Exact solution for the conditional entropy of Poissonian LDPC codes over the Binary Erasure Channel

Satish Babu KORADA, Shrinivas KUDEKAR and Nicolas MACRIS
Ecole Polytechnique Fédérale de Lausanne
School of Computer and Communication Science
EPFL, I&C, LTHC, Station 14
Lausanne CH-1015, Switzerland

***Abstract*— We consider communication over a binary erasure channel with low density parity check codes and optimal maximum a posteriori decoding. It is known that the problem of computing the average conditional entropy, over such code ensembles, in the asymptotic limit of large block length is closely related to computing the free energy of a mean field spin glass in the thermodynamic limit. Tentative, but explicit, formulas for these quantities have been derived thanks to the replica method (of spin glass theory) and are generally conjectured to be exact. In this contribution we show that the replica formulas are indeed exact in the case of Poissonian low density parity check ensembles. Our methods use ideas coming from the recent progress in the rigorous analysis of the Sherrington-Kirkpatrick model and their applications to the theory of error correcting codes.**

I. Introduction

Linear codes based on sparse random graphs have been very successful because of low-complexity decoding schemes and good performance [1]. Among the most popular code constructions one finds the low density parity check LDPC ensembles and it is conjectured that important quantities such as the average (over the code ensemble) conditional entropy per bit $h_n = n^{-1}H(X^n|Y^n)$ of the transmitted message X^n conditional to the received message Y^n can be computed exactly in the limit of large block length. This would give us exactly the noise threshold beyond which error free communication is not possible under optimal decoding. Even more, it is known that there is an intimate relationship between optimal maximum a posteriori (MAP) decoding and message passing decoding using belief propagation. Namely the so called generalized EXIT curves associated to the two decoders are thought to be equal beyond the MAP threshold.

Most of the theory has been so far developed for the binary erasure channel by using *combinatorial* methods. For example $\lim_{n\to+\infty}\mathbb{E}_{\mathcal{C}}[h_n]$ has been rigorously computed very recently [2] for a class of LDPC ensembles by computing the rate corresponding to the residual graphs left over after the completion of iterative decoding. In this contribution we use *non-combinatorial* methods coming from the rigorous analysis of mean field spin glasses to derive such a result for the special class of Poissonian LDPC ensemble. We believe that it will be possible to extend our proofs beyond the binary erasure channel because of their non-combinatorial nature. Our method uses the *two interpolations* first developed by Guerra and Toninelli [4] in the context of the Sherrington-Kirkpatrick model [8] of a mean field spin glass. The *first interpolation* has been adapted to diluted spin systems by Franz and Leone [5] and to error correcting codes by Montanari [6] for LDPC(n,Λ,P) ensembles[1] with any polynomial $\Lambda(x)$ but $P(x)$ restricted to be a convex polynomial in a region $-e \le x \le e$. The net result of the first interpolation is a *lower bound* on the conditional entropy which coincides with the "replica formula" and is thus conjectured to be tight. Note that the proof of this bound works for any memoryless binary-input output-symmetric channel. For some channels (BEC, BIAWGNC and BSC) the convexity requirement on $P(x)$ has been relaxed in [12]. For the Sherrington-Kirkpatrick model the *second interpolation* leads to an *upper bound* which coincides with the lower bound at least in the high temperature region of the phase diagram[2]. In the case of LDPC ensembles the second interpolation has not yet been developed. This is in essence what we do in this contribution for the simplest case of Poisson LDPC ensembles and the BEC. Part of the mathematical analysis involved in the second interpolation is reminiscent of the one we have developed recently for the (simpler) case of a "gauge symmetric p-spin model" [11].

II. Main result

In the sequel we consider communication through a BEC with transition probability $p_{Y|X}(y|x)$ and noise parameter ϵ. We consider Poisson LDPC(n,Λ,P) ensemble. The number of check nodes is a Poisson integer with mean $n\frac{\gamma}{P'(1)}$ and they are connected uniformly at random with n variable nodes. Here $\Lambda(x) = e^{\gamma(x-1)}$ and hence we denote this ensemble also as LDPC(n,γ,P). The design rate of this ensemble is given

[1]Here $\Lambda(x) = \sum_d \Lambda_d x^d$, $P(x) = \sum_k P_k x^k$ are the variable and check node degree distributions from the node perspective

