Concentration of Magnetization for Linear Block Codes

Satish Babu Korada, Shrinivas Kudekar, Nicolas Macris
School of Computer and Communication Sciences,
Ecole Polytechnique Fédérale de Lausanne
Email: {satish.korada, shrinivas.kudekar, nicolas.macris}@epfl.ch

***Abstract*— We consider communication over the binary erasure and the binary additive white gaussian noise channels using fixed linear block codes and also appropriate ensembles of such codes. We show concentration of the magnetization over the channel realizations and also over the code ensembles. The result has various implications. For the binary erasure channel, the result implies the concentration of the fraction of bits in error over the randomness in both noise and code realization, and that of the bit error probability under MAP decoding over the code ensemble. For both channels it implies concentration of the generalized EXIT function over code ensembles. Finally our results partly show that there is no replica symmetry breaking.**

I. Introduction and main results

Magnetization plays an important role in statistical mechanics [8] and proving its concentration is of interest. It is also important in communications due to its relation to various quantities which appear in the maximum a posteriori (MAP) decoding analysis of linear codes. In most relevant models in statistical mechanics or communications, the magnetization is believed to concentrate. But proving this fact is not trivial because it involves a bound on the second derivative of a free energy, and this quantity is not uniformly bounded (with respect to system size) at phase transition thresholds.

Recent convergence of statistical physics and communication has resulted in the application of various formal methods from statistical physics like the replica method, cavity method to the coding problem [3], [4], [5]. The communication system is interpreted as a random spin system and the replica method is applied to predict the performance. It is conjectured that this method yields correct results [5] and has been proved to be exact in some cases [12]. However proving the correctness of the replica or cavity methods is very difficult in general [8]. Even though the results in this paper do not prove any of these conjectures, they have been motivated by proofs of some of their aspects. In particular, as explained later they partly justify the (often accepted) assumption of absence of replica symmetry breaking for the coding problem with symmetric channels.

Our main result (Theorem 1.1) is valid for general fixed linear codes $\mathtt{C}$ of length n, and this is the main setting of the paper. This theorem basically states that, for a given linear code, for almost all values of the noise parameter, the magnetization concentrates on its average over the channel output realizations. By minor modifications in the proofs one can also obtain the corresponding statements for the standard parity check and generator matrix code ensembles. We note that the results cover the case of LDPC and LDGM ensembles but are not limited to low density ensembles.

We consider communication through a binary memoryless symmetric (BMS) channel with input $\{\pm 1\}$ and transition probability density $p_{Y|X}(y|x)$. The input to the channel is a codeword $\underline{x} = (x_1, x_2, \ldots, x_n)$ from a code $\mathtt{C}$ and the output is $\underline{y} = (y_1, y_2, \ldots, y_n)$. Let h_n denote the per-bit conditional entropy $n^{-1}H(\underline{X} \mid \underline{Y})$. Let $l_i = \frac{1}{2}\log\frac{p_{Y|X}(y_i|x=+1)}{p_{Y|X}(y_i|x=-1)}$ be the received half-loglikelihood for bit x_i and let $\underline{l} = (l_1, \ldots, l_n)$. The a posteriori distribution can be written as

$$p(\underline{x}|\underline{y}) = \frac{\mathbb{1}_{\{\underline{x}\in\mathtt{C}\}} e^{\sum_i l_i x_i}}{\sum_{\underline{x}} \mathbb{1}_{\{\underline{x}\in\mathtt{C}\}} e^{\sum_i l_i x_i}}$$

where $\mathbb{1}_{\{\underline{x}\in\mathtt{C}\}}$ is the code constraint. We denote by $\langle\cdot\rangle$, the average w.r.t the measure $p(\underline{x}|\underline{y})$. Since the channels considered are symmetric we can assume the transmission of all-one code word and the distribution of $\underline{l}$ is the distribution induced from such a transmission. The expectation with respect to the later is denoted by $\mathbb{E}_{\underline{l}}[\cdot]$. We define the partition function Z as

$$Z = \sum_{\underline{x}} \mathbb{1}_{\{\underline{x}\in\mathtt{C}\}} e^{\sum_i l_i x_i}$$

and the magnetization as

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

We have the following main result whose implications are explained in the next section.

Theorem 1.1: Consider communication over a BEC(ϵ), where ϵ is the erasure probability or a BAWGNC(ϵ) where ϵ^2 is the noise variance. There exists a finite positive constant c possibly depending on a, b and independent of n such that for any linear block code of length n,

$$\int_a^b d\epsilon\, \mathbb{E}_{\underline{l}}\big[\langle |m - \mathbb{E}_{\underline{l}}[\langle m\rangle]|\rangle\big] \leq \frac{c}{n^{\frac{1}{8}}}$$

where $[a, b] \in (0, 1)$ for BEC and $[a, b] \in (0, \infty)$ for BAWGNC.

