Sharp bounds on generalized EXIT functions

Nicolas Macris

LTHC - IC - EPFL
CH-1015 Lausanne, Switzerland
email: nicolas.macris@epfl.ch

Abstract
We consider communication over binary input memoryless symmetric channels with low density parity check codes. The relationship between maximum a posteriori and belief propagation decoding is investigated using a set of correlation inequalities that first appeared in statistical mechanics of gaussian spin glasses. We prove bounds on generalized EXIT functions, that are believed to be tight, and discuss their relationship with the ones obtained by the interpolation method.

Keywords. Low-density parity-check codes, belief propagation, EXIT curve, spin glasses, correlation inequalities.

1 Introduction

We consider communication with binary linear codes across a family of binary-input memoryless output-symmetric (BMS) channels that are ordered by physical degradation. This family is described by a transition probability $p^{\epsilon}_{Y|X}(y|x)$ depending on one parameter $\epsilon \geq 0$, where $x \in \{0,1\}$ and y belongs to the output alphabet $\mathcal{Y}$. We will think of ϵ as a noise level with $\epsilon = 0$ meaning zero noise. Ordering by physical degradation means that if $\epsilon > \epsilon'$ there exists a symmetric channel $q_{Y|Y'}$ such that

$$p^{\epsilon}_{Y|X}(y|x) = \sum_{y' \in \mathcal{Y}} q_{Y|Y'}(y|y') p^{\epsilon'}_{Y'|X}(y'|x) \tag{1.1}$$

We assume throughout that $p^{\epsilon}_{Y|X}(x|y)$ is differentiable with respect to ϵ. The superscript ϵ will be dropped for ease of notation except when necessary.

Suppose that we choose a codeword uniformly at random from a binary linear code of block length n, and that we observe the output $Y_1, ..., Y_n = Y^n$. The *Extrinsic Information Transfer* (EXIT) curve[1] is defined as

$$h(\epsilon) = \frac{1}{n}\sum_{i=1}^{n} H(X_i|Y^n \backslash Y_i)) \tag{1.2}$$

Since the channel is symmetric and the prior distribution on the codewords is uniform, the result does not depend on the input word and there is no loss in generality to suppose that the input is the all zero codeword. From now on we stick to this convention. When communication takes place over the binary erasure channel (BEC), with $0 < \epsilon < 1$, it was shown [3] that the area under the curve (1.2) is equal to the rate of the binary linear code: $\int_0^1 d\epsilon h(\epsilon) = r$. This has been exploited to give bounds on the MAP thresholds, ϵ_{MAP} of LDPC code ensembles [5]. Since for the BEC (1.2) is proportional to the bit error probability under MAP decoding it is always smaller than the corresponding quantity calculated with a belief propagation decoder: $h(\epsilon) \leq h_{BP}(\epsilon)$. Thus by looking at the area under the iterative curve and matching it to the code rate one can compute an upper bound $\bar{\epsilon}_{MAP}$ to the MAP threshold (i.e $\epsilon_{BP} < \epsilon_{MAP} < \bar{\epsilon}_{MAP}$). In fact much more is true: $\bar{\epsilon}_{MAP}$ agrees with the results of replica calculations [4]. Replica calculations are expected to be exact in this context so one should have that $\epsilon_{MAP} = \bar{\epsilon}_{MAP}$ and also that above ϵ_{MAP} the two EXIT curves, associated to MAP and BP decoding, should be equal. This equality has been proven recently, at least for some LDPC ensembles satisfying a special condition, in the work of Measson, Montanari, and Urbanke [7]. These authors show that this equality has a very nice connection with Maxwell's construction of first order phase transitions.

The picture that has emerged for the BEC channel should be valid for general channels. One reason to believe this is that the replica calculations, although mathematically uncontrolled, are the same on any channel and absence of replica symmetry breaking can be argued to be correct at least on general symmetric channels. This point of view is adopted in [6], [8] and is the motivation to introduce *Generalized* EXIT curves that satisfy an area theorem by construction. Although the usual MAP and BP EXIT curves are related by a simple inequality for general channels (this follows from the data processing inequality), there is no area theorem like in the case of the

[1] The present definition differs from the original one in [1]. Here we follow [2].

BEC. One would like to have *Generalized* EXIT functions that satisfy at the same time the area theorem and a simple inequality. This may then give an operational way to compute upper bounds on the MAP threshold. One may even be more optimistic and look for functions that are equal above the MAP threshold, so that one would have an operational way to compute this treshold exactly.

The *Generalized* EXIT curve associated to MAP decoding (MAP GEXIT) is defined so that it satisfies an "area theorem" by construction. It is given by the derivative of the conditional Shannon entropy with respect to the noise parameter, i.e

$$g_{\mathrm{MAP}}^{(n)}(\epsilon) = \frac{1}{n}\frac{d}{d\epsilon}H(X^n|Y^n) \tag{1.3}$$

There are at least three motivations for making such a definition. First, for the special case of the BEC channel this reduces to (1.2). Second this quantity satisfies an "area theorem" by construction. Third it is closely related to quantities of statistical mechanics (such as the free energy) for which the replica method is expected to be exact.

Let us give an alternative expression for (1.3) which will be useful to motivate the definition of the *Belief Propagation* GEXIT curve. Given that the all zero codeword is the input fed into the channel the observations are described by i.i.d random variables $Y_1, ..., Y_n$ whose common distribution is $p_{Y|X}(y|0)$. The distribution of the log-likelihood ratio $l_i = \ln\frac{p(y_i|0)}{p(y_i|1)}$ will be denoted by $c(l_i)$. The later depends on ϵ and is differentiable. We introduce the notation $Z^{n\backslash i}$ for $(Z_1, ..., Z_n)\backslash Z_i$. Consider the marginals of the *extrinsic* a posteriori distribution i.e $p_{X_i|Y^{n\backslash i}}(x_i|y^{n\backslash i})$, or equivalently the associated *extrinsic* log-likelihood ratios

$$L_i = \ln\frac{p_{X_i|Y^{n\backslash i}}(0|y^{n\backslash i})}{p_{X_i|Y^{n\backslash i}}(1|y^{n\backslash i})} \tag{1.4}$$

One has

$$g_{MAP}^{(n)}(\epsilon) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{l^{n\backslash i}}\left[\int_{-\infty}^{+\infty} dl_i\frac{dc(l_i)}{d\epsilon}\ln(1+e^{-l_i-L_i})\right] \tag{1.5}$$

The *BeliefPropagation* GEXIT curve is defined by a similar expression where instead of a MAP decoder we take a belief propagation decoder. In other words for each bit i one computes the extrinsic (that is setting $l_i = 0$) soft bit estimate with d iterations of the belief propagation decoder, call it Δ_i^d, and the associated likelihood variable, call it $L_i^{(d)}$. These are related through $\Delta_i^d = \tanh\frac{L_i^{(d)}}{2}$. The BP GEXIT curve is defined as

$$g_{BP}^{(n,d)}(\epsilon) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{l^n\setminus i}\left[\int_{-\infty}^{+\infty} dl_i \frac{dc(l_i)}{d\epsilon}\ln(1+e^{-l_i-\Lambda_i^{(d)}})\right] \tag{1.6}$$

In the next section we give a selfcontained derivation of (1.5) and also show several alternative representations for (1.5) and (1.6).

If the code is chosen uniformly at random from an LDPC ensemble one can prove concentration of (1.5) and (1.6) on their expectation value over the code ensemble [7]. This then gives an efficient way to compute the BP-GEXIT curve. Let $a_{DE}^{(d)}(\Lambda)$ the density of $\Lambda_i^{(d)}$ computed by the method of *density evolution.* As $n \to +\infty$, (1.6) concentrates on

$$\int_{-\infty}^{+\infty} d\Lambda a_{DE}^{(d)}(\Lambda)\int_{-\infty}^{+\infty} dl \frac{dc(l)}{d\epsilon}\ln(1+e^{-l-\Lambda}) \tag{1.7}$$

The main theme of this paper is to use *correlation inequalities* from statistical mechanics in order to prove:

Theorem 1.1. *Consider communication over a family of $BMS(\epsilon)$ channels that are ordered by physical degradation, using a* LDPC(λ,ρ,n) *code ensemble with bounded node degrees. For any sequence of* LDPC(λ,ρ,n) *codes of increasing block lengths,*

$$\limsup_{n\to+\infty}\mathbb{E}_{\mathcal{C}}[g_{MAP}^{(n)}(\epsilon)] \leq \lim_{d\to\infty}\int_{-\infty}^{+\infty} d\Lambda a_{DE}^{(d)}(\Lambda)\int_{-\infty}^{+\infty} dl \frac{dc(l)}{d\epsilon}\ln(1+e^{-l-\Lambda}) \tag{1.8}$$

where the limit on the right hand side exists.

Such bounds have already been used in [6], [8] and a detailled proof using other methods has been presented recently [9]. The correlation inequalities that we employ here are a slight extension of a class of Griffiths-Kelly-Sherman (GKS) like inequalities for gaussian spin glasses on their Nishimori line [11]. In the special case of the BEC it is sufficient to use the standard GKS [10] inequalities discovered in the framework of ferromagnetic spin systems. For a BIAWGNC one can directly use the inequalities in the form of [11]. Correlation inequalities are often a powerful tool of statistical mechanics and it is an interesting fact that they also apply to error correcting codes. They were employed by the author to give bounds on the growth rate of LDPC codes and also to prove the above theorem in the special case of a gaussian channel [15], [16].

The bound (1.8) is closely related to a bound on the conditional entropy itself that has been obtained by the *interpolation method* [12]. This relationship is discussed in section 5.

There are at least two reasons to believe that it is a sharp bound for $\epsilon > \epsilon_{MAP}$ (here we assume for simplicity that the GEXIT curves have only one threshold). For the BEC and under some condition on the LDPC ensemble the equality has been established [8]. Second, if one computes the MAP GEXIT curve by the replica symmetric method one finds the r.h.s, and the replica symmetric method is believed to be exact in the present context[2]

In section 2 we derive various representations for the GEXIT curves from the point of view of the underlying spin system. In section 3 we state the main correlation inequality and prove theorem (1.1). A discussion of the binary erasure and gaussian channels is the object of section 4 and the relationship with interpolation bounds is examined in section 5. Finally in section 6 we conclude with a version of Theorem 1.1 that is valid for each particular instance of a code. We also briefly discuss the close connection between correlation inequalities and the method of physical degradation in section 6. Streamlined proofs of generalized Nishimori identities and correlation inequalities are reported in the appendices.