[2]Thus in this case the replica solution is confirmed. Remarkably this has been extended to the whole phase diagram by Talagrand [7] thus proving the Parisi formula

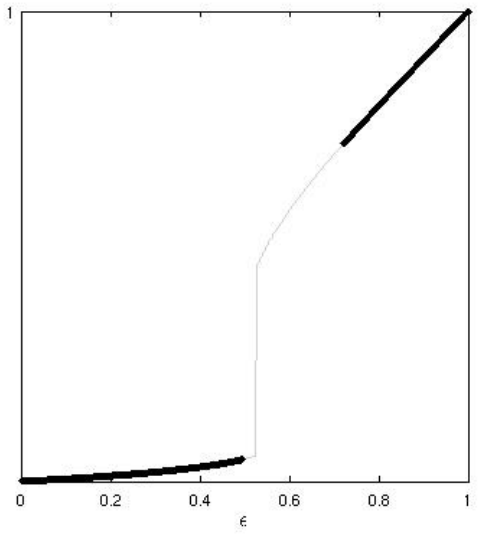

Fig. 1. The thin line is p_{RS} obtained from the replica solution for $\gamma = 3, P(x) = x^6$ and the thick line is the region where we prove its exactness

by $1 - \frac{\gamma}{P'(1)}$. We use the notation $\rho(x) = P'(x)/P'(1)$ and $\lambda(x) = \Lambda'(x)/\Lambda'(1)$.

The mathematically ill defined *replica calculations* of spin glass theory lead to a conjectured formula for the entropy which has the form of a variational problem $\sup_{d_V} h_{RS}[d_V]$ over a class of probability distributions satisfying[3] $d_V(v) = e^{2v} d_V(-v)$ where the functional $h_{RS}[d_V]$ is known as the *replica symmetric* entropy. For the BEC the probability distributions are of the form $d_V(v) = p\delta_0(v) + (1-p)\delta_\infty(v)$ with $0 \le p \le 1$ and

$$\begin{aligned}\frac{h_{RS}[p]}{\ln 2} =& \Lambda'(1)p\rho(1-p) + \epsilon\Lambda(1-\rho(1-p)) \\ &- \frac{\Lambda'(1)}{P'(1)}(1 - P(1-p))\end{aligned}$$

Hence the variational problem can be explicitly solved. The maximizer of $h_{RS}[p]$ is a solution of the stationary point equation $p = \epsilon\lambda(1-\rho(1-p))$, and we call the unique maximizer p_{RS}. Typically this is a curve as a function of ϵ with one or many discontinuities (see figure 1 for an example with one discontinuity). One of the results of this work (and of [2]) is that this curve is nothing else but the MAP probability of error.

When the suboptimal belief propagation (BP) decoder is used one can compute the corresponding BP probability of error by the method of density evolution. This method leads to the same stationary point equation which is solved iteratively starting from the initial condition $p = \epsilon$. The BP probability of error is the largest fixed point obtained from this initial condition and we will call it p_{BP}. This is again a curve as a function of ϵ with one or many discontinuities, which is closely related to p_{RS} (see the example of figure 1). For example it is possible to check that the two curves are identical for all ϵ above the last discontinuity of p_{RS} .

In order to state our theorem below we need to define an auxiliary function

$$\begin{aligned}f(z) =& \frac{\gamma}{P'(1)}\big(P(z) - zP'(z)\big) \\ &+ (1-p_{RS})\frac{\gamma}{P'(1)}\big(P'(z) - P'(1-p_{RS})\big) \\ &+ p_{RS}\ln\cosh\big[\frac{\gamma}{P'(1)}\big(P'(z) - P'(1-p_{RS})\big)\big]\end{aligned}$$

Theorem 1: Assume communication using a Poisson LDPC(n, γ, P) code ensemble, through a BEC with erasure probability ϵ. Assume that f has a unique maximizer $\hat{z}$. Then for all ϵ such that $p_{RS} = p_{BP}$ and $\epsilon \in C_{P,\gamma} = \{\epsilon \in [0,1] | \hat{z} = 1 - p_{RS}\}$ we have the exact expression for the average per bit conditional entropy

$$\lim_{n\to+\infty} \mathbb{E}_{\mathcal{C}}[h_n] = h_{RS}[p_{RS}]$$

The condition over the range of ϵ is not optimal and comes from the second interpolation that we use. One may check explicitly that $f'(\bar{p}_{RS}) = 0$, so that $\hat{z} = \bar{p}_{RS}$ is always a critical point. Furthermore, for some range of ϵ it is a global maximum and this range intersects the one where $p_{RS} = p_{BP}$. In [2] there is also a condition which is different from ours. A consequence of the theorem is that (at least over some range of ϵ) the MAP probability of error is given by p_{RS} which is computable from the replica formulas. One way to see this is to check that $\epsilon\frac{d}{d\epsilon}h_{RS}[p_{RS}] = p_{RS}\ln 2$.