Remark 1.1: For some code ensembles $\mathcal{C}$ which include standard parity check and generator matrix ensembles, LDPC ensembles and LDGM ensembles, we can show the same statement with $\mathbb{E}_{\underline{l}}$ replaced by $\mathbb{E}_{\mathcal{C},\underline{l}}$. In particular, by dominated convergence it follows that for Lebesgue almost every $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{C},\underline{l}}\langle|m - \mathbb{E}_{\mathcal{C},\underline{l}}\langle m\rangle|\rangle = 0 \tag{1}$$

The statement holds for almost every ϵ because it cannot be valid at phase transitions. In fact one expects that the phase transitions occur at isolated points so that the statement should hold for all ϵ away from these points.

II. Implications of Concentration of Magnetization

Fraction of bits in error. Consider the BEC and let $P_e(\underline{l})$ denote the fraction of bits in error for a given channel output realization $\underline{l}$,

$$P_e(\underline{l}) = \frac{1}{2}(1 - \langle m\rangle)$$

This formula is valid for the BEC because a bit x_i is either decoded correctly or not decoded, which implies $\langle x_i\rangle \in \{0, 1\}$ (for more general channels the right hand side has to be replaced by $\frac{1}{n}\sum_{i=1}^n \frac{1}{2}(1 - \mathrm{sgn}\langle x_i\rangle)$ with $\mathrm{sgn}(0) = 0$). The theorem implies concentration of the fraction of errors over its average which is nothing else than the bit MAP error probability,

$$P_e = \frac{1}{2}(1 - \mathbb{E}_{\underline{l}}\langle m\rangle)$$

This follows by bounding

$$\mathbb{E}_{\underline{l}}\big[|\langle m\rangle - \mathbb{E}_{\underline{l}}\langle m\rangle|\big] \le \mathbb{E}_{\underline{l}}\langle|m - \mathbb{E}_{\underline{l}}\langle m\rangle|\rangle$$

From Remark 1.1, the bit error probability of MAP decoding further concentrates on its average over the code ensemble $\mathbb{E}_{\mathcal{C}}[P_e]$.

GEXIT function. Another quantity related to magnetization is the Generalized EXIT (GEXIT) function introduced in [1], [2], for transmission over BMS channels using LDPC codes. Their motivation was primarily to elucidate the relationship between belief propagation (BP) and MAP decoding and provide upper bound to the MAP threshold. One wishes to know that the GEXIT function concentrates because this implies the upper bound for almost every code in the ensemble. In [2] the concentration is proved assuming that its derivative is uniformly bounded (away from thresholds). For BEC(ϵ) and BAWGNC(ϵ), we have the following formulas for the GEXIT function,

$$\text{BEC:}\quad h_n'(\epsilon) = \frac{dh_n}{d\epsilon} = \frac{\ln 2}{\epsilon}(1 - \mathbb{E}_{\underline{l}}[\langle m\rangle])$$
$$\text{BAWGNC:}\quad h_n'(\epsilon) = \frac{dh_n}{d\epsilon} = \epsilon^{-3}(1 - \mathbb{E}_{\underline{l}}[\langle m\rangle])$$

For these two channels we see that the magnetization is in some sense a more fundamental quantity that underlies the GEXIT function. Theorem 1.1 and (1) imply $\lim_{n\to\infty}\mathbb{E}_{\mathcal{C}}|h_n' - \mathbb{E}_{\mathcal{C}}[h_n']| = 0$.

Absence of replica symmetry breaking. Some random spin systems are believed to exhibit the phenomenon of replica symmetry breaking. In the present context this would mean that in the large block length limit the posterior measure becomes a convex combination of large number (exponential in the block length) of extremal measures. A criterion that signals such a phenomenon can be formulated thanks to the overlap parameter

$$q = \frac{1}{n}\sum_{i=1}^n x_i^{(1)} x_i^{(2)} \tag{2}$$

where the superscript is called the replica index and means that we take two independent samples $(x_1^{(a)}, ..., x_n^{(a)})$ from $p(\underline{x} \mid \underline{y})$. When the replica symmetry is broken the overlap parameter has a non trivial distribution [14]. Let $\mathbb{P}_m(x) = \mathbb{E}_{\mathcal{C},\underline{l}}\langle \mathbb{1}_{\{m=x\}}\rangle$, $\mathbb{P}_q(x) = \mathbb{E}_{\mathcal{C},\underline{l}}\langle \mathbb{1}_{\{q=x\}}\rangle_{1,2}$, where $\langle\cdot\rangle_{1,2}$ denotes the product measure $p(\underline{x}^{(1)} \mid \underline{y})\cdot p(\underline{x}^{(2)} \mid \underline{y})$ for the same output realization of $\underline{y}$. For any BMS channel and any linear block code, we have the Nishimori identity [4]

$$\mathbb{P}_m(x) = \mathbb{P}_q(x) \tag{3}$$

Thus concentration of m is equivalent to concentration of q so that the distribution of the overlap parameter remains a trivial delta function.