The main results of this work have been announced in [20].

2 MAP and BP EXIT functions from the spin system perspective

Let us first define general Gibbs measures over a set of Ising spin assignments $\sigma^n = (\sigma_1, ..., \sigma_n) \in \{-1, +1\}^n$. Consider a fixed bipartite factor (or Tanner) graph. There are n variable nodes denoted by latin lower case letters (i,j,...) and m check nodes denoted by upper case letters (A, B,...). We identify a check node A with the subset $A \subset \{1, ..., n\}$ of variable nodes that are connected to A. Thus a factor graph is defined by a certain collection $\mathcal{C}$ of subsets of $\{1, ..., n\}$. The Gibbs measures that will interest us are of the form

$$\mu_{\mathcal{C}}(\sigma^n) = \frac{1}{Z_{\mathcal{C}}} \exp(-H_{\mathcal{C}}(\sigma^n)), \qquad Z_{\mathcal{C}} = \sum_{\sigma^n} \exp(-H_{\mathcal{C}}(\sigma^n)) \tag{2.1}$$

where the hamiltonian (or cost) function is

$$H_{\mathcal{C}}(\sigma^n) = -\sum_{A \in \mathcal{C}} J_A(\sigma_A - 1) - \sum_{i=1}^{n} h_i \sigma_i, \qquad \sigma_A = \prod_{i \in A} \sigma_i \tag{2.2}$$

Here the coefficients J_A and h_i are real numbers. Note that the single spin terms may be associated to additional degree one check nodes but for us it

[2] Because the underlying spin system is dilute and gauge invariant.

is more convenient to separate them out from the other terms. Expectations with respect to the Gibbs measure are denoted by $\langle - \rangle_{\mathcal{C}}$. More precisely for any $X \subset \{1, ..., n\}$,

$$\langle \sigma_X \rangle_{\mathcal{C}} = \sum_{\sigma^n} \sigma_X \mu_{\mathcal{C}}(\sigma^n), \qquad \sigma_X = \prod_{i \in X} \sigma_i \tag{2.3}$$

In the communications problem the *a posteriori distribution* $p_{X^n|Y^n}(x^n|y^n)$ can be viewed as a Gibbs measure of a random spin system. Indeed using Bayes rule for a memoryless channel and assuming a uniform prior over the code words

$$p_{X^n|Y^n}(x^n|y^n) = \frac{1_{\mathcal{C}}(x^n) \prod_{i=1}^n p_{Y|X}(y_i|x_i)}{\sum_{x^n} 1_{\mathcal{C}}(x^n) \prod_{i=1}^n p(y_i|x_i)} \tag{2.4}$$

Now set $\sigma_i = (-1)^{x_i}$, and observe that

$$p_{Y|X}(y|x) = p(y|0) e^{-\frac{l}{2}} e^{\frac{l}{2}\sigma} \tag{2.5}$$

and also that the check node constraints become

$$1_{\mathcal{C}}(x^n) = \prod_{A \in \mathcal{C}} \frac{1}{2}(1 + \sigma_A) = \lim_{\{J_A \to +\infty, A \in \mathcal{C}\}} \prod_{A \in \mathcal{C}} e^{J_A(\sigma_A - 1)} \tag{2.6}$$

Then (2.4) becomes equal to

$$\frac{1}{Z_{\mathcal{C}}} \prod_{A \in \mathcal{C}} \frac{1}{2}(1 + \sigma_A) \prod_{i=1}^n e^{\frac{l_i}{2}\sigma_i}, \qquad Z_{\mathcal{C}} = \sum_{\sigma^n \in \{+1,-1\}^n} \prod_{A \in \mathcal{C}} \frac{1}{2}(1 + \sigma_A) \prod_{i=1}^n e^{\frac{l_i}{2}\sigma_i} \tag{2.7}$$

Obviously this is of the form (2.1) provided:

a) the collection $\mathcal{C}$ is identical with the check node constraints of the code and the associated coefficients $J_A = +\infty$,

b) the coefficients h_i are related to channel outputs by $h_i = \frac{l_i}{2}$.

The Gibbs measure defined by the a posteriori distribution is random in the sense that the channel outputs are random i.i.d with distribution $c(l)$. Furthermore there is also another source of randomness, namely the code which is sampled from an ensemble. The MAP estimate of the i-th bit is

$$\text{sign}[p_{X_i|Y^n}(0|y^n) - p_{X_i|Y^n}(1|y^n)] = \text{sign}\langle \sigma_i \rangle_{\mathcal{C}} \tag{2.8}$$

The soft estimate of the bit is simply $d_i = \langle \sigma_i \rangle_{\mathcal{C}}$, i.e the magnetization at node i. This depends on all observations l^n. Later on we will need the extrinsic soft bit estimate

$$D_i = p_{X_i|Y^{n\setminus i}}(0|y^{n\setminus i}) - p_{X_i|Y^{n\setminus i}}(1|y^{n\setminus i}) = \langle \sigma_i \rangle_{\mathcal{C}, l_i = 0} \tag{2.9}$$

There the Gibbs average is computed for $l_i = 0$.

The Gibbs entropy of the spin system is

$$-\sum_{\sigma^n} \mu_{\mathcal{C}}(\sigma^n) \ln \mu_{\mathcal{C}}(\sigma^n) \tag{2.10}$$

and its average over the channel outputs is nothing else than the Shannon conditional entropy $H(X^n|Y^n)$. Simple algebra shows that

$$H(X^n|Y^n) = \mathbb{E}_{l^n}\left[-\sum_{\sigma^n} \mu_{\mathcal{C}}(\sigma^n) \ln \mu_{\mathcal{C}}(\sigma^n)\right] = \mathbb{E}_{l^n}[\ln Z_{\mathcal{C}}] - \sum_{i=1}^{n} \mathbb{E}_{l^n}\left[\frac{l_i}{2}\langle\sigma_i\rangle_{\mathcal{C}}\right] \tag{2.11}$$

Furthermore channel symmetry implies [17],[4]

$$\mathbb{E}_{l^n}\big[\frac{l_i}{2}\langle\sigma_i\rangle_{\mathcal{C}}\big] = \mathbb{E}_{l_i}\big[\frac{l_i}{2}\big] = \int_{-\infty}^{+\infty} dl c(l)\frac{l}{2} \tag{2.12}$$

Therefore the evaluation of the Shannon conditional entropy reduces to that of the average free energy $\mathbb{E}_{l^n}[\ln Z_{\mathcal{C}}]$ of the corresponding spin system.

2.1 MAP GEXIT curve

Let us first derive identity (1.5) for the MAP-GEXIT function. Differentiating (2.11),

$$\frac{d}{d\epsilon}H(X^n|Y^n) = \sum_{i=1}^{n} \mathbb{E}_{l^{n\setminus i}}\left[\int_{-\infty}^{+\infty} dl_i \frac{dc(l_i)}{d\epsilon}(\ln Z_{\mathcal{C}} - \frac{l_i}{2})\right] \tag{2.13}$$

and using

$$e^{\frac{l_i}{2}\sigma_i} = (1 + \sigma_i \tanh\frac{l_i}{2})\cosh\frac{l_i}{2} = \frac{e^{\frac{l_i}{2}}}{2}\frac{1 + \sigma_i \tanh\frac{l_i}{2}}{1 + \tanh\frac{l_i}{2}} \tag{2.14}$$

we can rewrite the free energy as (for any i)

$$\ln\frac{Z_{\mathcal{C}}}{Z_{\mathcal{C},l_i=0}} = \frac{l_i}{2} - \ln 2 + \ln\frac{1 + \langle\sigma_i\rangle_{\mathcal{C},l_i=0}\tanh\frac{l_i}{2}}{1 + \tanh\frac{l_i}{2}} \tag{2.15}$$

Note that since $\ln Z_{\mathcal{C},l_i=0}$ does not depend on l_i and $\int_{-\infty}^{+\infty} c(l) = 1$,

$$\int_{-\infty}^{+\infty} dl_i \frac{dc(l_i)}{d\epsilon}(\ln Z_{\mathcal{C},l_i=0} - \ln 2) = 0 \tag{2.16}$$

Thus we obtain

$$g_{MAP}^{(n)}(\epsilon) \quad = \quad \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{l^n\setminus i}\Big[\int_{-\infty}^{+\infty} dl_i \frac{dc(l_i)}{d\epsilon}\ln\Big(\frac{1+\langle\sigma_i\rangle_{\mathcal{C},l_i=0}\tanh\frac{l_i}{2}}{1+\tanh\frac{l_i}{2}}\Big)\Big] \tag{2.17}$$

To obtain the expression in terms of the extrinsic log-likelihood ratio we note that $\langle\sigma_i\rangle_{\mathcal{C},l_i=0} = D_i = \tanh\frac{L_i}{2}$ which leads to (1.5).

An alternative form expresses it in terms of the soft bit MAP estimate $d_i = \langle\sigma_i\rangle_{\mathcal{C}}$. Using again (2.14) it is easily seen that

$$\langle\sigma_i\rangle_{\mathcal{C}} = \frac{\langle\sigma_i\rangle_{\mathcal{C},l_i=0}+\tanh\frac{l_i}{2}}{1+\langle\sigma_i\rangle_{\mathcal{C},l_i=0}\tanh\frac{l_i}{2}} \tag{2.18}$$

and inverting this equation leads to

$$g_{MAP}^{(n)}(\epsilon) = -\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{l^n\setminus i}\Big[\int_{-\infty}^{+\infty} dl_i \frac{dc(l_i)}{d\epsilon}\ln\Big(\frac{1-\langle\sigma_i\rangle_{\mathcal{C}}\tanh\frac{l_i}{2}}{1-\tanh\frac{l_i}{2}}\Big)\Big] \tag{2.19}$$

2.2 BP-EXIT curves

Let us begin with a description of the belief propagation decoder in the likelihood domain. Initial messages from variable to check nodes are set to channel observations $l_{i\to C}^{(0)} = l_i$, $i = 1, ..., n$. For $t = 0, ..., d$, messages from check to variable nodes are updated as

$$u_{C\to i}^{(t+1)} = 2\tanh^{-1}\Big(\prod_{j\in C\setminus i}\tanh\frac{l_{j\to C}^{(t)}}{2}\Big) \tag{2.20}$$

and from variable to check nodes

$$l_{i\to A}^{(t+2)} = l_i + \sum_{C\in V(i)\setminus A} u_{C\to i}^{(t+1)} \tag{2.21}$$

Here $V(i)$ is the set of checks connected to variable node i. The soft BP bit estimate after iteration d (d is even) is

$$\delta_i^{(d)} = \tanh\frac{\lambda_i^{(d)}}{2} = \tanh\frac{1}{2}(l_i + \sum_{C\in V(i)} u_{C\to i}^{(d-1)}) \tag{2.22}$$

The extrinsic BP estimate of the i-th bit does not take into account the observation l_i. This means that we define the extrinsic likelihood for the i-th bit at iteration d as,

$$\Lambda_i^{(d)} = \sum_{C\in V(i)} u_{C\to i}^{(d-1)} \tag{2.23}$$

Note that this definition is consistent with the message passing decoder described above because the messages involved in the computational graph of node i do not depend on l_i except for the last iteration. The extrinsic soft bit BP estimate is simply

$$\Delta_i^{(d)} = \tanh \frac{\Lambda_i^{(d)}}{2} \tag{2.24}$$

For a fixed factor graph these quantities have distributions that are induced by $c(l)$. The belief propagation decoder just described can then be used in order to compute the expression (1.6) for the BP-EXIT curve.

