Typical performance of low-density parity-check codes over general symmetric channels

Toshiyuki Tanaka[1,2] and David Saad[2]

[1] *Department of Electronics and Information Engineering, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji-shi, Tokyo, 192-0397 Japan*

[2] *Neural Computing Research Group, Aston University, Aston Triangle, Birmingham, B4 7ET, United Kingdom*

(Dated: April 29, 2002)

Typical performance of low-density parity-check (LDPC) codes over a general binary-input output-symmetric memoryless channel is investigated using methods of statistical mechanics. Theoretical framework for dealing with general symmetric channels is provided, based on which Gallager and MacKay-Neal codes are studied as examples of LDPC codes. It has been shown that the basic properties of these codes known for particular channels, including the property to potentially saturate Shannon's limit, hold for general symmetric channels. The binary-input additive-white-Gaussian-noise channel and the binary-input Laplace channel are considered as specific channel noise models.

I. INTRODUCTION

We investigate the typical performance of low-density parity-check (LDPC) codes over a general binary-input output-symmetric (BIOS) memoryless channel. Previous statistical physics based analyses of LDPC codes have discovered some interesting properties, including the fact that they can, in principle, saturate the information-theoretic upper bound (Shannon's bound defined by the channel coding theorem [1]) with low connectivity values. Existing statistical mechanical studies on the LDPC codes, however, have been mostly confined to the case of binary symmetric channel (BSC), which fits into the statistical-mechanical framework in a natural way [2, 3, 4, 5]. Notable exceptions are the work by Montanari [6] that discusses the case of binary-input additive-white-Gaussian-noise channel (BIAWGNC) as well as the BSC case and the study of Sourlas codes [7], a simple LDPC code, in which non-BSC channels are addressed [8, 9, 10]. From the statistical-mechanical point of view, LDPC codes are regarded as random spin systems; it is therefore natural to expect that they will exhibit some sort of universality, just as typical statistical-mechanical systems do, so that general properties of LDPC codes observed in the BSC case will be preserved when different communication channels are considered. In this paper we investigate the properties of LDPC codes in binary-input output-symmetric channels and show that this is generally the case. In particular, we show that the finite connectivity LDPC codes can saturate Shannon's bound for general BIOS channel.

The paper is organized as follows: In section II we introduce the general framework, notation, codes and the channels that we will focus on. In section III we will briefly describe the calculation for the various channels, while the results obtained will be described in section IV, followed by the conclusions.

II. THE GENERAL FRAMEWORK

A. Symmetric channels

We consider the general class of binary-input output-symmetric (BIOS) memoryless channel. The input of the channel is binary (± 1), and the output may take any real value. The characteristics of a channel is described by the channel transition probabilities, $P(y|x=1)$ and $P(y|x=-1)$. Let $p(y) \equiv P(y|x=1)$. A symmetric channel is characterized as a channel whose transition probabilities satisfy $P(y|x=-1) = P(-y|x=1) = p(-y)$. Various types of channel models of practical interest fall into the class of BIOS channels, including the binary symmetric channel (BSC)

$$p_{\rm BSC}(y) = (1-p)\delta(y-1) + p\delta(y+1), \tag{1}$$

the binary-input additive-white-Gaussian-noise channel (BIAWGNC)

$$p_{\rm BIAWGNC}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-1)^2/2\sigma^2}, \tag{2}$$

and the binary-input Laplace channel (BILC)

$$p_{\rm BILC}(y) = \frac{1}{\lambda} e^{-|y-1|/\lambda}, \tag{3}$$

Each of the parameters p, σ^2, and λ represents the degree of degradation induced by the channel noise. We call each of them the noise level and let d denotes the generic one.

An apparent technical difficulty in dealing with a general channel of real-valued output is that it is not at all obvious how to define the syndrome from the received signal: The modulo 2 arithmetic involved in computing syndrome in the BSC case is not directly applicable to the cases of real-valued received signal. This difficulty is resolved by using a truncation procedure [11]: We conceptually consider another *fictitious* binary-input binary-output channel in addition to the channel under study.

Let r be the (fictitious) output symbol of this fictitious channel. We can assign to r either of the values ± 1 arbitrarily, and the binary channel noise ζ for the fictitious channel is defined therefrom, via $r = x\zeta$. For the sake of making the argument simple, we assign $r = 1$ without loss of generality. Since the prior probability of ζ (*before* receiving y) should be $P(\zeta = \pm 1) = 1/2$, the joint distribution of y and ζ is given by

$$P(y, \zeta) = \frac{p(\zeta y)}{2} \tag{4}$$

since the truncation procedure used here yields $x = \zeta$.

B. Gallager code

LDPC codes have been originally introduced by Gallager in his seminal work from 1963 [12]. Gallager's original construction [12] is one of the most extensively studied LDPC codes in the information theory literature. It is defined by its parity-check matrix $A = [C_1|C_2]$ of dimensionality $(M-N) \times M$, which is taken to be random and very sparse. The submatrix C_2, of dimensionality $(M-N) \times (M-N)$, is assumed invertible.

In the encoding step, the encoder computes a codeword from the information vector $\boldsymbol{\xi} \in \{0,1\}^N$ by employing a generator matrix G

$$\boldsymbol{x} = G^T \boldsymbol{\xi} \mod 2, \tag{5}$$

where the generator matrix is defined by

$$G = [I|C_2^{-1}C_1] \mod 2. \tag{6}$$

This construction ensures $AG^T = 0 \mod 2$. The information code rate for unbiased messages is $R = N/M$.

In regular Gallager codes, the number of non-zero elements per row of A is fixed to be K. We call it the row constraint. Average number of non-zero elements per column is then $C \equiv K(M-N)/M$, whereas we will consider the case in which the number of non-zero elements in each column is forced to be exactly C, which we term the *column constraint*. *Irregular* Gallager codes can be defined by relaxing these constraints. It has been known that making code construction irregular may improve performance significantly [13], but we will not discuss irregular codes in the current paper. We call the resulting regular Gallager code a (C, K)-Gallager code.