III. Proof of Theorem 1.1 for BEC

To prove the theorem we first bound the fluctuation $|\langle m\rangle - \mathbb{E}_{\underline{l}}\langle m\rangle|$. For this we define the following perturbed partition function

$$Z(\lambda) = \sum_{\underline{x}} \mathbb{1}_{\{\underline{x}\in \mathtt{C}\}} \prod_{i=1}^n \frac{e^{l_i x_i}}{2\cosh l_i} e^{\lambda \sum_i x_i}$$

where $\lambda > 0$. Note that at $\lambda = 0$ we recover the original partition function. Let

$$p_\lambda(\underline{x}|\underline{y}) = Z(\lambda)^{-1} \mathbb{1}_{\{\underline{x}\in \mathtt{C}\}} \prod_{i=1}^n \frac{e^{l_i x_i}}{2\cosh l_i} e^{\lambda \sum_i x_i}.$$

Let $\langle\cdot\rangle_\lambda$ denote the average w.r.t the measure $p_\lambda(\underline{x}|\underline{y})$. We define the free energy $f(\lambda) = \frac{1}{n}\ln Z(\lambda)$ and the average free energy

$$\mathcal{F}(\lambda) = \frac{1}{n}\mathbb{E}_{\underline{l}}[\ln Z(\lambda)]$$

Note that $f(\lambda)$ depends also on the values of $\underline{l}$ and $\mathcal{F}(\lambda)$ depends on the channel parameter ϵ through the distribution of $\underline{l}$ which are i.i.d. as $\epsilon\delta_0 + (1-\epsilon)\delta_\infty$.

It is easy to see that $\langle m\rangle_\lambda$ is equal to $f'(\lambda)$ (the derivative of $f(\lambda)$ w.r.t λ). Therefore proving the concentration of $f'(\lambda)$ implies the concentration of $\langle m\rangle_\lambda$. For this we first need to prove the concentration of $f(\lambda)$. This follows from martingale arguments. More precisely, we have

Lemma 3.1: For any $\nu > 0$,

$$\mathbb{P}\{|f(\lambda) - \mathcal{F}(\lambda)| \le \nu\} \ge 1 - 4e^{-\frac{n\nu^2}{8(\ln 2)^2}}$$

Proof: Let the noise realization be $(l_1, l_2, \ldots, l_n)$. Consider the following martingale:

$$X_i = \mathbb{E}_{l_{i+1}^n}[f(\lambda)|l_1^i]$$

where l_i^j denotes the noise vector $(l_i, \dots, l_j)$. From standard arguments [6]

$$|X_{i+1} - X_i| \le \max_{l_{i+1}} \left|\mathbb{E}_{l_{i+2}^n}[f(\lambda)|l_1^{i+1}] - X_i\right|$$
$$\le \max_{l_{i+1}, l'_{i+1}} \left|\mathbb{E}_{l_{i+2}^n}[f(\lambda)|l_1^{i+1}] - \mathbb{E}_{l_{i+2}^n}[f(\lambda)|l_1^i, l'_{i+1}]\right|$$

Since the noise loglikelihoods can take values in $\{0, \infty\}$, the maximum is attained for $l_{i+1} \neq l'_{i+1}$. Therefore a small calculation shows

$$|X_{i+1} - X_i| \le \frac{1}{n} |\mathbb{E}_{l_{i+2}^n}[\ln(1 + \langle x_{i+1}\rangle_{\lambda, l_{i+1}=0})]|$$

By Griffiths-Kelly-Sherman (GKS) inequalities [13] we have $0 \le \langle x_{i+1}\rangle_{\lambda,0} \le 1$ and hence $|X_{i+1} - X_i| \le \frac{1}{n}\ln 2$. Since $X_0 = \mathbb{E}_{\underline{l}}[f(\lambda)], X_n = f(\lambda)$, using Azuma's inequality we get the lemma. ■

Remark 3.1: To prove (1) we need to prove the analog of this lemma with $\mathbb{E}_{\underline{l}}$ replaced by $\mathbb{E}_{\mathcal{C},\underline{l}}$. But this easily follows by combining the above argument with [5] where the concentration of $\mathcal{F}(\lambda)$ on $\mathbb{E}_{\mathcal{C}}[\mathcal{F}(\lambda)]$ is proved. The rest of the proof below is identical whether we use $\mathbb{E}_{\underline{l}}$ or $\mathbb{E}_{\mathcal{C},\underline{l}}$.