For completeness we give here two alternative representations of (1.6) in the *difference domain*. The relationship between δ_i and Δ_i follows by expanding the tanh in (2.22) (and is similar to (2.18))

$$\delta_i = \frac{\Delta_i + \tanh \frac{l_i}{2}}{1 + \Delta_i \tanh \frac{l_i}{2}} \tag{2.25}$$

Then using (2.24) and (2.25) we get the expressions

$$\begin{aligned} g_{BP}^{(n,d)}(\epsilon) &= \sum_{i=1}^{n} \mathbb{E}_{l^{n\setminus i}} \left[\int_{-\infty}^{+\infty} dl_i \frac{dc(l_i)}{d\epsilon} \ln\left(\frac{1 + \Delta_i^{(d)} \tanh \frac{l_i}{2}}{1 + \tanh \frac{l_i}{2}} \right) \right] \\ &= - \sum_{i=1}^{n} \mathbb{E}_{l^{n\setminus i}} \left[\int_{-\infty}^{+\infty} dl_i \frac{dc(l_i)}{d\epsilon} \ln\left(\frac{1 - \delta_i^{(d)} \tanh \frac{l_i}{2}}{1 - \tanh \frac{l_i}{2}} \right) \right] \end{aligned} \tag{2.26}$$

3 Correlation inequalities and their application

Originaly the GKS inequalities where derived for deterministic ferromagnetic spin systems (all $J_A \geq 0$ and $h_i \geq 0$). Remarkably it was shown recently [11] that they can be extended to a class of random spin systems where J_A and h_i are all independent gaussian random variables with equal mean and variance. These inequalities where exploited recently in [15], [16] in the case of a gaussian channel where they directly apply.

Here we adapt, generalize and present streamlined proofs of these inequalities for general output-symmetric channels. The key feature is channel symmetry which implies that for any reasonable function g,

$$\mathbb{E}_{l_i}[g(-l_i)] = \mathbb{E}_{l_i}[g(l_i)e^{-l_i}], \qquad i = 1, ...n \tag{3.1}$$

In the statistical physics literature this is known as *Nishimori*'s condition (translated as a condition on the h_i). When it is satisfied spin averages obey remarkable identities known as *Nishimori identities* [17]. In appendix A we give a simple proof of the following generalized version of these identities. Take any hamiltonian of the form (2.2) with $J_A = +\infty$ and $h_i = \frac{l_i}{2}$ satisfying (3.1) For any collection of subsets $X_1, ..., X_l \subset \{1, ..., n\}$ and integers $m_1, ..., m_l$,

$$\mathbf{E}_{l^n}\left[\langle\sigma_{X_1}\rangle_{\mathcal{C}}^{m_1}...\langle\sigma_{X_l}\rangle_{\mathcal{C}}^{m_l}\right] = \mathbf{E}_{l^n}\left[\langle\sigma_{X_1}^{m_1}...\sigma_{X_l}^{m_l}\rangle_{\mathcal{C}}\langle\sigma_{X_1}\rangle_{\mathcal{C}}^{m_1}...\langle\sigma_{X_l}\rangle_{\mathcal{C}}^{m_l}\right] \tag{3.2}$$

An immediate application is the analog of the first GKS inequality obtained by taking one set X_1 and $m_1 = 1$,

$$\mathbb{E}_{l^n}[\langle\sigma_{X_1}\rangle_{\mathcal{C}}] = \mathbb{E}_{l^n}[\langle\sigma_{X_1}\rangle_{\mathcal{C}}^2] \geq 0 \tag{3.3}$$

The following correlation inequality is the analog of the second GKS inequality. The proof is presented in appendix B.

Theorem 3.1. [Monotonicity under check node erasure.] *Fix a linear code and its asociated factor graph $\mathcal{C}$. Take any check node $B \in \mathcal{C}$ and consider the factor graph $\mathcal{C}\backslash B$ obtained by removing the check node B together with its outgoing edges. Consider the Gibbs measures $\langle - \rangle_{\mathcal{C}}$ and $\langle - \rangle_{\mathcal{C}\backslash B}$. For any subset $X \subset \{1, ..., n\}$ and any integer m we have*

$$\mathbb{E}_{l^n}[\langle\sigma_X\rangle_{\mathcal{C}}^m] \geq \mathbb{E}_{l^n}[\langle\sigma_X\rangle_{\mathcal{C}\backslash B}^m] \tag{3.4}$$

Remarks. This inequality states that we can start with the left hand side and form a monotone decreasing sequence by successively erasing check nodes. We will apply this inequality for the spin system with $l_i = 0$ for a given i. Then it becomes $\mathbb{E}_{l^n\backslash i}[\langle\sigma_X\rangle_{\mathcal{C}}^m] \geq \mathbb{E}_{l^n\backslash i}[\langle\sigma_X\rangle_{\mathcal{C}\backslash B}^m]$.

We are now ready to prove our main result.

Proof of theorem 1.1 The most convenient representation for us is expression (2.17). If we consider the average over the code ensemble, by symmetry we have

$$\mathbb{E}_{\mathcal{C}}[g_{MAP}^{(n)}(\epsilon)] = \mathbb{E}_{\mathcal{C},l^n\backslash 1}\left[\int_{-\infty}^{+\infty} dl_1 \frac{dc(l_1)}{d\epsilon} \ln\left(\frac{1 + \langle\sigma_1\rangle_{\mathcal{C},l_1=0}\tanh\frac{l_1}{2}}{1 + \tanh\frac{l_1}{2}}\right)\right] \tag{3.5}$$

Expanding the logarithm we obtain

$$\mathbb{E}_{\mathcal{C}}[g_{MAP}^{(n)}(\epsilon)] = \sum_{m=0}^{\infty} \frac{(-1)^m}{m} \Big(1 - \mathbb{E}_{\mathcal{C},l^{n\backslash 1}}[\langle \sigma_1 \rangle_{\mathcal{C},l_1=0}^{m}] \Big) \times \int_{-\infty}^{+\infty} dl \frac{dc(l)}{d\epsilon} (\tanh \frac{l_1}{2})^m \quad (3.6)$$

Using (3.2) with $X_1 = \{1\}$ and $m_1 = 2p-1$ we see that

$$\mathbb{E}_{\mathcal{C},l^{n\backslash 1}}[\langle \sigma_1 \rangle_{\mathcal{C},l_1=0}^{2p-1}] = \mathbb{E}_{\mathcal{C},l^{n\backslash 1}}[\langle \sigma_1 \rangle_{\mathcal{C},l_1=0}^{2p}] \quad (3.7)$$

Also, applying such an identity to a spin system with the simple Hamiltonian $\mathcal{H}(s) = \frac{l}{2} s$ we have

$$\int_{-\infty}^{+\infty} dl c(l) \big(\tanh \frac{l}{2}\big)^{2p-1} = \int_{-\infty}^{+\infty} dl c(l) \big(\tanh \frac{l}{2}\big)^{2p} \quad (3.8)$$

Then the sum (3.6) becomes

$$\mathbb{E}_{\mathcal{C}}[g_{MAP}^{(n)}(\epsilon)] = \sum_{p=1}^{\infty} \big(\frac{1}{2p} - \frac{1}{2p-1} \big) \Big(1 - \mathbb{E}_{\mathcal{C},l^{n\backslash 1}}[\langle \sigma_1 \rangle_{\mathcal{C},l_1=0}^{2p}] \Big) \times \int_{-\infty}^{+\infty} dl \frac{dc(l)}{d\epsilon} \big(\tanh \frac{l}{2}\big)^{2p} \quad (3.9)$$

Now consider node 1 and its neighborhood $\mathcal{T}_1^{(d)}$ of depth d, where d is an even integer. More precisely, $n \in \mathcal{T}_1^{(d)}$ (where n is a variable or a check node) if and only if the length of the shortest path from 1 to n is at most equal to d. The correlation inequality of theorem 3.1 can be used to show

$$\mathbb{E}_{l^{n\backslash 1}}[\langle \sigma_1 \rangle_{\mathcal{C},l_1=0}] \geq \mathbb{E}_{l^{n\backslash 1}}[\langle \sigma_1 \rangle_{\mathcal{T}_1^{(d)},l_1=0}] \quad (3.10)$$

Indeed, let $C_1^{(d)}$ denote the check nodes and $V_1^{(d)}$ the variable nodes in the complement of $\mathcal{T}_1^{(d)}$. Theorem 3.1 implies

$$\mathbb{E}_{l^{n\backslash 1}}[\langle \sigma_1 \rangle_{\mathcal{C},l_1=0}] \geq \mathbb{E}_{l^{n\backslash 1}}[\langle \sigma_1 \rangle_{\mathcal{C}\backslash C_1^{(d)},l_1=0}] \quad (3.11)$$

Now the Gibbs average $\langle - \rangle_{\mathcal{C}\backslash C_1^{(d)},l_1=0}$ contains the *free spin* terms[3]

$$\sum_{\sigma_i, i \in V_1^{(d)}} \prod_{i \in V_1^{(d)}} e^{\frac{l_i}{2}\sigma_i} = \prod_{i \in V_1^{(d)}} 2 \cosh \frac{l_i}{2} \quad (3.12)$$

[3] One can think of them as nodes of zero degree

in both the denominator and the numerator. These terms cancel which means $\langle\sigma_1\rangle_{\mathcal{C}\setminus C_1^{(d)},l_1=0} = \langle\sigma_1\rangle_{\mathcal{T}_1^{(d)},l_1=0}$ and we get (3.10). Next we use an important observation of Richardson and Urbanke [2]. Namely that if the channel family is ordered by physical degradation,

$$\int_{-\infty}^{+\infty} dl \frac{dc(l)}{d\epsilon}\big(\tanh\frac{l}{2}\big)^{2p} \leq 0 \tag{3.13}$$