C. MN code

We also discuss a variant of LDPC codes, called the MN code [11, 14]. The generator matrix G^T of the MN code is defined by

$$G^T = C_n^{-1} C_s \mod 2, \tag{7}$$

where C_s and C_n are sparse matrices of dimensionality $M \times N$ and $M \times M$, respectively; C_n is assumed invertible. The information rate for the code is $R = N/M$ for unbiased message.

In regular MN codes the row and column constraints are imposed on both matrices C_s and C_n. The number of non-zero elements per row of C_s and C_n should be exactly K and L, respectively. Also here, we do not discuss irregular MN codes [15] in this paper. The number of non-zero elements per column of C_s and C_n are set to C and L, respectively, where $C = KM/N$ holds. We call the resulting code a (K, C, L)-MN code.

III. ANALYSIS

A. Gallager code

The basic idea behind the statistical-mechanical treatment of the LDPC codes is the equivalence between the decoding problem and the thermal equilibrium distribution of a dilute Ising spin system. In order to see this in the Gallager code case, one should first note that the decoding problem is to find $\boldsymbol{\tau}$ which is best supported (i.e., most probable) by the received signal $\boldsymbol{y}$ among the set of $\boldsymbol{\tau}$ satisfying the parity-check equation ($A\boldsymbol{\zeta} = A\boldsymbol{\tau}$ mod 2 if we write it in the $\{0,1\}$-notation). The set is expressed as

$$\left\{ \boldsymbol{\tau} \;\middle|\; \lim_{\gamma \to \infty} \exp\left[-\gamma \sum_{\mu=1}^{M-N} \left(J_\mu \prod_{j \in \mathcal{L}(\mu)} \tau_j - 1 \right) \right] = 1 \right\}, \tag{8}$$

where

$$\mathcal{L}(\mu) = \{ j | A_{\mu j} = 1 \} \tag{9}$$

denotes the set of indices for which the parity-check matrix A has 1's in μ-th row, and

$$J_\mu = \prod_{j \in \mathcal{L}(\mu)} \zeta_j \tag{10}$$

is μ-th check. The posterior probability of $\boldsymbol{\tau}$ conditioned on the received signal $\boldsymbol{y}$ then acquires the following Gibbs-Boltzmann form:

$$P_\gamma(\boldsymbol{\tau}|\boldsymbol{y}) = \frac{1}{Z} \exp\left[-\beta \mathcal{H}_\gamma(\boldsymbol{\tau}; \boldsymbol{y}, \boldsymbol{J}) \right] \tag{11}$$

in which we have to take the limit $\gamma \to \infty$ and consider it at $\beta = 1$ (Nishimori's temperature [8, 16, 17, 18]) in order to obtain the true posterior. The Hamiltonian $\mathcal{H}_\gamma(\boldsymbol{\tau}; \boldsymbol{y}, \boldsymbol{J})$ is defined as

$$\begin{aligned} \mathcal{H}_\gamma(\boldsymbol{\tau}; \boldsymbol{y}, \boldsymbol{J}) \;=\; & -\gamma \sum_{\mu=1}^{M-N} \left(J_\mu \prod_{j \in \mathcal{L}(\mu)} \tau_j - 1 \right) \\ & - \sum_{j=1}^{M} \log p(\tau_j y_j), \end{aligned} \tag{12}$$

The channel characteristics enters into the Hamiltonian as the term $\log p(\tau_j y_j)$ which, by noting that τ_j takes ± 1, can be rewritten as

$$\log p(\tau_j y_j) = \tau_j \frac{1}{2} \log \frac{p(y_j)}{p(-y_j)} + \frac{1}{2} \log p(y_j) p(-y_j). \quad (13)$$

From this expression it immediately follows that it is the log-likelihood ratio $h_j \equiv (1/2)\log(p(y_j)/p(-y_j))$ of the channel noise y_j which serves as the external field acting on site j, and that the channel characteristics defines the field distribution. Analyzing the effect of having different communication channels on the code properties, therefore reduces to investigating the effect of different field distributions on the physical properties of the system. The field distributions $p(h)$ for various channel models are as follows:

- BSC:

$$p_{\rm BSC}(h) = (1-p)\delta\left(h - \frac{1}{2}\log\frac{1-p}{p}\right) + p\delta\left(h + \frac{1}{2}\log\frac{1-p}{p}\right) \quad (14)$$

- BIAWGNC:

$$p_{\rm BIAWGNC}(h) = \sqrt{\frac{\sigma^2}{2\pi}} e^{-(h-\sigma^{-2})^2/2\sigma^{-2}} \quad (15)$$

- BILC:

$$p_{\rm BILC}(h) = \frac{1}{2}\delta(h-\lambda^{-1}) + \frac{e^{-2\lambda^{-1}}}{2}\delta(h+\lambda^{-1}) + \chi[-\lambda^{-1} < h < \lambda^{-1}]\frac{1}{2}e^{h-\lambda^{-1}}, \quad (16)$$

  where $\chi[X]$ is the indicator function, taking 1 when X is true and 0 otherwise.

Sketches of these field distributions are given in Fig. 1.