III. Statistical mechanical formulation and first interpolation

The Tanner graph has variable nodes, denoted by i that are connected to check nodes denoted by c. We will work in terms of the *half*-loglikelihood variables $l = \frac{1}{2}\ln\frac{p_{Y|X}(y|1)}{p_{Y|X}(y|0)}$ and call their (ϵ dependent) distribution $c(l)$ assuming that the all zero codeword is transmitted.

The *posterior distribution* $p_{X^n|Y^n}(x^n|y^n)$ used in MAP decoding can be viewed as the Gibbs measure of a random spin system. For this it is convenient to use the mapping of bits to spins $\sigma_i = (-1)^{x_i}$. For a uniform prior over the code words and a memoryless binary-input output-symmetric channel, Bayes rule implies

$$p_{X^n|Y^n}(x^n|y^n) = \frac{1}{Z_{\mathcal{C}}}\prod_{c\in\mathcal{C}}\frac{1}{2}(1+\sigma_{\partial c})\prod_{i=1}^{n}\frac{e^{l_i\sigma_i}}{2\cosh l_i}$$

where $\sigma_{\partial c} = \prod_{i\in c}\sigma_i$ and $Z_{\mathcal{C}}$ is the normalization factor or partition function. Expectations with respect to the Gibbs measure for a fixed graph and a fixed channel output are denoted by the bracket $\langle - \rangle$. More precisely for any $X \subset \{1, ..., n\}$, $\langle \sigma_X \rangle = \sum_{\sigma^n} \sigma_X \mu_{\mathcal{C}}(\sigma^n)$ where $\sigma_X = \prod_{i\in X}\sigma_i$. Expectations with respect to the code ensemble and the channel outputs will be denoted by $\mathbb{E}_{\mathcal{C},l^n}[-]$.

[3] In coding this "symmetry" comes from channel symmetry and in statistical physics it is a special instance of gauge symmetry.

It is possible to show [9] in the case of BEC

$$\mathbb{E}_{\mathcal{C}}[h_n] = n^{-1}\mathbb{E}_{\mathcal{C},l^n}[\ln Z_{\mathcal{C}}] + \epsilon \ln 2$$

The quantity $n^{-1}\mathbb{E}_{\mathcal{C},l^n}[\ln Z_{\mathcal{C}}]$ is known as the average *free energy* in statistical mechanics. Since this differs from the average conditional entropy only by a constant, we will focus on the evaluation of the free energy.

A. First interpolation

The main idea behind the interpolation technique is to recursively remove the check node constraints and compensate for the change of rate with "extra observations" U_a. These are independent identically distributed random variables constructed as follows. Let V be a random variable with a *symmetric* density $d_V(v)$ (i.e $d_V(v) = e^{2v}d_V(-v)$). Here we deal with the BEC so it is sufficient to look at the space of distributions $d_V(v) = p\delta_0(v) + (1-p)\delta_\infty(v)$ where $0 \le p \le 1$ is for the moment an arbitrary parameter. Set $U_a = \tanh^{-1}\left[\prod_{i=1}^{k-1}\tanh V_i\right]$ where V_i are i.i.d copies of V and k is an integer distributed as P_k. Notice that this equation mimics a check node message in the belief propagation decoding algorithm.