Now taking into account $\frac{1}{2p} - \frac{1}{2p-1} < 0$, (3.10), (3.13) and summing the expansion of the logarithm we get

$$\mathbb{E}_{\mathcal{C}}[g_{MAP}^{(n)}(\epsilon)] \leq \mathbb{E}_{\mathcal{C},l^{n\setminus 1}}\Bigg[\int_{-\infty}^{+\infty} dl_1 \frac{dc(l_1)}{d\epsilon} \ln \frac{1+\langle\sigma_1\rangle_{\mathcal{T}_1^{(d)},l_1=0}\tanh\frac{l_1}{2}}{1+\tanh\frac{l_1}{2}}\Bigg] \tag{3.14}$$

The virtue of LDPC ensembles with bounded (say by k) node degrees is that with high probability, namely $1 - O(\frac{k^{4d}}{n})$, $\mathcal{T}_1^{(d)}$ is a tree. On a tree it is possible to compute exactly the average $\langle\sigma_1\rangle_{\mathcal{T}_1^{(d)},l_1=0}$ and one finds that

$$\langle\sigma_1\rangle_{\mathcal{T}_1^{(d)},l_1=0} = \Delta_1^{(d)} \tag{3.15}$$

where $\Delta_1^{(d)}$ is computed by the message passing procedure on the tree $\mathcal{T}_1^{(d)}$: the initial condition $l_{i\to C}^{(0)} = l_i$ is applied to leaf nodes and messages are passed until one reaches the root node 1. Therefore

$$\mathbb{E}_{\mathcal{C}}[g_{MAP}^{(n)}(\epsilon)] \leq \big(1-O(\frac{k^{4d}}{n})\big)\mathbb{E}_{\mathcal{C},l^{n\setminus 1}}\Bigg[\int_{-\infty}^{+\infty} dl_1 \frac{dc(l_1)}{d\epsilon} \ln \frac{1+\Delta_1^{(d)}\tanh\frac{l_1}{2}}{1+\tanh\frac{l_1}{2}}\Bigg]+O(\frac{k^{4d}}{n}) \tag{3.16}$$

The first term on the r.h.s is the probability that $\mathcal{T}_1^{(d)}$ is a tree times the expectation conditionned to that event and the second term comes from the probability that $\mathcal{T}_1^{(d)}$ is not a tree. Note that the expectation on the right hand side is independent of n since it involves quantities defined on random trees $\mathcal{T}_1^{(d)}$. The density of $\Delta_1^{(d)} = \tanh\frac{\Lambda_1^{(d)}}{2}$, can be inferred from the BP message passing equations on the trees, and satisfies the density evolution equations. Let us call $a_{DE}^{(d)}(\Lambda)$ the density of $\Lambda_1^{(d)}$ given by density evolution. We then consider the limit as $n \to \infty$ for d fixed on both sides, and express the right hand side in terms of the extrinsic log-likelihood ratio (2.24),

$$\limsup_{n\to+\infty} \mathbb{E}_{\mathcal{C}}[g_{MAP}^{(n)}(\epsilon)] \leq \int_{-\infty}^{+\infty} d\Lambda a_{DE}^{(d)}(\Lambda) \int_{-\infty}^{+\infty} dl \frac{dc(l)}{d\epsilon}\ln(1+e^{-l-\Lambda}) \tag{3.17}$$

Now it remains to check that the limit of the r.h.s when $d \to +\infty$ exists. This is again an easy consequence of theorem 3.1. Indeed consider trees of depth d and $d+2$. The correlation inequality applied to tree graphs implies that $\Delta_1^{(d+2)} \geq \Delta_1^{(d)}$. Thus considering again the expansion of the logarithm in (3.16) we conclude that the r.h.s of (3.17) is an increasing sequence. Since it is bounded it converges and this completes the proof of the theorem.

4 Binary erasure and gaussian channels

In the case of BEC and BIAWGNC most expressions can be simplified and the proofs are more transparent. The purpose of this section is to briefly discuss these simplifications.

4.1 BEC and classical GKS inequality

As shown here it turns out that a simpler correlation inequality of Griffiths-Kelly-Sherman (GKS) pertaining to *non-random spin systems* applies directly. The output alphabet is $0, e, 1$, with the corresponding log-likelihood ratios $l(0) = +\infty$, $l(e) = 0$, $l(1) = -\infty$. Thus $c(l) = (1-\epsilon)\delta_{+\infty}(l) + \epsilon\delta_0(l)$. The MAP and BP GEXIT curves become

$$g_{MAP}^{(n)}(\epsilon) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{l^n\setminus i}[\ln(1+e^{-L_i})] \tag{4.1}$$

and

$$g_{BP}^{(n,d)}(\epsilon) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{l^n\setminus i}[\ln(1+e^{-\Lambda_i^{(d)}})] \tag{4.2}$$

Here it is more convenient to use the expression (2.19) which becomes

$$g_{MAP}^{(n)}(\epsilon) = -\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{l^n\setminus i}\left[\ln\frac{1}{2}\big(1+\langle\sigma_i\rangle_{\mathcal{C},l_i=0}\big)\right] \tag{4.3}$$

where

$$\langle\sigma_1\rangle_{\mathcal{C},l_i=0} = \frac{1}{Z}\sum_{\sigma^n}\sigma_i\prod_{A\in\mathcal{C}}\frac{1}{2}(1+\sigma_A)\prod_{j\in E^c\setminus i}\frac{1}{2}(1+\sigma_i) \tag{4.4}$$

with Z the obvious normalisation factor and E the set of erased bits and E^c the set of received 0's (known bits). Obvioully this corresponds to a spin system defined by the hamiltonian (2.2) with $J_A = +\infty$ and $l_i = 0$ if $i \in E$ and $l_i = +\infty$ if $i \in E^c$. This spin sytem belongs to the class of ferromagnetic systems which are those for which *all* coefficients of the Hamiltonian are positive. For such systems we have [10]

Theorem 4.1. **[Griffiths-Kelly-Sherman]** *Given the hamiltonian* (2.2) *if all coeeficients J_A and h_i are non negative then for any $X \subset \{1, ..., n\}$ $\langle \sigma_X \rangle_{\mathcal{C}}$ is non negative and is an increasing function of each coefficient.*

Therefore for each individual instance of the channel outputs

$$\langle \sigma_i \rangle_{\mathcal{C}, l_i=0} \geq 0, \qquad \langle \sigma_i \rangle_{\mathcal{C}, l_i=0} \geq \langle \sigma_i \rangle_{\mathcal{C} \setminus B, l_i=0}, \qquad \text{any} B \in \mathcal{C} \tag{4.5}$$

An immediate application yields

$$\mathbb{E}_{\mathcal{C}}[g_{MAP}^{(n)}(\epsilon)] \leq -\mathbb{E}_{\mathcal{C}, l^{n \setminus 1}}\big[\ln \frac{1}{2}(1 + \langle \sigma_1 \rangle_{\mathcal{T}_1^{(d)}, l_1=0})\big] \tag{4.6}$$

where $\mathcal{T}_1^{(d)}$ is a neighborhood of depth d for variable node 1. Then, proceeding exactly as in section 3 we obtain the final estimate

$$\limsup_{n \to +\infty} \mathbb{E}_{\mathcal{C}}[g_{MAP}^{(n)}(\epsilon)] \leq \lim_{d \to +\infty} \int_{-\infty}^{\infty} d\Lambda a_{DE}^{(d)}(\Lambda) \ln(1 + e^{-\Lambda}) \tag{4.7}$$

For the BEC channel the right hand side can be computed exactly in terms of the degree distributions of the specific LDPC ensemble (see [2] for explicit formulas). Finaly let us remark that the inequality (4.7) is equivalent to the well known fact that the MAP decoder is better than the BP (or any other) decoder. Indeed from (4.1)

$$\begin{aligned}\frac{\epsilon}{\ln 2}\mathbb{E}_{\mathcal{C}}[g_{MAP}^{(n)}(\epsilon)] &= \epsilon \text{P}r(L_1 = 0 | l_1 = 0) = \epsilon \text{P}r(L_1 + l_1 = 0 | l_1 = 0) \\ &= \epsilon \text{P}r(L_1 + l_1 = 0 | l_1 = 0) + (1 - \epsilon)\text{P}r(L_1 + l_1 = 0 | l_1 = +\infty) \\ &= \text{P}r(L_1 + l_1 = 0) = P_{MAP}^{(n)}(\epsilon)\end{aligned} \tag{4.8}$$

and similarly

$$\frac{\epsilon}{\ln 2}\int_{-\infty}^{\infty} d\Lambda a_{DE}^{(d)}(\Lambda) \ln(1 + e^{-\Lambda}) = P_{BP}^{(n,d)}(\epsilon) \tag{4.9}$$

4.2 BIAWNG channel

It is in this case that the statistical mechanical formulation is most transparent because the MAP-GEXIT curve takes a very simple form

$$g_{MAP}^{(n)}(\epsilon) = \frac{\sigma^{-3}}{n}\sum_{i=1}^{n} \mathbb{E}_{l^n}[1 - d_i] = \frac{\sigma^{-3}}{n}\sum_{i=1}^{n} \mathbb{E}_{l^n}[1 - \langle \sigma_i \rangle_{\mathcal{C}}] \tag{4.10}$$

where σ^{-2} is the signal to noise ratio. We remark that this formula is analog to the relationship between mutual information and MMSE for gaussian channels [18], [6]. The difference is that here the alphabet is binary and we have a Nishimori identity, $\mathbb{E}_{l^n}[\langle\sigma_i\rangle_{\mathcal{C}}] = \mathbb{E}_{l^n}[\langle\sigma_i\rangle_{\mathcal{C}}^2]$, so that

$$\mathbb{E}_{l^n}[1 - \langle\sigma_i\rangle_{\mathcal{C}}] = \mathbb{E}_{l^n}[\langle\sigma_i^2\rangle_{\mathcal{C}} - \langle\sigma_i\rangle_{\mathcal{C}}^2] \tag{4.11}$$

This being said let us show how to obtain (4.10). For a BIAWGNC with signal to noise ratio σ^{-2} we have

$$c(l_i) = \frac{1}{\sqrt{8\pi\sigma^{-2}}} e^{-\frac{(l_i - 2\sigma^{-2})^2}{8\sigma^{-2}}} \tag{4.12}$$

and

$$\frac{dc(l_i)}{d\sigma} = -4\sigma^{-3}(-\frac{\partial}{\partial l_i} + \frac{\partial^2}{\partial l_i^2})c(l_i) \tag{4.13}$$

Replacing this expression in the formulas for the MAP-GEXIT curve one gets (4.10) after some calculus.