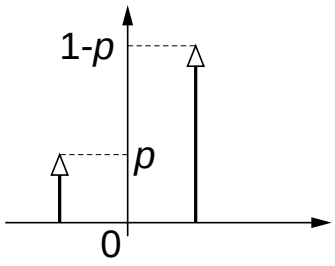
We assume that the free energy of the system is self-averaging, that is,

$$f = -\frac{1}{\beta} \lim_{M\to\infty} M^{-1} \langle \log Z \rangle_{A,\boldsymbol{y}}, \quad (17)$$

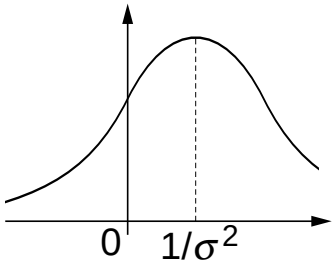
and evaluate the average $\langle\cdot\rangle_{A,\boldsymbol{y}}$ over the received signal $\boldsymbol{y}$ and the randomness of the parity-check matrix A using the replica method,

$$f = -\frac{1}{\beta} \lim_{M\to\infty} \lim_{n\to 0} M^{-1} \frac{\partial}{\partial n} \log \langle Z^n \rangle_{A,\boldsymbol{y}}. \quad (18)$$

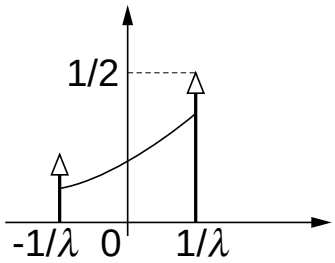
In calculating the free energy, we perform the gauge transformation $\tau_j \to \zeta_j \tau_j$, $y_j \to \zeta_j y_j$. The average over $\boldsymbol{y}$ can be taken with respect to $\prod_{j=1}^{M} p(y_j)$ after having performed the gauge transformation. We have to introduce a random tensor to take average over A.



(a) BSC



(b) BIAWGNC



(c) BILC

FIG. 1: Field distributions corresponding to various BIOS channels.

Following basically the same procedure as in [3] and exchanging the order of the two limits, taking the limit $M \to \infty$ first, one obtains

$$f = -\frac{1}{\beta} \lim_{n\to 0} \frac{\partial}{\partial n} \operatorname*{Extr}_{\boldsymbol{q},\hat{\boldsymbol{q}}} \left[\frac{C}{K} \mathcal{G}_1(\boldsymbol{q}) - \mathcal{G}_2(\boldsymbol{q},\hat{\boldsymbol{q}}) + \mathcal{G}_3(\hat{\boldsymbol{q}}) \right], \quad (19)$$

where

$$\begin{aligned} \mathcal{G}_1(\boldsymbol{q}) &\equiv \log \sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} q^K_{\alpha_1 \cdots \alpha_m} - n \log 2, \\ \mathcal{G}_2(\boldsymbol{q},\hat{\boldsymbol{q}}) &\equiv \sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} q_{\alpha_1 \cdots \alpha_m} \hat{q}_{\alpha_1 \cdots \alpha_m}, \\ \mathcal{G}_3(\hat{\boldsymbol{q}}) &\equiv \log \left[\sum_{\tau^1,\ldots,\tau^n} \left\langle \prod_{\alpha=1}^{n} p(\tau^\alpha y) \right\rangle_y \right. \\ &\quad \left. \times \frac{1}{C!} \left(\sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} \hat{q}_{\alpha_1 \cdots \alpha_m} \tau^{\alpha_1} \cdots \tau^{\alpha_m} \right)^C \right]. \end{aligned} \quad (20)$$

To proceed further we adopt the replica-symmetric (RS) ansatz and let

$$q_{\alpha_1\cdots\alpha_m} = q_0 \int u^m \pi(u)\, du, \quad \hat{q}_{\alpha_1\cdots\alpha_m} = \hat{q}_0 \int \hat{u}^m \hat{\pi}(\hat{u})\, d\hat{u}. \tag{21}$$

We will use the following simplifying notation.

$$\pi^K(\boldsymbol{u})\, d\boldsymbol{u} \equiv \prod_{j=1}^{K} \pi(u_j)\, du_j \tag{22}$$

The replica-symmetric free energy $f^{\rm RS}$ becomes

$$\begin{aligned} f^{\rm RS} &= \frac{1}{\beta} \mathop{\rm Extr}_{\pi,\hat{\pi}} \Bigg\{ \frac{C}{K} \log 2 \\ &+ C \iint \log(1+u\hat{u})\, \pi(u)\, \hat{\pi}(\hat{u})\, du\, d\hat{u} \\ &- \frac{C}{K} \int \log\Big(1 + \prod_{j=1}^{K} u_j\Big)\, \pi^K(\boldsymbol{u})\, d\boldsymbol{u} \\ &- \int \Big\langle \log\Big[p(y) \prod_{l=1}^{C} (1+\hat{u}_l) + p(-y) \prod_{l=1}^{C} (1-\hat{u}_l) \Big] \Big\rangle_y \\ &\qquad \times \hat{\pi}^C(\hat{\boldsymbol{u}})\, d\hat{\boldsymbol{u}} \Bigg\}, \end{aligned} \tag{23}$$

in which q_0 and $\hat{q}_0$ have been eliminated using the extremization condition $q_0\hat{q}_0 = C$. Heuristic construction of a sufficient condition to the extremization problem with respect to π and $\hat{\pi}$ is possible, and it gives the following saddle-point equations.

$$\begin{aligned} \pi(u) &= \int \Big\langle \delta\Big[u - \tanh\Big(h(y) + \sum_{l=1}^{C-1} \tanh^{-1} \hat{u}_l \Big) \Big] \Big\rangle_y \\ &\qquad \times \hat{\pi}^{C-1}(\hat{\boldsymbol{u}})\, d\hat{\boldsymbol{u}} \\ \hat{\pi}(\hat{u}) &= \int \delta\Big(\hat{u} - \prod_{j=1}^{K-1} u_j \Big)\, \pi^{K-1}(\boldsymbol{u})\, d\boldsymbol{u} \end{aligned} \tag{24}$$

We have let

$$h(y) \equiv \frac{1}{2} \log \frac{p(y)}{p(-y)}. \tag{25}$$

The performance of the code is quantified by the overlap $m = M^{-1} \sum_{k=1}^{M} \zeta_j \langle \tau_j \rangle$, which is given as

$$m = \int \mathrm{sign}(z)\, P(z)\, dz, \tag{26}$$

where

$$\begin{aligned} P(z) &= \int \Big\langle \delta\Big[z - \tanh\Big(h(y) + \sum_{l=1}^{C} \tanh^{-1} \hat{u}_l \Big) \Big] \Big\rangle_y \\ &\qquad \times \hat{\pi}^C(\hat{\boldsymbol{u}})\, d\hat{\boldsymbol{u}}. \end{aligned} \tag{27}$$