Let $t \in [0,1]$ be an interpolating parameter and consider the Tanner graphs $\mathcal{C}_t$ from the ensemble LDPC$(n, \gamma t, P)$. Thus at "time" t the number of check nodes is a Poisson r.v with mean $n\gamma t$. As said before the loss of check nodes is compensated by "extra observations". Variable nodes i receive d_i i.i.d copies $\{U_a^i\}$, $a = 1, ..., d_i$ of the r.v U_a. Here d_i are i.i.d copies of a random Poisson integer with mean $n\gamma(1-t)$. The interpolating Gibbs measure is

$$\frac{1}{Z(t)}\prod_{c\in\mathcal{C}_t}\frac{1}{2}(1+\sigma_{\partial c})\prod_{i=1}^{n}\frac{e^{(l_i+\sum_{a=1}^{d_i}u_a^i)\sigma_i}}{2\cosh l_i\prod_{a=1}^{d_i}2\cosh u_a^i}$$

Expectations with respect to this measure will be denoted by $\langle - \rangle_t$ and for the corresponding average free energy we set $\alpha_n(t) = n^{-1}\mathbb{E}_{\mathcal{C}_t,l^n,\{u_a^i\}}[\ln Z(t)]$. At $t = 1$ one recovers the original free energy $\alpha_n(1) = n^{-1}\mathbb{E}_{\mathcal{C},l^n}[\ln Z_{\mathcal{C}}]$ while at $t = 0$ we have a simple product measure which is tailored to yield the replica symmetric free energy (or up to a constant term $h_{RS}[p]$).

In order to lighten the formulas *from now on we consider the case* $P(x) = x^r$ but it is straightforward to extend the arguments to general polynomials. Also, we use the simplified notation $\mathbb{E}_{\mathcal{C}_t,l^n,\{u_a^i\}} = \mathbb{E}_t$ and $\bar{p} = 1 - p$.

From [6] we have that $\alpha_n(1)$ can be written as,

$$\alpha_n(1) = h_{RS}[p] + \int_0^1 dt\mathcal{R}(t)$$

$$\mathcal{R}(t) = \frac{\gamma}{r}\sum_{l\ge1}\frac{(-1)^{l+1}}{l}\mathbb{E}_t[\langle R(\bar{p}, Q_l)\rangle_t] \qquad (1)$$

where $Q_l = n^{-1}\sum_i \sigma_i^1\sigma_i^2\cdots\sigma_i^l$ are called the overlap parameters and $R(a,b) = (r-1)a^r - ra^{r-1}b + b^r$. One crucial property of the later polynomial is that it is positive for all r for $a \ge 0$, $b \ge 0$.

There are two major simplifications that we can make on $\mathcal{R}(t)$. First it was shown in [12] that for almost every ϵ

$$\mathcal{R}(t) = \frac{\gamma}{r}\sum_{l\ge1}\frac{(-1)^{l+1}}{l}\mathbb{E}_t[R(\bar{p}, \langle Q_l\rangle_t)] + o_n(1) \qquad (2)$$

Note that in equation (1), $\langle R(\bar{p}, Q_l)\rangle_t$ involves the three quantities $\bar{p}, \langle Q_l^r\rangle_t, \langle Q_l\rangle_t$ whereas in equation (2), $R(\bar{p}, \langle Q_l\rangle_t)$ is a polynomial in two variables $\bar{p}, \langle Q_l\rangle_t$. Furthermore, for the BEC, we either receive a bit perfectly or it is erased. This is the content of the lemma which can be given a formal proof.

Lemma 1: For the BEC the random variables $\langle\sigma_i\rangle_t$ take values 0 or 1.

Proof: From the GKS inequality [10] we have $\langle\sigma_i\rangle_t \ge 0$. Moreover from channel (or Nishimori) symmetry we have $\mathbb{E}_t[\langle\sigma_i\rangle_t] = \mathbb{E}_t[\langle\sigma_i\rangle_t^2]$. Thus $\langle\sigma_i\rangle_t(1 - \langle\sigma_i\rangle_t)$ is a positive random variable with zero mean, even for n finite. Thus it is zero for all the graph and noise realizations. ■

The lemma implies

$$\langle Q_l\rangle_t = \frac{1}{n}\sum_{i=1}^{n}\langle\sigma_i\rangle_t^l = \frac{1}{n}\sum_{i=1}^{n}\langle\sigma_i\rangle_t = \langle m\rangle_t$$

where $m = \frac{1}{n}\sum_{i=1}^{n}\sigma_i$ is the "total magnetization", which leads to

$$\alpha_n(1) = h_{RS}[p] + \ln 2\frac{\gamma}{r}\int_0^1 dt\mathbb{E}_t[R(\bar{p}, \langle m\rangle_t)] + o_n(1)$$

In particular this equality implies the lower bound $\mathbb{E}_{\mathcal{C}}[h_n] \ge h_{RS}[p_{RS}] + o_n(1)$ for all r and almost every ϵ.