However there is a simpler calculation starting directly from (2.11). First of all we note that (2.12) is equal to σ^{-2}. Thus using (4.13) and integration by parts

$$\frac{d}{d\sigma} H(X^n|Y^n) = -4\sigma^{-3} \sum_{i=1}^{n} \mathbb{E}_{l^n} \left[(\frac{\partial}{\partial l_i} + \frac{\partial^2}{\partial l_i^2}) \ln Z_{\mathcal{C}} \right] - 2\sigma^{-3} \tag{4.14}$$

The definition of Gibbs averages implies

$$\frac{\partial}{\partial l_i} \ln Z_{\mathcal{C}} = \frac{1}{2}\langle\sigma_i\rangle_{\mathcal{C}}, \qquad \frac{\partial^2}{\partial l_i^2} \ln Z_{\mathcal{C}} = \frac{1}{4}(\langle\sigma_i^2\rangle_{\mathcal{C}} - \langle\sigma_i\rangle_{\mathcal{C}}^2) = \frac{1}{4}(1 - \langle\sigma_i\rangle_{\mathcal{C}}^2) \tag{4.15}$$

Replacing these identities in (4.14) and using the Nishimori identity we immediately obtain (4.10).

In order to get the bound on the GEXIT curve we apply theorem 3.1,

$$\mathbb{E}_{\mathcal{C}}[g_{MAP}^{(n)}(\epsilon)] = \sigma^{-3}\mathbb{E}_{\mathcal{C},l^n}[1 - \langle\sigma_1\rangle_{\mathcal{C}}] \leq \sigma^{-3}\mathbb{E}_{\mathcal{C},l^n}[1 - \langle\sigma_1\rangle_{\mathcal{T}_1^{(d)}}] \tag{4.16}$$

If the neighborhood of node 1, namely $\mathcal{T}_1^{(d)}$, is a tree the Gibbs average can be computed recursively

$$\langle\sigma_1\rangle_{\mathcal{T}_1^{(d)}} = \tanh\frac{1}{2}(l_1 + \Lambda_1^{(d)}) \tag{4.17}$$

Since the graph is a tree with high probability, we can proceed as in section 3 to get the final result

$$\limsup_{n\to+\infty} \mathbb{E}_{\mathcal{C}}[g_{MAP}^{(n)}(\epsilon)] \leq \lim_{d\to+\infty} \sigma^{-3} \int_{-\infty}^{+\infty} d\Lambda a_{DE}^{(d)}(\Lambda) \int_{-\infty}^{+\infty} dl c(l) \tanh\frac{1}{2}(l + \Lambda) \tag{4.18}$$

One may check that for the gaussian channel the right hand side of (1.8) and (4.18) are the same.

5 Relationship with Bounds from Interpolation Method

It turns out that the bounds discussed in this paper are closely related to the ones obtained by the interpolation methods. This is interesting in its own rigth but also means that correlation inequalities might be used to approach in a rigorous way other problems where the replica method is successful. Here the discussion will remain at a formal level due to some technicalities.

We denote the degree distributions of the LDPC ensemble from the edge perspective as $\lambda(x) = \sum_m \lambda_m x^{m-1}$ and $\rho(x) = \sum_k \rho_k x^k$, and from the node perspective as $\Lambda(x) = \sum_m \Lambda_m x^m$ and $P(x) = \sum_k P_k x^k$. In terms of the latter the design rate is $r = 1 - \frac{\Lambda'(1)}{P'(1)}$. The bounds involve a functional[4] of two probability distributions $\zeta(l)$ and $\hat{\zeta}(u)$,

$$\begin{aligned} f[\zeta, \hat{\zeta}; \epsilon] = &-\frac{\Lambda'(1)}{P'(1)} \ln 2 - \Lambda'(1) \int dl \zeta(l) \int du \hat{\zeta}(u) \ln(1 + \tanh \frac{l}{2} \tanh \frac{u}{2}) \\ &+ \sum_m \Lambda_m \int dl' c(l') \int \prod_{c=1}^{m} du_c \hat{\zeta}(u_c) \ln\Big(e^{\frac{l'}{2}} \prod_{c=1}^{m} (1 + \tanh \frac{u_c}{2}) \\ &+ e^{-\frac{l'}{2}} \prod_{c=1}^{m} (1 - \tanh \frac{u_c}{2}) \Big) + \frac{\Lambda'(1)}{P'(1)} \sum_k P_k \int \prod_{i=1}^{k} dl_i \zeta(l_i) \ln(1 + \prod_{i=1}^{k} \tanh \frac{l_i}{2}) \end{aligned} \tag{5.1}$$

We emphasize that in this expression the ϵ dependence enters only through $c(l')$. The *replica symmetric solution* to the free energy is

$$f_{RS}(\epsilon) = \sup_{\zeta \in S} f[\zeta, \hat{\zeta}(\zeta); \epsilon] \tag{5.2}$$

and is believed to be exact. In this last formula the supremum is taken over the set S of "symmetric" probability distributions satisfying $\zeta(-l) = \zeta(l)e^{-l}$ and it is understood that the conjugate variable $\hat{\zeta}$ is replaced by

$$\hat{\zeta}(u) = \sum_k \rho_k \int \prod_{i=1}^{k-1} dl_i \zeta(l_i) \delta(u - 2 \tanh^{-1}(\prod_{i=1}^{k-1} \tanh l_i)) \tag{5.3}$$

More precisely we have the following conjecture:

[4] In the language of statistical mechanics f is a "Landau functional" and ζ, $\hat{\zeta}$ the "order parameters". See [19] for an introduction to these concepts.

Conjecture. *Given a sequence of $LDPC(\lambda,\rho,n)$ ensembles we have*

$$\lim_{n\to+\infty}\frac{1}{n}\mathbb{E}_{\mathcal{C}}[H(X^n|Y^n)] = f_{RS}(\epsilon) - \int_{-\infty}^{+\infty} dl c(l)\frac{l}{2} \tag{5.4}$$

Montanari [12] has obtained the lower bound by an application of the *interpolation method* invented by Guerra and Toninelli for the Sherrington-Kirkpatrick model [13], and further developped in [14] for dilute spin systems. The precise statement is that $LDPC(\lambda,\rho,n)$ ensembles with *convex* P (e.g. for regular check node degree this degree is even)[5]

$$\liminf_{n\to\infty}\frac{1}{n}\mathbb{E}_{\mathcal{C}}[H(X^n|Y^n)] \geq f_{RS}(\epsilon) - \int_{-\infty}^{+\infty} dl c(l)\frac{l}{2} \tag{5.5}$$

The critical points of this functional are the solutions of the equations $\frac{\delta f}{\delta\zeta}=0$ and $\frac{\delta f}{\delta\hat{\zeta}}=0$, whose iterative version are the density evolution equations

$$\hat{\zeta}^{(d+2)}(u) = \sum_k \rho_k \int \prod_{i=1}^{k-1} dl_i \zeta^{(d+1)}(l_i)\delta(u - 2\tanh^{-1}(\prod_{i=1}^{k-1}\tanh l_i)) \tag{5.6}$$

$$\zeta^{(d+1)}(l) = \sum_m \lambda_m \int dl' c(l') \int \prod_{c=1}^{m-1} du_c \hat{\zeta}^{(d)}(u_c)\delta(l - l' - \sum_{c=1}^{m-1} u_c) \tag{5.7}$$

with the initial condition $\zeta^{(1)}(l) = c(l)$ and $\hat{\zeta}^{(0)}(u) = \delta(u)$. We define the iterative or BP free energy as

$$f_{BP}^{(d)}(\epsilon) = f[\zeta^{(d+1)}, \hat{\zeta}^{(d+2)}; \epsilon] \tag{5.8}$$

We have

$$\begin{aligned}\lim_{d\to\infty}\frac{\partial}{\partial\epsilon}\left(f_{BP}^{(d)}(\epsilon) - \int_{-\infty}^{+\infty} dl c(l)\frac{l}{2}\right)&\\ = \lim_{d\to\infty}\int_{-\infty}^{+\infty} d\Lambda a_{DE}^{(d)}(\Lambda)\int_{-\infty}^{+\infty}\frac{dc(l)}{d\epsilon}\ln(1+e^{-l-\Lambda})&\end{aligned} \tag{5.9}$$

This was shown for the BIAWNGC in [16]. Let us briefly give the main steps for general output symmetric channels. Using that $\frac{d}{d\epsilon}\int_{-\infty}^{+\infty} dl' c(l') = 0$ we

[5] This has been extended to any P for the BEC, BIAWGNC (any noise level) and the BSC (high noise)

easily obtain

$$\frac{\partial}{\partial\epsilon}f[\zeta,\hat{\zeta};\epsilon]=\int_{-\infty}^{+\infty}dl'\frac{dc(l')}{d\epsilon}\sum_m\Lambda_m\int\prod_{c=1}^m du_c\hat{\zeta}(u_c)\ln\Big(e^{\frac{l'}{2}}\prod_{c=1}^m(1+\tanh\frac{u_c}{2})$$
$$+e^{-\frac{l'}{2}}\prod_{c=1}^m(1-\tanh\frac{u_c}{2})\Big)$$
$$=\int_{-\infty}^{+\infty}dl'\frac{dc(l')}{d\epsilon}\sum_m\Lambda_m\int\prod_{c=1}^m du_c\hat{\zeta}(u_c)\ln\Big(1+e^{-l'-\sum_{c=1}^m u_c}\Big)$$
$$+\int_{-\infty}^{+\infty}dlc(l)\frac{l}{2}\quad(5.10)$$

Replacing now ζ and $\hat{\zeta}$ by $\zeta^{(d)}$ and $\hat{\zeta}^{(d)}$ we obtain (5.9).