Using convexity of $f(\lambda)$ and $\mathcal{F}(\lambda)$ with λ we bound the the fluctuation $\langle m\rangle_\lambda$ as follows.

Lemma 3.2: For any $0 < \delta < \lambda$ and any $\nu > 0$ we have

$$|\langle m\rangle_\lambda - \mathbb{E}_{\underline{l}}\langle m\rangle_\lambda| \le \frac{2\nu}{\delta} + \int_0^{\lambda+\delta} d\lambda \frac{d^2\mathcal{F}(\lambda)}{d\lambda^2}$$

with probability $1 - 16e^{-\frac{n\nu^2}{8(\ln 2)^2}}$

Proof: Consider $\lambda > 0$ and $0 < \delta < \lambda$. Note that we have

$$\langle m\rangle_\lambda = \frac{df(\lambda)}{d\lambda}, \qquad \mathbb{E}_{\underline{l}}\langle m\rangle_\lambda = \frac{d\mathcal{F}(\lambda)}{d\lambda} \tag{4}$$

From the convexity of $f(\lambda)$ and $\mathcal{F}(\lambda)$, we get

$$\frac{df(\lambda)}{d\lambda} - \frac{d\mathcal{F}(\lambda)}{d\lambda} \le \frac{d\mathcal{F}(\lambda+\delta)}{d\lambda} - \frac{d\mathcal{F}(\lambda)}{d\lambda}$$
$$+ \frac{1}{\delta}\{f(\lambda+\delta) - \mathcal{F}(\lambda+\delta)\} - \frac{1}{\delta}\{f(\lambda) - \mathcal{F}(\lambda)\}$$

From Lemma 3.1 we have

$$|f(\lambda+\delta) - \mathcal{F}(\lambda+\delta)| \le \nu \quad \text{and} \quad |f(\lambda) - \mathcal{F}(\lambda)| \le \nu$$

with probability greater than $1 - 8e^{-\frac{n\nu^2}{8(\ln 2)^2}}$. Therefore,

$$\frac{df(\lambda)}{d\lambda} - \frac{d\mathcal{F}(\lambda)}{d\lambda} \le \frac{2\nu}{\delta} + \int_\lambda^{\lambda+\delta} d\lambda \frac{d^2\mathcal{F}(\lambda)}{d\lambda^2}$$

with probability greater than $1 - 8e^{-\frac{n\nu^2}{8(\ln 2)^2}}$. Since $\frac{d^2\mathcal{F}(\lambda)}{d\lambda^2} \ge 0$ we get

$$\frac{df(\lambda)}{d\lambda} - \frac{d\mathcal{F}(\lambda)}{d\lambda} \le \frac{2\nu}{\delta} + \int_0^{\lambda+\delta} d\lambda \frac{d^2\mathcal{F}(\lambda)}{d\lambda^2}$$

Using similar arguments as before, with probability greater than $1 - 8e^{-\frac{n\nu^2}{8(\ln 2)^2}}$

$$\frac{df(\lambda)}{d\lambda} - \frac{d\mathcal{F}(\lambda)}{d\lambda} \ge -\frac{2\nu}{\delta} - \int_0^{\lambda+\delta} d\lambda \frac{d^2\mathcal{F}(\lambda)}{d\lambda^2}$$

The proof is completed by combining the last two inequalities with (4). ■

Proof of Theorem 1.1. We relate the fluctuation of $\langle m\rangle$ to $\langle m\rangle_\lambda$ by

$$\mathbb{E}_{\underline{l}}\big|\langle m\rangle - \mathbb{E}_{\underline{l}}\langle m\rangle\big| \le \mathbb{E}_{\underline{l}}\big|\langle m\rangle - \langle m\rangle_\lambda\big| + |\mathbb{E}_{\underline{l}}\big[\langle m\rangle_\lambda - \langle m\rangle\big]|$$
$$+ \mathbb{E}_{\underline{l}}\big|\langle m\rangle_\lambda - \mathbb{E}_{\underline{l}}\langle m\rangle_\lambda\big| \tag{5}$$