B. MN code

The decoding problem for the MN code is to find $\boldsymbol{S}$ and $\boldsymbol{\tau}$ which are the best suitable in view of the received signal $\boldsymbol{y}$ among the sets of $\boldsymbol{S}$ and $\boldsymbol{\tau}$ satisfying the parity-check equation ($C_s\boldsymbol{S} + C_n\boldsymbol{\tau} = C_s\boldsymbol{\xi} + C_n\boldsymbol{\zeta} \mod 2$ if written in the $\{0,1\}$-notation). Defining the μth component of the check $\boldsymbol{J}$ as

$$J_\mu = \prod_{j \in \mathcal{L}_s(\mu)} \xi_j \prod_{l \in \mathcal{L}_n(\mu)} \zeta_l, \tag{28}$$

where

$$\mathcal{L}_s(\mu) = \{j | (C_s)_{\mu j} = 1\}, \quad \mathcal{L}_n(\mu) = \{l | (C_n)_{\mu l} = 1\}, \tag{29}$$

the posterior probability of $\boldsymbol{S}$ and $\boldsymbol{\tau}$ conditioned on the received signal $\boldsymbol{y}$ and the check $\boldsymbol{J}$ is given by

$$P_\gamma(\boldsymbol{S}, \boldsymbol{\tau} | \boldsymbol{y}, \boldsymbol{J}) = \frac{1}{Z} \exp\big[-\beta \mathcal{H}_\gamma(\boldsymbol{S}, \boldsymbol{\tau}; \boldsymbol{y}, \boldsymbol{J}) \big], \tag{30}$$

in the limit $\gamma \to \infty$ and at $\beta = 1$, where the Hamiltonian $\mathcal{H}_\gamma(\boldsymbol{S}, \boldsymbol{\tau}; \boldsymbol{y}, \boldsymbol{J})$ is defined as

$$\begin{aligned} \mathcal{H}_\gamma(\boldsymbol{S}, \boldsymbol{\tau}; \boldsymbol{y}, \boldsymbol{J}) &= -\gamma \sum_{\mu=1}^{M} \Big(J_\mu \prod_{j \in \mathcal{L}_s(\mu)} S_j \prod_{l \in \mathcal{L}_n(\mu)} \tau_l - 1 \Big) \\ &\quad - F_s \sum_{j=1}^{N} S_j - \sum_{l=1}^{M} \log p(\tau_l y_l), \end{aligned} \tag{31}$$

where F_s is a parameter representing the bias of the information vector $\boldsymbol{\xi}$ in such a way that $P(\xi_j = \pm 1) = (1 \pm \tanh F_s)/2$ holds. The form of Eq. (31) clearly shows that the channel characteristics again acts as random field on $\{\tau_l\}$, where the log likelihood ratio gives the actual value of the field.

The replica calculation can be done along the same way as in the case of the Gallager code. We have performed the gauge transformation $S_j \to \xi_j S_j$, $\tau_j \to \zeta_j \tau_j$, and $y_j \to \zeta_j \tau_j$. The free energy f becomes

$$\begin{aligned} f &= -\frac{1}{\beta} \lim_{n\to 0} \frac{\partial}{\partial n} \mathop{\rm Extr}_{\boldsymbol{q},\hat{\boldsymbol{q}},\boldsymbol{r},\hat{\boldsymbol{r}}} \Big[\frac{C}{K} \mathcal{G}_1(\boldsymbol{q}, \boldsymbol{r}) \\ &\qquad - \mathcal{G}_2(\boldsymbol{q}, \hat{\boldsymbol{q}}, \boldsymbol{r}, \hat{\boldsymbol{r}}) + \mathcal{G}_3(\hat{\boldsymbol{q}}, \hat{\boldsymbol{r}}) \Big], \end{aligned} \tag{32}$$

where

$$\begin{aligned} \mathcal{G}_1(\boldsymbol{q}, \boldsymbol{r}) &\equiv \log \sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} q^K_{\alpha_1\cdots\alpha_m} r^L_{\alpha_1\cdots\alpha_m} \\ &\quad - n \log 2, \\ \mathcal{G}_2(\boldsymbol{q}, \hat{\boldsymbol{q}}, \boldsymbol{r}, \hat{\boldsymbol{r}}) &\equiv \sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} q_{\alpha_1\cdots\alpha_m} \hat{q}_{\alpha_1\cdots\alpha_m} \\ &\quad + \frac{M}{N} \sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} r_{\alpha_1\cdots\alpha_m} \hat{r}_{\alpha_1\cdots\alpha_m}, \end{aligned} \tag{33}$$

and

$$\mathcal{G}_3(\hat{\boldsymbol{q}},\hat{\boldsymbol{r}}) \equiv \log\Bigg[\sum_{S^1,\ldots,S^n}\Big\langle e^{F_s\sum_{\alpha=1}^n\xi S^\alpha}\Big\rangle_\xi \times\frac{1}{C!}\Bigg(\sum_{m=0}^n\sum_{\langle\alpha_1\cdots\alpha_m\rangle}\hat{q}_{\alpha_1\cdots\alpha_m}S^{\alpha_1}\cdots S^{\alpha_m}\Bigg)^C\Bigg] + \frac{M}{N}\log\Bigg[\sum_{\tau^1,\ldots,\tau^n}\Big\langle\prod_{\alpha=1}^n p(\tau^\alpha y)\Big\rangle_y \times\frac{1}{L!}\Bigg(\sum_{m=0}^n\sum_{\langle\alpha_1\cdots\alpha_m\rangle}\hat{r}_{\alpha_1\cdots\alpha_m}\tau^{\alpha_1}\cdots\tau^{\alpha_m}\Bigg)^L\Bigg]. \tag{34}$$