We claim that for almost all ϵ, the partial derivative in (5.9) can be replaced by a total derivative. Indeed formaly,

$$\frac{d}{d\epsilon}f_{BP}^{(d)}(\epsilon)=\frac{\partial}{\partial\epsilon}f[\zeta^{(d+1)},\hat{\zeta}^{(d+2)};\epsilon]$$
$$+\int dl\frac{\delta f}{\delta\zeta(l)}[\zeta^{(d+1)},\hat{\zeta}^{(d+2)};\epsilon]\frac{\partial}{\partial\epsilon}\zeta^{(d+1)}(l)+\int du\frac{\delta f}{\delta\hat{\zeta}(u)}[\zeta^{(d+1)},\hat{\zeta}^{(d+2)};\epsilon]\frac{\partial}{\partial\epsilon}\zeta^{(d+2)}(u)\quad(5.11)$$

As long as the critical points of the functional (5.1) are unique and behave smoothly with respect to ϵ one expects that the integrals tend to zero as $d\to\infty$. This is because the functional derivatives tend to zero and the ϵ derivatives are bounded. At threshhold points however, the ϵ derivatives become "infinite" so that the integrals will have a non trivial contribution. This justifies the claim that away from discontinuity points,

$$\lim_{d\to\infty}\frac{d}{d\epsilon}\Big(f_{BP}^{(d)}(\epsilon)-\int_{-\infty}^{+\infty}dlc(l)\frac{l}{2}\Big)$$
$$=\lim_{d\to\infty}\int_{-\infty}^{+\infty}d\Lambda a_{DE}^{(d)}(\Lambda)\int_{-\infty}^{+\infty}dl\frac{dc(l)}{d\epsilon}\ln(1+e^{-l-\Lambda})\quad(5.12)$$

Because of this identity we know explicitely a primitive of the BP-EXIT function. Therefore an integration of both sides of (1.8) leads to bounds on the average conditional entropy. To keep the discussion simple we assume that the MAP and BP GEXIT curves each have only one discontinuity point ϵ_{MAP} and ϵ_{BP} (of course $\epsilon_{BP}<\epsilon_{MAP}$). Then for $\epsilon>\epsilon_{MAP}$ integrating (1.8) from ϵ to $+\infty$ we get

$$\liminf_{n\to\infty}\frac{1}{n}\mathbb{E}_{\mathcal{C}}[H(X^n|Y^n)]\geq\lim_{d\to\infty}\Big(f_{BP}^{(d)}(\epsilon)-\int_{-\infty}^{+\infty}dlc(l)\frac{l}{2}\Big)\quad(5.13)$$

We expect that for $\epsilon > \epsilon_{MAP}$ $f_{RS}(\epsilon) = \lim_{d\to+\infty} f_{BP}^{(d)}(\epsilon)$ so that this lower bound is the same as the interpolation bound. On the other hand for $\epsilon < \epsilon_{BP}$ we can integrate from 0 to ϵ which yields

$$\liminf_{n\to\infty} \frac{1}{n}\mathbb{E}_{\mathcal{C}}[H(X^n|Y^n)] \leq \lim_{d\to\infty}\big(f_{BP}^{(d)}(\epsilon) - \int_{-\infty}^{+\infty} dlc(l)\frac{l}{2}\big) \tag{5.14}$$

Combining with the interpolation bound (5.5) we find that for $\epsilon < \epsilon_{BP}$,

$$\begin{aligned}\liminf_{n\to+\infty} \frac{1}{n}\mathbb{E}_{\mathcal{C}}[H(X^n|Y^n)] &= f_{RS}(\epsilon) - \int_{-\infty}^{+\infty} dlc(l)\frac{l}{2} \\ &= \lim_{d\to\infty}\big(f_{BP}^{(d)}(\epsilon) - \int_{-\infty}^{+\infty} dlc(l)\frac{l}{2}\big) \end{aligned} \tag{5.15}$$

This confirms the above conjecture for $\epsilon < \epsilon_{BP}$. However the content of this equality is trivial since below the BP threshold both sides vanish.

Note that this last equality also holds, and is non trivial, for $LDPC(\lambda, \rho, n)$ ensembles with *no discontinuity* for the GEXIT curves. An example is given by the case of regular right degree and a Poisson left degree which was considered in [16] for the gaussian channel. These do not constitute good codes since there are always $O(n)$ errors, but it is an interesting theoretical result since it confirms the above conjecture for all ϵ.

6 Concluding remarks

The check erasure inequality of Theorem 3.1 is valid for each fixed linear code. This implies a version of Theorem 1.1 that holds for non-averaged GEXIT curves. If one considers a bit dependent noise level ϵ_i one can define $g_{MAP}^{(n,i)}(\epsilon_1, ..., \epsilon_n) = \frac{1}{n}\frac{d}{d\epsilon_i}H(X^n|Y^n)$. This equals the i-th term of the MAP-GEXIT formula (1.5). Analogously the i-th term of the BP-GEXIT expression (1.6) defines $g_{BP}^{(n,d,i)}(\epsilon_1, ..., \epsilon_n)$ and is computed from the BP decoder for bit i. As before we consider a neighborhood of node i and erase all checks outside. As long as the neighborhood of i is a tree the Gibbs average is computed exactly by the BP decoder. Therefore with probability at least $1 - O(\frac{k^{4d}}{n})$ we have

$$g_{MAP}^{(n,i)}(\epsilon_1, ..., \epsilon_n) \leq g_{BP}^{(n,d,i)}(\epsilon_1, ..., \epsilon_n) \tag{6.1}$$

We wish to conclude with a few remarks about the connection between the present approach and the method of physical degradation which we first explain. Consider two BMS channels with transition probabilities $q_{X|Y}^{\epsilon_1}(x|y)$

and $q^{\epsilon_2}_{X|Y}(x|y)$ ordered by physical degradation $\epsilon_1 < \epsilon_2$. Let $w^\epsilon_i = q^\epsilon_{X|Y}(0|y) - q^\epsilon_{X|Y}(1|y)$. A basic observation in [2] (chap 3) is

$$\mathbb{E}_{w_i^{\epsilon_1}}[w_i^{\epsilon_1}|w_i^{\epsilon_2}] = w_i^{\epsilon_2} \tag{6.2}$$

From this it follows that for any *concave* function F

$$\mathbb{E}_{w_i^{\epsilon_1}}[F(w_i^{\epsilon_1})] \leq \mathbb{E}_{w_i^{\epsilon_2}}[F(w_i^{\epsilon_2})] \tag{6.3}$$

An application of this result to uncoded transmission where $w_i = \tanh\frac{l_i}{2}$ immediately yields

$$\int_{-\infty}^{+\infty} dl_i \frac{dc(l_i)}{d\epsilon}\big(\tanh\frac{l_i}{2}\big)^{2p} \leq 0 \tag{6.4}$$

One may also apply (6.3) to soft bit estimates from MAP decoding namely $w_i = d_i = \langle\sigma_i\rangle_{\mathcal{C}}$. For example consider $d_i = \langle\sigma_i\rangle_{\mathcal{C}}$ for a given ϵ and the physicaly degraded version $D_i = \langle\sigma_i\rangle_{\mathcal{C},l_i=0}$ corresponding to the concatenation of the channel with another channel which erases bit i with probability 1. Then (6.3) yields as a special case

$$\mathbb{E}_{l^n}[\langle\sigma_i\rangle_{\mathcal{C}}^{2p}] \geq \mathbb{E}_{l^{n\setminus i}}[\langle\sigma_i\rangle_{\mathcal{C},l_i=0}^{2p}] \tag{6.5}$$

Alternatively one may consider physical degradation as a function of the channel parameter to obtain

$$\frac{d}{d\epsilon}\mathbb{E}_{l^n}[\langle\sigma_i\rangle_{\mathcal{C}}^{2p}] \leq 0 \tag{6.6}$$

In fact (6.5) and (6.6) are correlation inequalities that closely ressemble (3.4). In appendix C we prove the following generalizations, by the same methods used to prove (3.4). For any subset $X \subset \{1, ..., n\}$ and any $i = 1, ..., n$

$$\mathbb{E}_{l^n}[\langle\sigma_X\rangle_{\mathcal{C}}] \geq \mathbb{E}_{l^{n\setminus i}}[\langle\sigma_X\rangle_{\mathcal{C},l_i=0}] \tag{6.7}$$

For a family of physicaly degraded channels and any subset $X \subset \{1, ..., n\}$,

$$\frac{d}{d\epsilon}\mathbb{E}_{l^n}[\langle\sigma_X\rangle_{\mathcal{C}}] \leq 0 \tag{6.8}$$

Acknowledgment. The author would like to thank Cyril Méasson, Henry Pfister, Ruediger Urbanke for useful discussions and one of the referees for a useful remark.

A The generalised Nishimori identities

We start with the left hand side of (3.2) and perform a *gauge transformation*. By this we mean that we take a fixed codeword $(\tau_1, ..., \tau_n) \in \mathcal{C}$ and do the local change of variables $\sigma_i \to \tau_i\sigma_i$, $l_i \to \tau_i l_i$ for $i = 1, ..., n$. Because of channel symmetry (3.1)

$$\mathbf{E}_{l^n}\left[\langle\sigma_{X_1}\rangle_{\mathcal{C}}^{m_1}...\langle\sigma_{X_l}\rangle_{\mathcal{C}}^{m_l}\right] = \mathbf{E}_{l^n}\left[\prod_{i=1}^{n} e^{\frac{l_i}{2}(\tau_i-1)}\tau_{X_1}^{m_1}...\tau_{X_l}^{m_l}\langle\sigma_{X_1}\rangle_{\mathcal{C}}^{m_1}...\langle\sigma_{X_l}\rangle_{\mathcal{C}}^{m_l}\right] \quad \text{(A.1)}$$

Next sum over all possible codewords. Denoting by $|\mathcal{C}|$ the number of code words we have,

$$\begin{aligned}\mathbf{E}_{l^n}\left[\langle\sigma_{X_1}\rangle_{\mathcal{C}}^{m_1}...\langle\sigma_{X_l}\rangle_{\mathcal{C}}^{m_l}\right] &= \frac{1}{|\mathcal{C}|}\mathbf{E}_{l^n}\left[Z_{\mathcal{C}}\prod_{i=1}^{n} e^{-\frac{l_i}{2}}\langle\tau_{X_1}^{m_1}...\tau_{X_l}^{m_l}\rangle_{\mathcal{C}}\langle\sigma_{X_1}\rangle_{\mathcal{C}}^{m_1}...\langle\sigma_{X_l}\rangle_{\mathcal{C}}^{m_l}\right] \\ &= \frac{1}{|\mathcal{C}|}\sum_{\rho^n\in\mathcal{C}}\mathbf{E}_{l^n}\left[\prod_{i=1}^{n} e^{-\frac{l_i}{2}(\rho_i-1)}\langle\tau_{X_1}^{m_1}...\tau_{X_l}^{m_l}\rangle_{\mathcal{C}}\langle\sigma_{X_1}\rangle_{\mathcal{C}}^{m_1}...\langle\sigma_{X_l}\rangle_{\mathcal{C}}^{m_l}\right] \quad \text{(A.2)}\end{aligned}$$

Finaly we perform a second gauge transformation. For each term in the above sum we do $\sigma_i \to \rho_i\sigma_i$, $\tau_i \to \rho_i\tau_i$, $l_i \to \rho_i l_i$. Again due to channel symmetry

$$\begin{aligned}\mathbf{E}_{l^n}\left[\langle\sigma_{X_1}\rangle_{\mathcal{C}}^{m_1}...\langle\sigma_{X_l}\rangle_{\mathcal{C}}^{m_l}\right] = \frac{1}{|\mathcal{C}|}\sum_{\rho^n\in\mathcal{C}}\mathbf{E}_{l^n}\Bigg[&\prod_{i=1}^{n} e^{-\frac{l_i}{2}(\rho_i-1)(\rho_i+1)} \\ &\times \langle\tau_{X_1}^{m_1}...\tau_{X_l}^{m_l}\rangle_{\mathcal{C}}\langle\sigma_{X_1}\rangle_{\mathcal{C}}^{m_1}...\langle\sigma_{X_l}\rangle_{\mathcal{C}}^{m_l}\Bigg] \quad \text{(A.3)}\end{aligned}$$

Since $(\rho_i - 1)(\rho_i + 1) = 0$ we have obtained the desired identity.