From convexity of $f(\lambda)$ we have $\langle m\rangle_\lambda - \langle m\rangle \ge 0$ for every code $\mathtt{C}$ and noise realization $\underline{l}$. Therefore we can replace the first term in (5) by $\mathbb{E}_{\underline{l}}[\langle m\rangle_\lambda] - \mathbb{E}_{\underline{l}}[\langle m\rangle]$. Using (4), $\big|m\big| \le 1$ and Lemma 3.2 we get

$$\mathbb{E}_{\underline{l}}\big|\langle m\rangle - \mathbb{E}_{\underline{l}}\langle m\rangle\big|$$
$$\le 3\int_0^{\lambda+\delta} \frac{d^2\mathcal{F}(\lambda)}{d\lambda^2}\, d\lambda + \frac{2\nu}{\delta} + 16e^{-\frac{n\nu^2}{8(\ln 2)^2}} \tag{6}$$

where we used the fact that $\frac{d^2\mathcal{F}(\lambda)}{d\lambda^2} \ge 0$ to obtain $\int_0^{\lambda+\delta} \frac{d^2\mathcal{F}(\lambda)}{d\lambda^2} \ge \int_0^{\lambda} \frac{d^2\mathcal{F}(\lambda)}{d\lambda^2}$. Differentiating the average free energy twice,

$$\frac{d^2\mathcal{F}(\lambda)}{d\lambda^2} = \frac{1}{n}\sum_{i,j} \mathbb{E}_{\underline{l}}\big[\langle x_i x_j\rangle_\lambda - \langle x_i\rangle_\lambda \langle x_j\rangle_\lambda\big]$$

Using the above equality along with Lemma 3.3 of [11], the second derivative w.r.t λ can be related to the second derivative w.r.t ϵ as

$$\frac{d^2\mathcal{F}(\lambda)}{d\lambda^2} \le 1 + 4\frac{d^2\mathcal{F}(\lambda)}{d\epsilon^2} \tag{7}$$

Substituting the bound (7) in (6) and integrating over $\epsilon \in [a, b] \subset (0, 1)$,

$$\int_a^b d\epsilon \mathbb{E}_{\underline{l}}\big|\langle m\rangle - \mathbb{E}_{\underline{l}}\langle m\rangle\big|$$
$$\le 3(\lambda+\delta) + 12\int_0^{\lambda+\delta} d\lambda \int_a^b d\epsilon \frac{d^2\mathcal{F}(\lambda)}{d\epsilon^2} + \frac{2\nu}{\delta} + 16e^{-\frac{n\nu^2}{8(\ln 2)^2}}$$

The first derivative of $\mathcal{F}(\lambda)$ w.r.t ϵ can be bounded as

$$\Big|\frac{d\mathcal{F}(\epsilon)}{d\epsilon}\Big| = \Big|\frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\underline{l}\setminus l_i}\left[\ln(1 + \langle x_i\rangle_{\lambda, l_i=0})\right]\Big| \le \ln 2$$

where the inequality follows from $\langle x_i\rangle_{\lambda, l_i=0} > 0$. Therefore the integral of the double derivative of $\mathcal{F}(\lambda)$ can be bounded by $2\ln 2$. Take $\delta = \sqrt{\nu}$ and $\lambda = 2\sqrt{\nu}$ to get

$$\int_a^b d\epsilon \mathbb{E}_{\underline{l}}\big|\langle m\rangle - \mathbb{E}_{\underline{l}}\langle m\rangle\big|$$
$$\le \sqrt{\nu}(11 + 72\ln 2) + 16e^{-\frac{n\nu^2}{8(\ln 2)^2}}$$

The proof is completed by substituting $\nu = n^{-\frac{1}{4}}$ along with the following concentration result from [11]

$$\int_a^b d\epsilon \mathbb{E}_{\underline{l}}\langle|m - \langle m\rangle|\rangle \le O\Big(\frac{1}{n^{\frac{1}{8}}}\Big)$$

■

IV. Proof of Theorem 1.1 for BAWGNC

Here we closely follow the approach developed in [10] for communication using Code Division Multiple Access. Note that the analysis developed for BEC does not apply here because the system is not ferromagnetic ($l_i \not\geq 0$). This prevents us from applying the GKS inequalities which we need to derive inequality similar to (7).

We parametrize the channels with their SNR $= \frac{1}{\epsilon^2}$, where ϵ^2 is the noise variance. Under the assumption of all-one codeword, the half log-likelihoods are distributed as $l_i \sim \mathcal{N}(\frac{1}{\epsilon^2}, \frac{1}{\epsilon^2})$. Therefore, the average free energy can be defined as

$$\tilde{\mathcal{F}}(u) = \frac{1}{n}\mathbb{E}_{\underline{l}}[\ln \sum_{\underline{x}} \mathbb{1}_{\{\underline{x}\in \mathtt{C}\}} e^{\sqrt{u}\sum_i l_i x_i + u\sum_i x_i}]$$

where $u = \frac{1}{\epsilon^2}$ and $l_i \sim \mathcal{N}(0,1)$. Note that SNR comes out explicitly in the expression of free energy which is not the case for BEC. This enables us to work without any additional variable like λ which was introduced in the case of the BEC. We define the following perturbed partition function,