B. Belief propagation for the interpolating system

It will prove useful to collect here a few properties of the interpolating system at time t. It can be thought of as a communication system where code words from $\mathcal{C}_t \in$ LDPC$(n, \gamma t, P)$ are sent via a BEC. The receiver also collects "extra observations" U distributed as $d_U(u) = (1 - \rho(\bar{p}))\delta_0(u) + \rho(\bar{p})\delta_\infty(u)$ Alternatively one can view the system as codewords from $\mathcal{C}_t$ transmitted through a channel with effective erasure probability $\epsilon\lambda_{1-t}(1 - \rho(\bar{p}))$. We have the following recursive equation for the density evolution (of erasures) analysis of the BP decoding (at iteration ℓ)

$$x_{l+1,t} = \epsilon\lambda_{1-t}(1 - \rho(\bar{p}))\lambda_t(1 - \rho(1 - x_{l,t})) \qquad (3)$$

where $\lambda_t(x) = e^{\gamma t(x-1)}$. By explicit analysis of (3) we can show

Lemma 2: For any δ small enough, if we set $p = p_{BP} + \delta$ then the fixed point of the recursion (3) obtained from the initial condition $x_{0,t} = \epsilon$ satisfies $x_{\infty,t} = p_{BP} + O(\delta)$.

Now we consider the *BP estimate* of the spin (or bit) σ_i after l iterations for the interpolated code. It is possible to regard the BP estimate as a statistical mechanical average on a *computational tree* $\mathcal{T}_i(\ell)$ [1] of depth ℓ for node i. One simply considers the Gibbs measure with appropriate check node

constraints and observations associated to all nodes appearing in the labeled tree graph $\mathcal{T}_i(\ell)$. We denote this average by $\langle\sigma_i\rangle_{t,\mathcal{T}_i(\ell)}$.

Lemma 3: For any t, and any $\delta > 0$, one can find a depth $\ell(\delta)$ and a block length $n(\delta)$ independent of t such that for all $n \geq n(\delta)$, if $p = p_{BP} + \delta$, we have $\mathbb{E}_t[\langle\sigma_i\rangle_{t,\mathcal{T}_i(\ell)}] \geq \bar{p}_{BP} - g(\delta)$, where $0 < g(\delta) < \delta$.

Proof: On the Tanner graph $\mathcal{C}_t$, let $\mathcal{N}_i(\ell)$ denote the neighborhood of depth l for node i. With high probability this is a tree so that one easily computes $\mathbb{E}_t[\langle\sigma_i\rangle_{t,\mathcal{T}_i(\ell)}]$ by (3). The inequality then follows from lemma 2. ∎

Finally we will need a concentration property for the BP estimate of the interpolating system.

Lemma 4: One can find a numerical constant $\beta > 0$ such that for any δ small enough, any fixed ℓ, if n is large enough

$$\mathbb{P}_t\Big[|\sum_{i=1}^{n}\langle\sigma_i\rangle_{t,\mathcal{T}_i(\ell)} - \sum_{i=1}^{n}\mathbb{E}_t\langle\sigma_i\rangle_{t,\mathcal{T}_i(\ell)}| \geq n\delta\Big] \leq e^{-\frac{n\beta\delta^2}{(\ln n)^{2\ell}}} + n^{1-\frac{1}{2}\ln(\ln n)}$$

Proof: Adapt the proof of [3]. ∎

IV. Upper bound on the remainder $\mathcal{R}(t)$

To prove the exactness of RS solution it remains to show that $\mathbb{E}_t\left[R(\bar{p}_{RS}, \langle m\rangle_t)\right] \leq o_n(1)$. The idea is that for $p = p_{RS}$ the removal of the check nodes is perfectly compensated by the addition of the cavity biases and $\mathbb{E}_t\left[R(\bar{p}_{RS}, \langle m\rangle_t)\right]$ does not change with t and hence the remainder term is 0 (since at $t = 0$ one can explicitly verify this). We are unable to show directly this perfect compensation, but we can show that for any $\delta > 0$ and $p = p_\delta = p_{RS} + \delta$ the compensation is almost perfect in the sense that $\mathbb{E}_t\left[R(\bar{p}_\delta, \langle m\rangle_t)\right] \leq o_n(1) + O(\delta)$. Therefore,

$$h_{RS}[p_{RS}] \leq \lim_{n\to\infty}\mathbb{E}_{\mathcal{C}}[h_n] \leq h_{RS}[p_{RS} + \delta] + O(\delta)$$

Since δ can be made as small as desired, using the continuity of $h_{RS}[p]$ we get the equality of the theorem.