B Proof of Theorem 3.1

Let us start with a simple proof for the case $m = 1$. We remark that for any $X \subset \{1, ..., n\}$ and any check node B,

$$\langle\sigma_X\rangle_{\mathcal{C}} = \frac{\langle\sigma_X\rangle_{\mathcal{C}\setminus B} + \langle\sigma_X\sigma_B\rangle_{\mathcal{C}\setminus B}}{1 + \langle\sigma_B\rangle_{\mathcal{C}\setminus B}} \quad \text{(B.1)}$$

Expanding the denominator and grouping terms appropriately leads to

$$\begin{aligned}\langle\sigma_X\rangle_{\mathcal{C}} = \langle\sigma_X\rangle_{\mathcal{C}\backslash B} \\ &+ \sum_{p\geq 0}\Big(\langle\sigma_B\rangle^{2p}_{\mathcal{C}\backslash B}\langle\sigma_X\sigma_B\rangle_{\mathcal{C}\backslash B} - \langle\sigma_B\rangle^{2p+1}_{\mathcal{C}\backslash B}\langle\sigma_X\sigma_B\rangle_{\mathcal{C}\backslash B} \\ &\qquad - \langle\sigma_B\rangle^{2p+1}_{\mathcal{C}\backslash B}\langle\sigma_X\rangle_{\mathcal{C}\backslash B} + \langle\sigma_B\rangle^{2p+2}_{\mathcal{C}\backslash B}\langle\sigma_X\rangle_{\mathcal{C}\backslash B}\Big) \qquad \text{(B.2)}\end{aligned}$$

Applying (3.2) to each term in the sum yields the four identities

$$\mathbb{E}_{l^n}[\langle\sigma_B\rangle^{2p}_{\mathcal{C}\backslash B}\langle\sigma_X\sigma_B\rangle_{\mathcal{C}\backslash B}] = \mathbb{E}_{l^n}[\langle\sigma_B\rangle^{2p}_{\mathcal{C}\backslash B}\langle\sigma_X\sigma_B\rangle^2_{\mathcal{C}\backslash B}]$$

$$\mathbb{E}_{l^n}[\langle\sigma_B\rangle^{2p+1}_{\mathcal{C}\backslash B}\langle\sigma_X\sigma_B\rangle_{\mathcal{C}\backslash B}] = \mathbb{E}_{l^n}[\langle\sigma_B\rangle^{2p+1}_{\mathcal{C}\backslash B}\langle\sigma_X\sigma_B\rangle_{\mathcal{C}\backslash B}\langle\sigma_X\rangle_{\mathcal{C}\backslash B}]$$

$$\mathbb{E}_{l^n}[\langle\sigma_B\rangle^{2p+1}_{\mathcal{C}\backslash B}\langle\sigma_X\rangle_{\mathcal{C}\backslash B}] = \mathbb{E}_{l^n}[\langle\sigma_B\rangle^{2p+1}_{\mathcal{C}\backslash B}\langle\sigma_X\rangle_{\mathcal{C}\backslash B}\langle\sigma_X\sigma_B\rangle_{\mathcal{C}\backslash B}]$$

$$\mathbb{E}_{l^n}[\langle\sigma_B\rangle^{2p+2}_{\mathcal{C}\backslash B}\langle\sigma_X\rangle_{\mathcal{C}\backslash B}] = \mathbb{E}_{l^n}[\langle\sigma_B\rangle^{2p+2}_{\mathcal{C}\backslash B}\langle\sigma_X\rangle^2_{\mathcal{C}\backslash B}]$$

Taking the expectation of (B.2) and using these four identities

$$\begin{aligned}\mathbb{E}_{l^n}[\langle\sigma_X\rangle_{\mathcal{C}}] = \mathbb{E}_{l^n}[\langle\sigma_X\rangle_{\mathcal{C}\backslash B}] \\ &+ \sum_{p\geq 0}\mathbb{E}_{l^n}\left[\langle\sigma_B\rangle^{2p}_{\mathcal{C}\backslash B}\Big(\langle\sigma_X\sigma_B\rangle_{\mathcal{C}\backslash B} - \langle\sigma_B\rangle_{\mathcal{C}\backslash B}\langle\sigma_X\rangle_{\mathcal{C}\backslash B}\Big)^2\right] \qquad \text{(B.3)}\end{aligned}$$

Thus

$$\mathbb{E}_{l^n}[\langle\sigma_X\rangle_{\mathcal{C}}] \geq \mathbb{E}_{l^n}[\langle\sigma_X\rangle_{\mathcal{C}\backslash B}] \qquad \text{(B.4)}$$

The case of general $m \geq 1$ can be dealt with the above method. However this is quite cumbersome and we prefer to adapt the technique of [11] which uses gaussian integration by parts. To this end we introduce a soft version of the check node constraints. Let us denote by $\langle - \rangle_{\mathcal{C},J}$ the Gibbs average corresponding to the Hamiltonian (2.2) where now J_A are independent gaussian random variables with

$$\mathbb{E}_J[J_A] = \mathbb{E}_J[J_A^2] - \mathbb{E}_J[J_A]^2 = t_A \qquad \text{(B.5)}$$

These Gibbs averages satisfy the same Nishimori identities than (3.2) and (see [16])

$$\mathbb{E}_{l^n}[\langle\sigma_X\rangle^m_{\mathcal{C}}] = \lim_{\{t_A\to+\infty, A\in\mathcal{C}\}}\mathbb{E}_{l^n,J}[\langle\sigma_X\rangle^m_{\mathcal{C},J}] \qquad \text{(B.6)}$$

From now on we work with the soft check node constraints and will use the $t_A + \infty$ limit to go back to the original case of interest.

The choice of a gaussian distribution with equal mean and variance for J_A is very convenient because of the identity

$$\frac{\partial}{\partial t_A}\frac{e^{-\frac{(J_A-t_A)^2}{2t_A}}}{2\pi t_A} = \big(-\frac{\partial}{\partial J_A} + \frac{1}{2}\frac{\partial^2}{\partial J_A^2}\big)\frac{e^{-\frac{(J_A-t_A)^2}{2t_A}}}{2\pi t_A} \tag{B.7}$$

and the integration by parts formula,

$$\frac{\partial}{\partial t_A}\mathbb{E}_{l^n,J}[\langle\sigma_X\rangle^m_{\mathcal{C},J}] = \mathbb{E}_{l^n,J}\big[\big(\frac{\partial}{\partial J_A} + \frac{1}{2}\frac{\partial^2}{\partial J_A^2}\big)\langle\sigma_X\rangle^m_{\mathcal{C},J}\big] \tag{B.8}$$

Straighforward algebra leads to

$$\begin{aligned}\frac{\partial}{\partial t_A}\mathbb{E}_{l^n,J}[\langle\sigma_X\rangle^m_{\mathcal{C},J}] = m\mathbb{E}_{l^n,J}\Bigg[\langle\sigma_X\rangle^{m-1}_{\mathcal{C},J}\Bigg(&\langle\sigma_X\sigma_A\rangle_{\mathcal{C},J} - \langle\sigma_X\rangle_{\mathcal{C},J}\langle\sigma_A\rangle_{\mathcal{C},J}\\ &- \langle\sigma_X\sigma_A\rangle_{\mathcal{C},J}\langle\sigma_B\rangle_{\mathcal{C},J} + \langle\sigma_X\rangle_{\mathcal{C},J}\langle\sigma_A\rangle^2_{\mathcal{C},J}\Bigg)\Bigg]\\ &+ \frac{1}{2}m(m-1)\mathbb{E}_{l^n,J}\Bigg[\langle\sigma_X\rangle^{m-2}_{\mathcal{C},J}\Bigg(\langle\sigma_X\sigma_A\rangle^2_{\mathcal{C},J}\\ &- 2\langle\sigma_X\sigma_A\rangle_{\mathcal{C},J}\langle\sigma_X\rangle_{\mathcal{C},J}\langle\sigma_A\rangle_{\mathcal{C},J} + \langle\sigma_X\rangle^2_{\mathcal{C},J}\langle\sigma_A\rangle^2_{\mathcal{C},J}\Bigg)\Bigg]\end{aligned} \tag{B.9}$$

The next step is to apply (3.2) to *all* terms of the above expression,

$$\begin{aligned}\frac{\partial}{\partial t_A}\mathbb{E}_{l^n,J}[\langle\sigma_X\rangle^m_{\mathcal{C},J}] = m\mathbb{E}_{l^n,J}\Bigg[\langle\sigma_X\rangle^{m-1}_{\mathcal{C},J}\Bigg(&\langle\sigma_X\sigma_A\rangle_{\mathcal{C},J}\langle\sigma_X^m\sigma_A\rangle_{\mathcal{C},J}\\ - \langle\sigma_X\rangle_{\mathcal{C},J}\langle\sigma_A\rangle_{\mathcal{C},J}\langle\sigma_X^m\sigma_A\rangle_{\mathcal{C},J} &- \langle\sigma_X\sigma_A\rangle_{\mathcal{C},J}\langle\sigma_A\rangle_{\mathcal{C},J}\langle\sigma_X^m\rangle_{\mathcal{C},J}\\ &+ \langle\sigma_X\rangle_{\mathcal{C},J}\langle\sigma_A\rangle^2_{\mathcal{C},J}\langle\sigma_X^m\rangle_{\mathcal{C},J}\Bigg)\Bigg]\\ + \frac{1}{2}m(m-1)\mathbb{E}_{l^n,J}\Bigg[\langle\sigma_X\rangle^{m-2}_{\mathcal{C},J}\langle\sigma_X^{m-2}\rangle_{\mathcal{C},J}\Bigg(&\langle\sigma_X\sigma_A\rangle^2_{\mathcal{C},J}\\ - 2\langle\sigma_X\sigma_A\rangle_{\mathcal{C},J}\langle\sigma_X\rangle_{\mathcal{C},J}\langle\sigma_A\rangle_{\mathcal{C},J} &+ \langle\sigma_X\rangle^2_{\mathcal{C},J}\langle\sigma_A\rangle^2_{\mathcal{C},J}\Bigg)\Bigg]\end{aligned} \tag{B.10}$$