We adopt the RS ansatz as before, under which we have

$$r_{\alpha_1\cdots\alpha_m}=r_0\int v^m\rho(v)\,dv,\quad \hat{r}_{\alpha_1\cdots\alpha_m}=\hat{r}_0\int\hat{v}^m\hat{\rho}(\hat{v})\,d\hat{v}, \tag{35}$$

in addition to Eq. (21). The replica-symmetric free energy $f^{\rm RS}$ becomes

$$\begin{aligned} f^{\rm RS} &= \frac{1}{\beta}\operatorname*{Extr}_{\pi,\hat{\pi},\rho,\hat{\rho}}\Bigg\{\frac{C}{K}\log 2 \\ &+ C\iint\log(1+u\hat{u})\,\pi(u)\,\hat{\pi}(\hat{u})\,du\,d\hat{u} \\ &+ \frac{CL}{K}\iint\log(1+v\hat{v})\,\rho(v)\,\hat{\rho}(\hat{v})\,dv\,d\hat{v} \\ &- \frac{C}{K}\iint\log\Big(1+\prod_{k=1}^K u_k\prod_{l=1}^L v_l\Big) \times\pi^K(\boldsymbol{u})\,d\boldsymbol{u}\,\rho^L(\boldsymbol{v})\,d\boldsymbol{v} \\ &- \int\Big\langle\log\Big[\sum_{S=\pm1}e^{F_s\xi S}\prod_{k=1}^C(1+S\hat{u}_k)\Big]\Big\rangle_\xi\hat{\pi}^C(\hat{\boldsymbol{u}})\,d\hat{\boldsymbol{u}} \\ &- \frac{C}{K}\int\Big\langle\log\Big[\sum_{\tau=\pm1}p(\tau y)\prod_{l=1}^L(1+\tau\hat{v}_l)\Big]\Big\rangle_y \times\hat{\rho}^L(\hat{\boldsymbol{v}})\,d\hat{\boldsymbol{v}}\Bigg\}, \end{aligned} \tag{36}$$

in which q_0, $\hat{q}_0$, r_0, and $\hat{r}_0$ have been eliminated using the extremization conditions, $q_0\hat{q}_0=C$ and $r_0\hat{r}_0=L$.

Construction of a heuristic solution to the extremization problem can be done in the same manner, which yields the following saddle-point equations:

$$\begin{aligned} \pi(u) &= \int\Big\langle\delta\Big[u-\tanh\Big(F_s\xi+\sum_{l=1}^{C-1}\tanh^{-1}\hat{u}_l\Big)\Big]\Big\rangle_\xi \times\hat{\pi}^{C-1}(\hat{\boldsymbol{u}})\,d\boldsymbol{u} \\ \hat{\pi}(\hat{u}) &= \iint\delta\Big(\hat{u}-\prod_{k=1}^{K-1}u_k\prod_{l=1}^L v_l\Big)\,\pi^{K-1}(\boldsymbol{u})\,d\boldsymbol{u}\,\rho^L(\boldsymbol{v})\,d\boldsymbol{v} \\ \rho(v) &= \int\Big\langle\delta\Big[v-\tanh\Big(h(y)+\sum_{l=1}^{L-1}\tanh^{-1}\hat{v}_l\Big)\Big]\Big\rangle_y \times\hat{\rho}^{L-1}(\hat{\boldsymbol{v}})\,d\hat{\boldsymbol{v}} \\ \hat{\rho}(\hat{v}) &= \iint\delta\Big(\hat{v}-\prod_{k=1}^{K}u_k\prod_{l=1}^{L-1}v_l\Big)\,\pi^{K}(\boldsymbol{u})\,d\boldsymbol{u}\,\rho^{L-1}(\boldsymbol{v})\,d\boldsymbol{v} \end{aligned} \tag{37}$$

The overlap is then evaluated by

$$m=\int\operatorname{sign}(z)\,P(z)\,dz, \tag{38}$$

where

$$P(z) = \int\Big\langle\delta\Big[z-\tanh\Big(F_s\xi+\sum_{l=1}^{C}\tanh^{-1}\hat{u}_l\Big)\Big]\Big\rangle_\xi \times\hat{\pi}^C(\hat{\boldsymbol{u}})\,d\hat{\boldsymbol{u}}. \tag{39}$$

It is worthwhile mentioning that, when the message is unbiased ($F_s=0$) and K is even, saddle-point solutions have the following symmetry: For each solution $\{\pi(u),\hat{\pi}(\hat{u}),\rho(v),\hat{\rho}(\hat{v})\}$ there is another solution $\{\pi(-u),\hat{\pi}(-\hat{u}),\rho(v),\hat{\rho}(\hat{v})\}$. The latter has the same overlap as that of the former with the opposite sign.

IV. RESULTS

A. Gallager code

1. Analytical solutions

Of particular interest is the ferromagnetic state, which corresponds to an error-free communication. One can see that the assertion

$$\pi(u)=\delta(u-1),\quad \hat{\pi}(\hat{u})=\delta(\hat{u}-1) \tag{40}$$

always satisfies the saddle-point equation (24) irrespective of the values of K and C (provided that $K,C\geq 2$), thereby providing a solution. The overlap and the free energy of the solution at $\beta=1$ are $m_{\rm ferro}=1$ and $f_{\rm ferro}=-\langle\log p(y)\rangle_y$, respectively. One can therefore identify this as the ferromagnetic solution.