Notice that we need $p = p_{BP} + \delta$ in lemmas 2, 3. Thus we see that for the present proof to work ϵ has to be in the range $p_{RS} = p_{BP}$. This is the first condition appearing in the theorem.

A. Second Interpolation

In fact from [12] it is enough to show that $\mathbb{E}_t\left[\langle R(\bar{p}_\delta, m)\rangle_t\right] \leq o_n(1) + O(\delta)$. We will use a second interpolation in a similar fashion than in [11]. Consider the following partition function,

$$Z(t,\mu,\bar{p}) = \sum_{\sigma^n} e^{\mu n R(\bar{p},m)} \prod_{c\in\mathcal{C}_t}\big(\frac{1+\sigma_{\partial c}}{2}\big) \prod_{i=1}^{n}\frac{e^{(h_i+\sum_{a=1}^{d_i}u_a^i)\sigma_i}}{2\cosh(h_i)\prod_{a=1}^{d_i}2\cosh(u_a^i)}$$

We denote $\langle\cdot\rangle_{t,\mu}$ the average associated to this partition function and $\alpha_n(t,\mu,\bar{p}) = n^{-1}\ln Z(t,\mu,\bar{p})$ the associated free energy. The motivation for defining $Z(t,\mu,\bar{p})$ is to use the fact that the fluctuations of $R(\bar{p}_{RS}, m)$ are very small and the free energy $\alpha_n(t,\mu,\bar{p}_{RS})$ is close to $\alpha_n(t,0,\bar{p}_{RS})$.

The convexity of $\alpha_n(t,\mu,\bar{p})$ with respect to μ implies for any $\mu > 0$,

$$\begin{aligned}\mathbb{E}_t[\langle R(\bar{p},m)\rangle_t] &= \frac{\partial}{\partial\mu}\mathbb{E}_t[\alpha_n(t,\mu,\bar{p})]\big|_{\mu=0} \\ &\leq \frac{1}{\mu}\mathbb{E}_t[\alpha_n(t,\mu,\bar{p}) - \alpha_n(t,0,\bar{p})]\end{aligned}$$

It turns out that the good choice for μ is $\mu(t) = \frac{\gamma}{r}(1-t)$. Note that with this choice a priori the t-integral of the remainder term diverges. But it is easy to circumvent this problem by splitting the integral in two intervals $[0, 1-\delta]$ and $[1-\delta, 1]$. The second interval easily leads to a contribution $O(\delta)$ and for the first one the rest of the proof will yield a contribution of the form $o_n(1) + O(\delta^2\ln\delta)$.

Using the fundamental theorem of calculus, some algebra, and positivity (of part of the remainders as in the first interpolation), we finally obtain the estimate

$$\mathbb{E}_t[\alpha_n(t,\mu(t),\bar{p}) - \alpha_n(t,0,\bar{p})] \leq \Delta_n\big(\frac{\gamma}{r},\bar{p}\big) \tag{4}$$

$$+\frac{\gamma}{r}\int_0^t\sum_{l\geq 2}\frac{(-1)^{l+1}}{l}\mathbb{E}_s\langle R(\bar{p},Q_l)\rangle_{s,\mu(s)}ds \tag{5}$$

where

$$\Delta_n\big(\frac{\gamma}{r},\bar{p}\big) = \mathbb{E}[\alpha_n(0,\frac{\gamma}{r},\bar{p}) - \alpha_n(0,0,\bar{p})]$$

Remark that the term $l = 1$ in the sum has canceled due to the judicious choice of $\mu(t)$.