We notice that for even m we have

$$\begin{aligned}\frac{\partial}{\partial t_A}\mathbb{E}_{l^n,J}[\langle\sigma_X\rangle^m_{\mathcal{C},J}] = \frac{1}{2}m(m-1)\mathbb{E}_{l^n,J}\Bigg[&\langle\sigma_X\rangle^{m-2}_{\mathcal{C},J}\\ &\times\Bigg(\langle\sigma_X\sigma_A\rangle_{\mathcal{C},J} - \langle\sigma_X\rangle_{\mathcal{C},J}\langle\sigma_A\rangle_{\mathcal{C},J}\Bigg)^2\Bigg]\end{aligned} \tag{B.11}$$

which is positive. On the other hand for odd m we have

$$\frac{\partial}{\partial t_A}\mathbb{E}_{l^n,J}[\langle\sigma_X\rangle_{\mathcal{C},J}^m] = \frac{1}{2}m(m+1)\mathbb{E}_{l^n,J}\Bigg[\langle\sigma_X\rangle_{\mathcal{C},J}^{m-1} \times \Big(\langle\sigma_X\sigma_A\rangle_{\mathcal{C},J} - \langle\sigma_X\rangle_{\mathcal{C},J}\langle\sigma_A\rangle_{\mathcal{C},J}\Big)^2\Bigg] \quad \text{(B.12)}$$

which is also positive. Thus for any m the average $\mathbb{E}_{l^n,J}[\langle\sigma_X\rangle_{\mathcal{C},J}]$ is an increasing function of t_A, for all A. Therefore for any given check node B, the limit of this quantity as $t_A \to +\infty$ for all $A \in \mathcal{C}$, is greater than the limit as $t_A \to +\infty$ for all $A \in \mathcal{C}\backslash B$ and $t_B = 0$. This is precisely the desired inequality.

C Proof of (6.7) and (6.8)

We begin with the correlation inequality (6.7). The method is the same than in appendix B. First we notice that

$$\langle\sigma_X\rangle_{\mathcal{C}} = \frac{\langle\sigma_X\rangle_{\mathcal{C},l_i=0} + \tanh\frac{l_i}{2}\langle\sigma_X\sigma_i\rangle_{\mathcal{C},l_i=0}}{1 + \tanh\frac{l_i}{2}\langle\sigma_i\rangle_{\mathcal{C},l_i=0}} \quad \text{(C.1)}$$

Expanding the denominator and grouping terms appropriately we get

$$\begin{aligned}\langle\sigma_X\rangle_{\mathcal{C}} &= \langle\sigma_X\rangle_{\mathcal{C},l_i=0} \\ &+ \sum_{p\geq 0}\Bigg(\big(\tanh\frac{l_i}{2}\big)^{2p+2}\langle\sigma_X\rangle_{\mathcal{C},l_i=0}\langle\sigma_i\rangle_{\mathcal{C},l_i=0}^{2p+2} - \big(\tanh\frac{l_i}{2}\big)^{2p+1}\langle\sigma_X\rangle_{\mathcal{C},l_i=0}\langle\sigma_i\rangle_{\mathcal{C},l_i=0}^{2p+1} \\ &+ \big(\tanh\frac{l_i}{2}\big)^{2p+1}\langle\sigma_X\sigma_i\rangle_{\mathcal{C},l_i=0}\langle\sigma_i\rangle_{\mathcal{C},l_i=0}^{2p} - \big(\tanh\frac{l_i}{2}\big)^{2p+2}\langle\sigma_X\sigma_i\rangle_{\mathcal{C},l_i=0}\langle\sigma_i\rangle_{\mathcal{C},l_i=0}^{2p+1}\Bigg)\end{aligned} \quad \text{(C.2)}$$

Applying the Nishimori identities to *all terms* in the sum we finaly obtain

$$\begin{aligned}\mathbb{E}_{l^n}[\langle\sigma_X\rangle_{\mathcal{C}}] = \langle\sigma_X\rangle_{\mathcal{C},l_i=0} + &\sum_{p\geq 0}\int_{-\infty}^{+\infty} dl_i c(l_i)\big(\tanh\frac{l_i}{2}\big)^{2p+2} \\ &\times \langle\sigma_i\rangle_{\mathcal{C},l_i=0}^{2p}\Big(\langle\sigma_X\rangle_{\mathcal{C},l_i=0}\langle\sigma_i\rangle_{\mathcal{C},l_i=0} - \langle\sigma_X\sigma_i\rangle_{\mathcal{C},l_i=0}\Big)^2\end{aligned} \quad \text{(C.3)}$$

Obviously the sum on the r.h.s is positive and we get (6.7). For (6.8) we have

$$\frac{d}{d\epsilon}\mathbb{E}_{l^n}[\langle\sigma_X\rangle_{\mathcal{C}}] = \sum_{i=1}^{n}\int_{-\infty}^{+\infty} dl_i \frac{dc(l_i)}{d\epsilon}\mathbb{E}_{l^n\backslash i}[\langle\sigma_X\rangle_{\mathcal{C}}] \tag{C.4}$$

Expanding and applying the Nishimori identities as above we obtain

$$\begin{aligned}\frac{d}{d\epsilon}\mathbb{E}_{l^n}[\langle\sigma_X\rangle_{\mathcal{C}}] = \sum_{i=1}^{n}\sum_{p\geq 0}\int_{-\infty}^{+\infty} dl_i \frac{dc(l_i)}{d\epsilon}\big(\tanh\frac{l_i}{2}\big)^{2p+2} \\ \times \langle\sigma_i\rangle_{\mathcal{C},l_i=0}^{2p}\Big(\langle\sigma_X\rangle_{\mathcal{C},l_i=0}\langle\sigma_i\rangle_{\mathcal{C},l_i=0} - \langle\sigma_X\sigma_i\rangle_{\mathcal{C},l_i=0}\Big)^2\end{aligned} \tag{C.5}$$

The result of the theorem follows because of (6.4).

References

[1] S.ten Brink, *Convergence behaviour of iteratively decoded parallel concatenated codes*, IEEE Transactions on Communications vol 49, no 10.pp.1727-1737 (2001).

[2] R. Urbanke, T. Richardson, in *Modern Coding Theory*, Cambridge University Press (in preparation).

[3] A. Ashikhmin, G. Kramer and S. ten Brink, *Code rate and the area under extrinsic information transfer curves*, Proc.of IEEE Int. Symp. Inf. Theory, Lausanne, Switzerland, June 2002, p 115.

[4] A. Montanari, *The glassy phase of Gallager codes*, European Physical Journal, **23** (2001).

[5] C. Méasson and R. Urbanke, *An upper-bound for the ML threshold of iterative coding systems over the BEC*, Proc. of the 41st Allerton Conference on Communications, Control and Computing, Allerton House, Monticello, USA, October 2003 p.3.

[6] C. Méasson, A. Montanari, T. Richardson and R.Urbanke, *Life Above Threshold: From List Decoding to Area Theorem and MSE*, IEEE Information Theory Workshop, San Antonio October 2004.

[7] C. Méasson, A. Montanari,and R.Urbanke, *Maxwell Construction: The Hidden Bridge between Iterative and Maximum a Posteriori Decoding*, submitted to IEEE Trans. Inf. Theory, 2005.

[8] C. Méasson, A. Montanari,and R.Urbanke, Proc of IEEE Int. Symp Inf. Theory, Adelaide, Australia (2005)

[9] C. Méasson, A. Montanari,and R.Urbanke, *The generalized area theorem and some of its consequences*, submitted to trans. inf. Theory (2005)

[10] R. B. Griffiths, in *Phase Transitions and Critical Phenomena*, vol 1 eds. C. Domb and M. S. Green, Academic Press (1972).

[11] S. Morita, H. Nishimori, P. Contucci, *Griffiths inequalities for the Gaussian spin glass*, J. Phys. A **37** L203 (2004).

[12] A. Montanari, *Tight bounds for LDPC and LDGM codes under MAP decoding*, IEEE Trans. Inf. theory, **51** 3221 - 3246 (2005)

[13] F. Guerra, F. Toninelli, *Quadratic replica coupling in the Sherrington-Kirkpatrick mean field spin glass model*, J. Math. Phys **43**, p 3704 (2002)

[14] S. Franz, M. Leone, *Replica bounds for optimisation problems and diluted spin systems*, J. Stat. Phys. **111** (2003) 535-564.

[15] N. Macris, *Correlation inequalities: a useful tool in the theory of LDPC codes*, *Proc. IEEE Int. Symp. Inf. Theory* Adelaide, Australia (2005)

[16] N. Macris, *Griffiths-Kelly-Sherman correlation inequalities: a useful tool in the theory of error correcting codes*, IEEE Trans. Inf. Theory Volume 53, Issue 2, Feb. 2007 Page(s):664 - 683

[17] H. Nishimori, in *Statistical Physics of Spin Glasses and Information Processing: An Introduction*, Oxford Science Publications (2001).

[18] D. Guo, S. Shamai and S. Verdu, *Mutual information and MMSE in Gaussian channels*, Proc. 2004 ISIT, Chicago, IL, USA, p. 347.

[19] P.M. Chaikin, T.C. Lubensky, *Principles of Condensed Matter Physics*, Cambridge University Press (2000), Chap 8

[20] N. Macris, *On the relation between MAP and BP GEXIT functions of low density parity check codes*, Proc. Information Theory Workshop IEEE, Uruguay 2006, p 312-316

[21] S. Kudekar, N. Macris, *Sharp Bounds for MAP Decoding of General Irregular LDPC Codes* Proc. International Symposium on Information Theory, Seattle July 2006 Page(s) 2259 - 2263