Another solution, which can be found in the limit $K\to\infty$, is the sub-optimal ferromagnetic solution

$$\pi(u)=\big\langle\delta\big[u-\tanh h(y)\big]\big\rangle_y,\quad \hat{\pi}(\hat{u})=\delta(\hat{u}), \tag{41}$$

for which

$$m_{\rm sf} = \big\langle {\rm sign}\big[p(y) - p(-y)\big]\big\rangle_y \tag{42}$$

and

$$f_{\rm sf} = \frac{C}{K}\log 2 - \big\langle \log\big[p(y) + p(-y)\big]\big\rangle_y. \tag{43}$$

The difference of the free energy is expressed as

$$f_{\rm sf} - f_{\rm ferro} = \mathsf{C} - R\log 2, \tag{44}$$

where C is the channel capacity of the BIOS channel defined as

$$\mathsf{C} = \log 2 - \big\langle \log\big[p(y) + p(-y)\big]\big\rangle_y + \big\langle \log p(y)\big\rangle_y. \tag{45}$$

This proves that the thermodynamic transition between the ferromagnetic and sub-optimal ferromagnetic solutions (no other solution has been identified in this case) occurs at the theoretical limit. This means that the maximum rate $R_{\rm max}$, up to which error-free communication is theoretically possible, asymptotically achieves the theoretical limit as $K \to \infty$. This result has been known for BSC channel [4, 5] in the physics literature and is in agreement with results reported in the information theory literature [11]. The current result is an extension to the case of a general BIOS channel.

2. *Numerical solutions of saddle-point equations*

In finite-K cases no simple analytical solution exists other than the ferromagnetic one, so one has to solve the saddle-point equations numerically. We have done it for BIAWGNC and BILC. The dependence of the overlap m on the noise level d (σ^2 for BIAWGNC, and λ for BILC) is qualitatively the same as that observed in BSC: For $K \geq 3$ the ferromagnetic solution is locally stable over the whole range of noise levels. At $d = d_s$, another solution with $m < 1$ appears, which defines the spinodal point. At a higher noise level $d = d_t > d_s$ thermodynamic transition takes place, beyond which the ferromagnetic solution with $m = 1$ becomes metastable (see Fig. 2). Table I summarizes the results for the BIAWGNC case, showing the spinodal point σ_s^2 (the value of the variance at which new, non ferromagnetic, solutions emerge), the thermodynamic transition point σ_t^2 (at which the thermodynamic transition occurs), and σ_0^2, the information-theoretic upper bound of the variance allowing error-free communication.

Table II summarizes the results for the BILC case, showing the values of the spinodal point λ_s, the thermodynamic transition point λ_t, and the information-theoretic upper bound λ_0.

It should be noted that the results for the spinodal point agree well with the results obtained by the density evolution approach [19], as expected, since the saddle-point equations by the replica analysis happen to coincide with the time evolution equations in the density evolution.

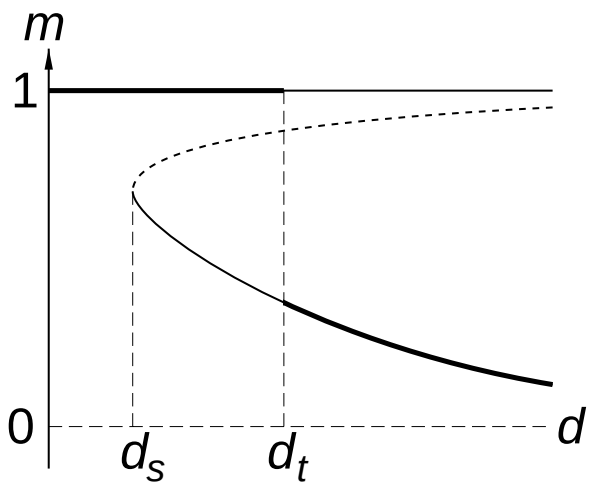


FIG. 2: Noise-overlap diagram for Gallager code. Thick solid lines stand for the stable state, thin solid lines for metastable state, and broken lines for unstable states. The ferromagnetic solution is characterized by the $m = 1$ solution, while $m < 1$ defines the suboptimal ferromagnetic solution.

TABLE I: The variances σ_s^2 and σ_t^2 at the spinodal point and thermodynamic transition, respectively, for the BIAWGNC for various code parameters; σ_0^2, denoting the information-theoretical upper bound for error-free communication, is also shown.

C	K	R	σ_s^2	σ_t^2	σ_0^2
3	6	0.5	0.775	0.899	0.958
4	8	0.5	0.701	0.943	0.958
5	10	0.5	0.629	0.952	0.958
3	5	0.4	1.017	1.253	1.321
4	6	0.333	1.020	1.666	1.681
3	4	0.25	1.598	2.325	2.401

B. MN code

1. *Analytical solutions*

In the following we restrict our discussion of the MN code to the unbiased case $F_s = 0$. The ferromagnetic

TABLE II: The parameter values λ_s and λ_t at the spinodal point and thermodynamic transition, respectively, for the BILC with various code parameters; λ_0, denoting the information-theoretical upper bound for error-free communication, is also shown.

C	K	R	λ_s	λ_t	λ_0
3	6	0.5	0.651	0.712	0.752
4	8	0.5	0.618	0.741	0.752
5	10	0.5	0.581	0.746	0.752
3	5	0.4	0.773	0.875	0.914
4	6	0.333	0.782	1.045	1.055
3	4	0.25	1.018	1.260	1.298

solution, corresponding to the error-free communication, can be constructed for the MN code with $L \geq 2$. (In fact, in the case $L = 1$ the matrix C_n reduces to a simple permutation matrix, so that we have to estimate each element of noise separately. This case is not at all interesting and therefore we will not discuss it any more.) It is given by

$$\begin{aligned} &\pi(u) = \delta(u-1), \quad \hat{\pi}(\hat{u}) = \delta(\hat{u}-1), \\ &\rho(v) = \delta(v-1), \quad \hat{\rho}(\hat{v}) = \delta(\hat{v}-1), \end{aligned} \tag{46}$$

for which $m_{\rm ferro} = 1$ and

$$f_{\rm ferro} = -\frac{C}{K}\big\langle \log p(y) \big\rangle_y. \tag{47}$$

The MN code has the following paramagnetic solution for $K \geq 2$:

$$\begin{aligned} &\pi(u) = \delta(u), && \hat{\pi}(\hat{u}) = \delta(\hat{u}), \\ &\rho(v) = \big\langle \delta\big[v - \tanh h(y)\big] \big\rangle_y, && \hat{\rho}(\hat{v}) = \delta(\hat{v}), \end{aligned} \tag{48}$$

which yields $m_{\rm para} = 0$ and

$$f_{\rm para} = \left(\frac{C}{K} - 1\right)\log 2 - \frac{C}{K}\big\langle \log\big[p(y) + p(-y)\big] \big\rangle_y. \tag{49}$$