B. Estimate of (4)

One can explicitly compute (4). Indeed the free energy $\alpha_n(0,0,\bar{p})$ corresponds to a Gibbs measure with a product form. For the other free energy $\alpha_n(0,\frac{\gamma}{r},\bar{p})$ the situation is more complicated but the code $\mathcal{C}_t$ is absent and the problem is similar to the computation of a free energy of a non-random complete p-spin model. This can be computed by the saddle point methods much like in [11]. The net result of this long calculation is

$$\Delta_n\big(\frac{\gamma}{r},\bar{p}_\delta\big) = \frac{\gamma}{r}(r-1)\bar{p}_{RS}^r + \max_z f(z) + o_n(1)$$

where $f(z)$ was defined in the second section. Thus for ϵ such that the maximizer $\hat{z} = \bar{p}_{RS}$, this contribution vanishes as $n \to +\infty$.

C. Estimate of (5)

It is difficult to decide what is the sign of (3) because unlike the case $\mu = 0$ we do not have tools such as the GKS inequality [10] or Nishimori symmetry. This is why we establish a relation between the $\mu \neq 0$ system and $\mu = 0$ system through the following two lemmas the (easy) proofs of which we omit here.

Lemma 5: If $\langle \sigma_i \rangle_t = 1$ then also $\langle \sigma_i \rangle_{t,\mu} = 1$

Lemma 6: For $P(x) = x^r$ we have $\langle \sigma_0 \rangle_{t,\mu} \geq -1 + e^{-4n\mu r}$.

Thanks to these we can show the crucial result

Lemma 7: Let ϵ such that $p_{RS} = p_{BP}$. For any $\delta > 0$ there exists $n(\delta)$ such that for all $n \geq n(\delta)$,

$$\sum_{l \geq 2} \frac{(-1)^{l+1}}{l} \mathbb{E}_s \langle R(\bar{p}_\delta, Q_l) \rangle_{s,\mu(s)} \leq 0$$

Proof: Define the random variable $q = \langle m \rangle_s$ and the random set of variable nodes $S = \{i : \langle \sigma_i \rangle_t = 0\}$. From lemma 5, $Q_l = q + \frac{1}{n} \sum_{i \in S} \sigma_i^{(1)} ... \sigma_i^{(l)}$ which implies

$$\begin{aligned} &\langle R(\bar{p}_\delta, Q_l) \rangle_{s,\mu(s)} \\ &\quad = \sum_{j=0}^{r} \binom{r}{j} q^{r-j} \frac{1}{n^j} \sum_{i_1, ... i_j \in S} \langle \sigma_{i_1} ... \sigma_{i_j} \rangle_{s,\mu}^l \\ &\quad - r\bar{p}_\delta^{r-1} q - r\bar{p}_\delta^{r-1} \frac{1}{n} \sum_{i \in S} \langle \sigma_i \rangle_{s,\mu(s)}^l + (r-1)\bar{p}_\delta^r \end{aligned}$$

Using $\langle \sigma_X \rangle_{s,\mu}^{2k} \geq \langle \sigma_X \rangle_{s,\mu}^{2k+1}$ and $\langle \sigma_X \rangle_{s,\mu}^{2k} \geq 0$ the $j \geq 2$ part of the j-sum yields a negative contribution when we combine $l = 2k$ and $l = 2k+1$ as follows

$$\begin{aligned} &-\frac{1}{2k} \mathbb{E}_s \langle R(\bar{p}_\delta, Q_{2k}) \rangle_{s,\mu} + \frac{1}{2k+1} \mathbb{E}_s \langle R(\bar{p}_\delta, Q_{2k+1}) \rangle_{s,\mu} \\ &\quad \leq -\frac{1}{2k(2k+1)} \mathbb{E}_s [R(q, \bar{p}_\delta)] \\ &\quad - \frac{1}{2k} \mathbb{E}_s [\mathcal{T}_{2k}] + \frac{1}{2k+1} \mathbb{E}_s [\mathcal{T}_{2k+1}] \end{aligned}$$

where

$$\mathcal{T}_l = r(q^{r-1} - \bar{p}_\delta^{r-1}) \frac{1}{n} \sum_{i \in S} \langle \sigma_i \rangle_{s,\mu(s)}^l$$

