models" *Prob. Theory. Relat. Fields , vol 130, (2004) 319-336.*

[10] S. Korada, N. Macris; "Exact solution of a p-spin model and its relationship to error correcting codes" *Int. Symp. Inf. Theory, Seattle (2006) 2264-2268*

[11] C. Measson, A. Montanari, R. Urbanke; "Asymptotic rate versus design rate" *Int. Symp. Inf. Theory, Nice (2007) 1541-1545*

[12] S. Korada, S. Kudekar, N. Macris; "Exact solution for the conditional entropy of Poissonian LDPC codes over the binary erasure channel" *Int. Symp. Inf. Theory, Nice (2007) 1016-1020*

[13] S. Kudekar, N. Macris; "Decay of correlations: An application to low density parity check codes" *5th Int Symp. on Turbo Codes and Related Topics, Lausanne (2008)*

[14] S. Kudekar, N. Macris; "Proof of replica formulas in the high noise regime for communication using LDGM codes" *Inf. Theory Workshop, Porto (2007)*

[15] F. Guerra, F. Toninelli; "Quadratic Replica Coupling in the Sherrington-Kirkpatrick Mean Field Spin Glass Model" *J. Math. Phys. vol 43 (2002) 3704 - 3716.*

[16] F. Guerra; "Replica broken bounds in the mean field spin glass model" *Commun. Math. Phys. vol 233 (2003) 1 - 12*

[17] M. Talagrand; "Spin glasses: a Challenge for Mathematicians" *Spinger-Verlag (2003)*

[18] N. Macris; "Griffiths-Kelly-Sherman correlation inequalities: a useful tool in the theory of LDPC codes" *IEEE Trans. Inf. Theory, vol 53 no 2 (2007) 664 - 683.*

[19] N. Macris; "Sharp bounds on generalized EXIT functions" *IEEE Trans. Inf. Theory, vol 53 no 7 (2007) 2365 - 2375.*

[20] S. Kudekar, N. Macris; "Sharp bounds for MAP decoding of general irregular LDPC codes" *Int. Symp. Inf. Theory, Seattle (2006) 2259 - 2263*

[21] H. Nishimori; "Statistical Physics of Spin Glasses and Information Processing: An Introduction" *Oxford University Press (2001)*