$$Z(u) = \sum_{\underline{x}} \mathbb{1}_{\{\underline{x}\in \mathtt{C}\}} e^{\sqrt{u}\sum_i l_i x_i + u\sum_i x_i - \sqrt{u}\sum_i |l_i|} \tag{8}$$

where $l_i \sim \mathcal{N}(0,1)$. The free energy corresponding to this partition function is $f(u) = \frac{1}{n}\ln Z(u)$ and its average is denoted $\mathcal{F}(u) = \mathbb{E}_{\underline{l}}[f(u)]$. Note the extra perturbation $\sqrt{u}\sum_i |l_i|$ in this expression whose reason will become clear later. The idea is to use the concentration of first derivative of $f(u)$ to prove the concentration of magnetization.

As in the case of the BEC we first need to prove the concentration of the free energy. The proof relies on a general concentration theorem for suitable Lipschitz functions of many Gaussian random variables [7], [8]. In the version that we use here we need functions that are Lipschitz with respect to the Euclidean distance. More precisely we say that a function $f:\mathbb{R}^M \to \mathbb{R}$ is a Lipschitz function with constant L_M if for all $(\underline{y},\underline{z}) \in \mathbb{R}^M \times \mathbb{R}^M$

$$|f(\underline{y}) - f(\underline{z})| \leq L_M \|\underline{y} - \underline{z}\|$$

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

Theorem 4.1: [8] Let $Y_1, ..., Y_M$ be M independent identically distributed Gaussian random variables with distribution $\mathcal{N}(0, v^2)$ and let $f:\mathbb{R}^M \to \mathbb{R}$ be Lipschitz with respect to the Euclidean distance, with constant L_M. Then f satisfies

$$\mathbb{P}[|f(y_1, ..., y_M) - \mathbb{E}[f(y_1, ..., y_M)]| \geq t] \leq 2e^{-\frac{t^2}{2v^2 L_M^2}}$$

Lemma 4.1: For any $\nu > 0$ we have,

$$\mathbb{P}[|f(u) - \mathcal{F}(u)| \geq \nu] = 2e^{-\frac{n\nu^2}{8u}}$$

Proof: Here we use concentration of measure properties of gaussian random variables. For this we use the Lipschitz property of f with respect to $\underline{l}$. For a given code $\mathtt{C}$ and channel realizations $\underline{l}$, let $f(\underline{l})$ be the free energy $f(u)$. Let $\underline{l}_1, \underline{l}_2$ be two channel realizations and let $\langle\cdot\rangle_2$ denote the average w.r.t the measure

$$\frac{\mathbb{1}_{\{\underline{x}\in \mathtt{C}\}} e^{\sqrt{u}\sum_i l_{2i} x_i + u\sum_i x_i - \sqrt{u}\sum_i |l_{2i}|}}{\sum_{\underline{x}} \mathbb{1}_{\{\underline{x}\in \mathtt{C}\}} e^{\sqrt{u}\sum_i l_{2i} x_i + u\sum_i x_i - \sqrt{u}\sum_i |l_{2i}|}}$$

Then

$$\begin{aligned}
|f(\underline{l}_1) - f(\underline{l}_2)| &= \frac{1}{n}\ln\langle e^{\sqrt{u}\sum_i (l_{1i} - l_{2i})x_i - \sqrt{u}\sum_i (|l_{1i}| - |l_{2i}|)}\rangle_2 \\
&\leq \frac{2\sqrt{u}}{n}\sum_i |l_{1i} - l_{2i}| \\
&\leq \frac{2\sqrt{u}}{\sqrt{n}}\sqrt{\sum_i (l_{1i} - l_{2i})^2}
\end{aligned}$$

Using Theorem 4.1, we get the lemma. ■

In the sequel by $\mathbb{E}[\cdot]$ we mean $\mathbb{E}_{\mathcal{C},\underline{l}}[\cdot]$ and $\langle\cdot\rangle_u$ denotes the average w.r.t the measure

$$\frac{\mathbb{1}_{\{\underline{x}\in \mathtt{C}\}} e^{\sqrt{u}\sum_i l_i x_i + u\sum_i x_i - \sqrt{u}\sum_i |l_i|}}{\sum_{\underline{x}} \mathbb{1}_{\{\underline{x}\in \mathtt{C}\}} e^{\sqrt{u}\sum_i l_i x_i + u\sum_i x_i - \sqrt{u}\sum_i |l_i|}}$$

The perturbation term $\sqrt{u}\sum_i |l_i|$ in (8) has been chosen carefully so that the following holds,

Lemma 4.2: $f(u)$ is convex in u.