Let $\mathcal{A}_s$ be the event $\{q \geq \bar{p}_{BP} - \frac{\delta}{2} - \frac{g(\delta)}{2}\}$. Since in the set $\mathcal{A}_s$ $\mathcal{T}_{2k} \geq \mathcal{T}_{2k+1}$, we have

$$\begin{aligned} &-\frac{1}{2k} \mathbb{E}_s \langle R(\bar{p}_\delta, Q_{2k}) \rangle_{s,\mu} + \frac{1}{2k+1} \mathbb{E}_s \langle R(\bar{p}_\delta, Q_{2k+1}) \rangle_{s,\mu} \\ &\leq \mathbb{E}_{\mathcal{A}_s^c} \left[\frac{-\mathcal{T}_{2k}}{2k} + \frac{\mathcal{T}_{2k+1}}{2k+1} \right] - \frac{r(r-1)\delta^2}{2k(2k+1)} \bar{p}_{BP}^{r-2} \mathbb{P}[\mathcal{A}_s] \end{aligned}$$

This yields a negative contribution for any fixed δ and n large enough provided that we show $\mathbb{P}[\mathcal{A}_s^c] \to 0$ fast enough (in fact one has also to use lemma 6 to control the k-sum which yields a contribution $\ln(1 + \langle \sigma_i \rangle_{s,\mu})$).

To bound the probability of the event $\mathcal{A}_s^c$ we note that for any realization of the randomness $\langle \sigma_i \rangle_s \geq \langle \sigma_i \rangle_{s,\mathcal{T}_i(l)}$. Indeed if the iterative decoder succeeds and the BP estimate is 1 then the MAP estimate is also necessarily 1. On the other hand if the iterative decoder fails then the BP estimate is 0 and is surely less than the MAP estimate which is 0 or 1. Therefore lemma 3 implies that there is a $n(\delta)$ such that for $n \geq n(\delta)$,

$$\begin{aligned} \mathbb{P}[\mathcal{A}_s^c] \leq \mathbb{P}\Big[\sum_i \langle \sigma_i \rangle_{s,\mathcal{T}_i(l)} \leq \sum_i \mathbb{E}_s \langle \sigma_i \rangle_{s,\mathcal{T}_i(l)} \\ - \frac{n}{2}(\delta - g(\delta)) \Big] \end{aligned}$$

and the result follows from lemma 4. ∎

V. Conclusions

Similar results can be obtained for the Poisson LDGM ensembles. The most important open problems are the extension of the present methods to any standard irregular LDPC ensembles and other binary memoryless symmetric channels. Because of the non-combinatorial nature of the proof it should be possible to go beyond the BEC. The main problem is to show that lemma 7 holds for other channels as well.

Acknowledgement

The work of S. Kudekar has been supported by a grant from the Swiss National Science Foundation number 200020-113412. The work of S. Korada is supported by NCCR-MICS, a center supported by the Swiss National Science Foundation under grant number 5005-67322.

References

[1] T. Richardson, R. Urbanke "Modern Coding Theory," *Cambridge University Press*, in preparation.

[2] C. Meason, A. Montanari, R. Urbanke "Maxwell Construction: The Hidden Bridge between Iterative and Maximum a Posteriori Decoding", preprint 2005 *arxiv:cs.IT/0506083*

[3] T. Richardson, R. Urbanke "The Capacity of LDPC codes under Message-Passing Decoding", *IEEE Trans. Inf. Theory.*, pp. 638–656, 2001.

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

[5] S. Franz, M. Leone "Replica Bounds for Optimization Problems and Diluted Spin Systems," *J. Stat. Phys.*, **111** p. 535-564 (2003).

[6] A. Montanari, "Tight Bounds for LDPC and LDGM Codes Under MAP Decoding," *IEEE Trans. Inf. Theory.*, **51**, no. 9, pp. 3221–3246, (2005).

[7] M. Talagrand, "The Parisi formula", Annals of Mathematics, **163** (2006), 221-263

[8] M. Mezard, G. Parisi, M. Virasoro, "Spin Glass Theory and Beyond", (1987) World Scientific

[9] A. Montanari, "The glassy phase of Gallager codes", *European Phys. Journal*, **23** 2001.

[10] N. Macris, "Griffith-Kelly-Sherman Correlation Inequalities: A Useful Tool in the Theory of Error Correcting Codes," *IEEE Trans. Inf. Theory.*, **Vol. 53**, No. 2, February 2007.

[11] S. Korada, N.Macris, "Exact solution of a p-spin model and its relationship to error correcting codes", Proc. *ISIT, Seattle,* 2006.

[12] S. Kudekar, N.Macris, "Sharp Bounds for MAP Decoding of General Irregular LDPC Codes", Proc. *ISIT, Seattle,* 2006.