Again, since

$$f_{\rm para} - f_{\rm ferro} = \frac{C}{K}(\mathsf{C} - R\log 2) \tag{50}$$

holds, we conclude that for the MN code the maximum rate $R_{\rm max}$, theoretically allowing error-free communication, achieves the theoretical limit as long as $K \geq 2$, $L \geq 2$, provided that there is no locally stable solution other than the ferromagnetic and paramagnetic solutions. This result is an extension of the result reported in [2, 3] to the case of a general BIOS channel.

It should be noted that the paramagnetic solution (46) is also valid in the limit $L \to \infty$ for the case $K = 1$. This means that the above-mentioned result also holds for the case of $K = 1$ asymptotically in the limit $L \to \infty$.

2. *Numerical solutions of saddle-point equations*

In order to explore solutions other than the ferromagnetic and paramagnetic solutions, we have to solve the saddle-point equations numerically. We have done it for the BIAWGNC and BILC cases. We observed qualitatively the same characteristics as those reported in [3].

The obtained numerical results suggest that the qualitative physical properties are categorized into three types according to the K value: cases with $K = 1$, $K = 2$ and $K \geq 3$, whereas it is only affected quantitatively by the values of C and L, as described in the following.

The structure of noise-overlap diagram for the MN code with $K = 1$ is qualitatively the same as that for Gallager code (see Fig. 2): At very low noise level only the ferromagnetic solution with $m = 1$ exists. At a certain noise level $d = d_s$ another metastable solution with $m < 1$ appears, and it becomes dominant beyond $d = d_t > d_s$. Since the latter solution is obtained only numerically, there is no guarantee that the thermodynamical transition d_t is equal to the information-theoretical limit d_0. Numerical results show that in general d_t *is* smaller than d_0: However, it is also observed that, for fixed C, increasing L makes d_s smaller and d_t larger, the latter of which approaches the information-theoretical limit d_0 as $L \to \infty$, as discussed at the end of the previous subsection. Even for finite L the value of d_t may be numerically very close to d_0, especially when the rate R is small. These properties have already been reported for the BSC case [3], so that our finding implies that they also hold for the BIAWGNC and BILC cases, revealing some sort of universality.

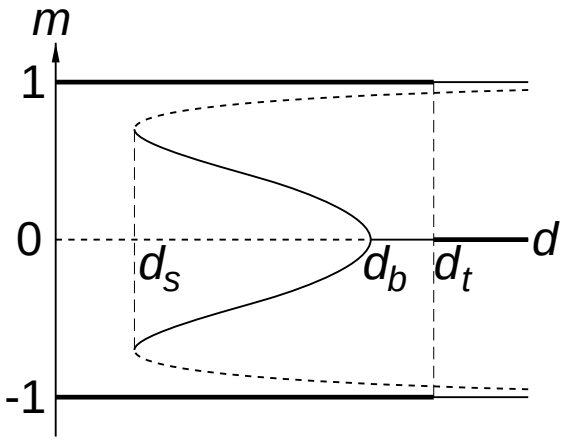
The noise-overlap diagram for the cases with $K = 2$

TABLE III: The variances σ_s^2, σ_t^2, and σ_b^2 at the spinodal point and thermodynamic transition, and at bifurcation of paramagnetic solution, respectively, for (K, C, L)-MN codes over the BIAWGNC and various code parameters; σ_0^2, denoting the information-theoretical upper bound for error-free communication, is also shown.

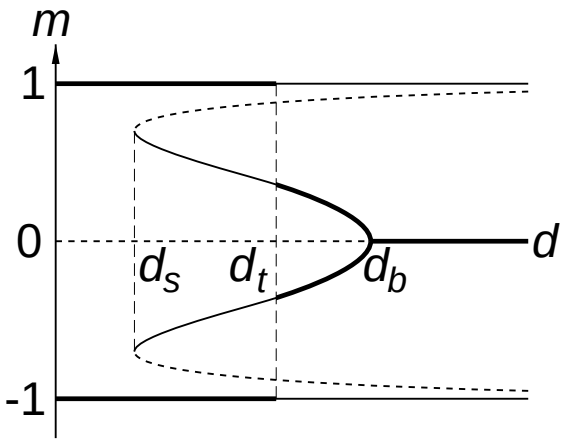
K	C	L	R	σ_s^2	σ_t^2	σ_b^2	σ_0^2
1	2	3	0.5	0.775	0.901	–	0.958
1	2	4	0.5	0.703	0.944	–	0.958
1	2	5	0.5	0.630	0.955	–	0.958
1	3	2	0.333	1.338	1.423	–	1.681
1	3	3	0.333	1.129	1.659	–	1.681
1	3	4	0.333	0.913	1.672	–	1.681
2	3	2	0.667	0.536	0.587	0.612	0.588
2	3	3	0.667	0.430	0.588	0.459	0.588
2	3	4	0.667	0.368	0.588	0.385	0.588
2	4	2	0.5	0.809	0.958	0.919	0.958
2	5	2	0.4	1.039	1.321	1.175	1.321

TABLE IV: The parameter values λ_s, λ_t, and λ_b at the spinodal point and thermodynamic transition, and at bifurcation of paramagnetic solution, respectively, for (K, C, L)-MN codes over the BILC and various code parameters; λ_0, denoting the information-theoretical upper bound for error-free communication, is also shown.