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

$$\frac{df(u)}{du} = \langle L(\underline{x})\rangle_u - \frac{1}{2\sqrt{u}}\sum_i |l_i|$$

where we have defined

$$L(\underline{x}) = \frac{1}{2n\sqrt{u}}\sum_i l_i x_i + \frac{1}{n}\sum_i x_i$$

Differentiating again,

$$\begin{aligned}
\frac{d^2 f(u)}{du^2} &= -\frac{1}{4nu^{3/2}}\Big\langle \sum_i l_i x_i \Big\rangle_u + \frac{1}{4nu^{3/2}}\sum_i |l_i| \\
&\quad + n(\langle L(\underline{x})^2\rangle_u - \langle L(\underline{x})\rangle_u^2) \geq 0
\end{aligned} \tag{9}$$

■

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

Lemma 4.3: For any $[a,b] \subset (0,\infty)$ fixed,

$$\int_a^b du \mathbb{E}\Big\langle \Big| L(\underline{x}) - \langle L(\underline{x})\rangle_u \Big| \Big\rangle_u = O\Big(\frac{1}{\sqrt{n}}\Big)$$

Proof: From equation (9), we have

$$\begin{aligned}
\int_a^b du \mathbb{E}\Big\langle \Big(L(\underline{x}) - \langle L(\underline{x})\rangle_u \Big)^2 \Big\rangle_u &\leq \int_a^b du \frac{1}{n}\frac{d^2}{du^2}\mathcal{F}(u) \\
&\leq \frac{1}{n}\Big(|\frac{d}{du}\mathcal{F}(b) - \frac{d}{du}\mathcal{F}(a)|\Big) = O\Big(\frac{1}{n}\Big)
\end{aligned}$$

In the very last equality we use that the first derivative of $\mathcal{F}$ is bounded for $u \geq a > 0$. Using Cauchy-Schwartz inequality we obtain the lemma. ■

Lemma 4.4: For any $[a,b] \subset (0,\infty)$ fixed,

$$\int_a^b du\mathbb{E}\Big|\langle L(\underline{x})\rangle_u - \mathbb{E}\langle L(\underline{x})\rangle_u\Big| \leq O\Big(\frac{1}{n^{\frac{1}{4}}}\Big)$$

Proof: From convexity of $f(u)$ with respect to u (Lemma 4.2) we have for any $\delta > 0$,

$$\frac{d}{du}f(u) - \frac{d}{du}\mathcal{F}(u) \leq \frac{d}{du}\mathcal{F}(u+\delta) - \frac{d}{du}\mathcal{F}(u) + \frac{1}{\delta}\{f(u+\delta) - \mathcal{F}(u+\delta)\} - \frac{1}{\delta}\{f(u) - \mathcal{F}(u)\}$$

A similar lower bound holds with δ replaced by $-\delta$. Now from Lemma 4.1 we know that the first two terms are $O(n^{-\frac{1}{2}})$. Thus from the formula for the first derivative in the proof of Lemma 4.2 and the fact that the fluctuations of $\frac{1}{n}\sum_{i=1}^n |l_i|$ are $O(n^{-\frac{1}{2}})$ we get

$$\mathbb{E}\Big|\langle L(\underline{x})\rangle_u - \mathbb{E}\langle L(\underline{x})\rangle_u\Big| \leq \frac{c}{\delta\sqrt{n}} + \frac{d}{du}\mathcal{F}(u+\delta) - \frac{d}{du}\mathcal{F}(u)$$

for some positive constant c. We will choose $\delta = n^{-\frac{1}{4}}$. Note that we cannot assume that the difference of the two derivatives is small because the first derivative of the free energy is not uniformly continuous in n (as $n \to \infty$ it may develop jumps at the phase transition points). The free energy itself is uniformly continuous (because of convexity). For this reason if we integrate with respect to u, we get

$$\int_a^b du\mathbb{E}\Big|\langle L(\underline{x})\rangle_u - \mathbb{E}\langle L(\underline{x})\rangle_u\Big| \leq O\Big(\frac{1}{n^{\frac{1}{4}}}\Big)$$

■

Proof of Theorem 1.1 for BAWGNC. Combining the concentration lemmas we get

$$\int_a^b du\mathbb{E}\langle|L(\underline{x}) - \mathbb{E}\langle L(\underline{x})\rangle_u|\rangle_u \leq O\Big(\frac{1}{n^{\frac{1}{4}}}\Big)$$

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

$$\int_a^b du|\mathbb{E}\langle L(\underline{x})g(\underline{x})\rangle_u - \mathbb{E}\langle L(\underline{x})\rangle_u\mathbb{E}\langle g(\underline{x})\rangle_u|\rangle_u \leq \int_a^b du\mathbb{E}\langle|L(\underline{x}) - \mathbb{E}\langle L(\underline{x})\rangle_u|\rangle_u$$

More generally the same thing holds if one takes a function depending on many replicas such as $g(\underline{x}^{(1)}, \underline{x}^{(2)}) = q_{12}$. Here q_{12} is equal to the q defined in (2). The subscript denotes the replica indices involved in the overlap parameter. Using integration by parts formula with respect to l_i like in [10],

$$\begin{aligned}\mathbb{E}\langle L(\underline{x})q_{12}\rangle_u &= \mathbb{E}\Big\langle\frac{1}{n}\frac{1}{2\sqrt{u}}\sum_i l_i x_i q_{12}\Big\rangle_u + \mathbb{E}\langle mq_{12}\rangle_u \\ &= \frac{1}{2}\mathbb{E}\langle(1+q_{12})q_{12}\rangle_u - \frac{1}{2}\mathbb{E}\langle(q_{13}+q_{14})q_{12}\rangle_u + \mathbb{E}\langle mq_{12}\rangle_u \\ &= \frac{1}{2}\mathbb{E}\langle(1+q_{12})q_{12}\rangle_u = \frac{1}{2}\mathbb{E}\langle m + m^2\rangle_u \qquad (10)\end{aligned}$$

where in the last two equalities we used the Nishimori identity (3). By a similar calculation,

$$\begin{aligned}\mathbb{E}\langle L(\underline{x})\rangle_u\mathbb{E}\langle q_{12}\rangle_u &= \frac{1}{2}\mathbb{E}\langle 1 - q_{12} + 2m\rangle_u\mathbb{E}\langle q_{12}\rangle_u \\ &= \frac{1}{2}(\mathbb{E}\langle m\rangle + (\mathbb{E}\langle m\rangle)^2) \qquad (11)\end{aligned}$$

From equations (10) and (11), we get

$$\int_a^b du\mathbb{E}\langle(m - \mathbb{E}\langle m\rangle_u)^2\rangle_u \leq O\Big(\frac{1}{n^{\frac{1}{4}}}\Big)$$

Using Cauchy-Schwartz inequality we get the result. ■

V. Discussion

Both the proof techniques used above do not extend to the case of binary symmetric channel (BSC). The first approach fails again because the system is not ferromagnetic. The second approach relies on the integration by parts formula specific for gaussian random variables, which can not be used for BSC.

Acknowledgment

The work of S. Kudekar is 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] C. Measson, A. Montanari, T. Richardson, R. Urbanke, "The Generalized Area Theorem and Some of its Consequences," *E-print: cs.IT/0511039. Submitted to IEEE Transactions on Information Theory.*

[2] C. Measson, A. Montanari, R. Urbanke, " Maxwell Construction: The Hidden Bridge between Iterative and Maximum a Posteriori Decoding", *E-print: cs.IT/0506083. Submitted to IEEE Transactions on Information Theory.*

[3] N. Sourlas,"Spin glass models as error correcting codes", *Nature*, **339** pp. 693–695, 1989.

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

[5] A. Montanari, "Tight Bounds for LDPC and LDGM codes," *IEEE Transactions on Information Theory,* **51**, Issue 9, (2005) pp:3221 - 3246

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

[7] M. Talagrand, "A New look at independence", *The Annals of Probability* vol. 24, Issue. 1, (1996) pp. 1–34.

[8] M. Talagrand, *"Spin Glasses: A Challenge for Mathematicians"* Springer-Verlag (2003)

[9] S.B. Korada, N. Macris, "On the concentration of capacity of a Code Division Multiple Access System", *Proc. IEEE Int. Symp. Information Theory*, Nice, Jul. 2007.

[10] S.B. Korada, N. Macris, "On the capacity of a Code Division Multiple Access System", in *Proc. of the Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, USA, 2007.

[11] S. Kudekar, N. Macris, "Sharp bounds for MAP decoding of general irregular code s", in *Proc. IEEE Int. Symp. Information Theory*, Seatle, WA, Jul. 2006, pp2259-2263.

[12] S.B. Korada, S. Kudekar, N. Macris, "Exact solution for the conditional entropy of Poissonian LDPC codes over the Binary Erasure Channel", in *Proc. IEEE Int. Symp. Information Theory*, Nice, Jul. 2007, pp2259-2263.

[13] N. Macris, "Griffith-Kelly-Sherman Correlation Inequalities: A Useful Tool in the Theory of Error Correcting Codes," *IEEE Trans. Inf. Theory.*, 53(2):664-683, (2007).

[14] M. Mézard, G. Parisi, M. Virasoro, *"Spin lass Theory and Beyond"*, World Scientific (1984)