K	C	L	R	λ_s	λ_t	λ_b	λ_0
1	2	3	0.5	0.652	0.714	–	0.752
1	2	4	0.5	0.619	0.740	–	0.752
1	2	5	0.5	0.582	0.748	–	0.752
1	3	2	0.333	0.903	0.934	–	1.055
1	3	3	0.333	0.831	1.040	–	1.055
1	3	4	0.333	0.735	1.051	–	1.055
2	3	2	0.667	0.525	0.551	0.597	0.553
2	3	3	0.667	0.464	0.553	0.493	0.553
2	3	4	0.667	0.419	0.553	0.437	0.553
2	4	2	0.5	0.689	0.751	0.771	0.752
2	5	2	0.4	0.807	0.914	0.894	0.914



(a)



(b)

FIG. 3: Noise-overlap diagram for the cases with $K = 2$.

has the general structure shown in Fig. 3. The diagram is characterized by three transition points: the spinodal point d_s, the thermodynamic transition d_t, and the bifurcation point d_b. The order of the thermodynamic transition d_t and the bifurcation point d_b varies with the values of C and L, so that the bifurcation pattern for the cases with $K = 2$ is further divided into two sub-categories depending on the order of the two transitions: $d_s < d_b < d_t$ for the first group, and $d_s < d_t < d_b$ for the second group. The noise-overlap diagrams for these groups are illustrated in Fig. 3 (a) and (b), respectively. By the local stability analysis the bifurcation point d_b is determined by

$$\int v\rho(v)\,dv = (C-1)^{-1/L}, \tag{51}$$

which allows us to decide the type of bifurcation of a particular case. See the appendix for derivation of Eq. (51). As a result, we found that only a few cases with small values of C and L fall into the second category. The values of C and L for which the $(2, C, L)$-MN code fall into the second category depend on the channel characteristics; as far as we have observed, only the cases

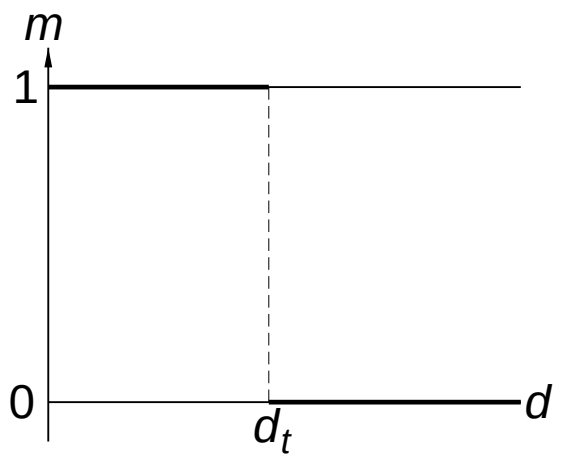


FIG. 4: Noise-overlap diagram for the cases with $K \geq 3$.

with $L = 2$ fall into the second group. For the BIAWGNC case, the $(2, 3, 2)$-MN code is the only one instance, whereas for the BILC case, both $(2, 3, 2)$- and $(2, 4, 2)$-MN codes belong to this group. (For the BSC case, $(2, 3, 2)$-, $(2, 4, 2)$-, and $(2, 5, 2)$-MN codes belong to this group.) All the $(2, C, L)$-MN codes but those mentioned above are in the first group. For the cases in the second group, the thermodynamic transition d_t must be less than the information-theoretic limit d_0: However, it turns out numerically that d_t is very close to d_0.

We observed that the noise-overlap diagram for the cases with $K \geq 3$ is relatively simple for the BIAWGNC and BILC cases, just as in the BSC case (Fig. 4): The ferromagnetic solution with $m = 1$ (and its mirror image with $m = -1$ when K is even) and the paramagnetic solution are the only stable solutions found, both of which are locally stable over the whole range of the noise level. The system exhibits a first-order transition at the information-theoretic limit d_t. We did not find any solutions other than the ferromagnetic and paramagnetic solutions.

V. CONCLUSIONS

We have analyzed typical performance of LDPC codes over BIOS channel using statistical mechanics. We have shown for the case of LDPC codes that the log-likelihood ratio of the received signal serves as an external random field acting on each site, and that channel characteristics define the distribution of the random field. The Gallager and MN codes are analyzed, to find that the basic properties of these codes remain unchanged regardless of channel characteristics. In particular, it has been shown that these codes potentially saturate Shannon's limit asymptotically, as $K \to \infty$, for the Gallager code; and when $K, L \geq 2$ — with a few exceptions with small C and L values — and asymptotically as $L \to \infty$ for $K = 1$, for the MN code. Saddle-point solutions have also been

numerically evaluated extensively for the cases of BIAWGNC and BILC channels, from which noise-overlap diagrams, as well as the transition and bifurcation points, have been characterized.

Acknowledgments

We would like to thank Yoshiyuki Kabashima for his helpful suggestions, Jort van Mourik for providing computer programs, and Nikos Skantzos for helpful discussions. Support from EPSRC research grant GR/N00562 is acknowledged.

*

APPENDIX A: STABILITY OF PARAMAGNETIC SOLUTION FOR $K \geq 2$

To probe the stability of paramagnetic solution, which exists for $K \geq 2$, we analyze the stability with respect to $\boldsymbol{q}$ and $\boldsymbol{r}$ only, and do not consider stability with respect to $\hat{\boldsymbol{q}}$ and $\hat{\boldsymbol{r}}$; these conjugate variables are subsidiary to their counterparts, $\boldsymbol{q}$ and $\boldsymbol{r}$, respectively, so that the former should not be considered as independent variables.

Let A, B, ... denote sets of replica indices such as $\langle \alpha_1 \cdots \alpha_m \rangle$, $m \geq 1$. We first evaluate the Hessian of the free energy (32) with respect to $4 \times (2^n - 1)$ variables $\{q_A, \hat{q}_A, r_A, \hat{r}_A\}$:

$$H = \begin{pmatrix} H_{\boldsymbol{qq}} & H_{\boldsymbol{q}\hat{\boldsymbol{q}}} & O & \\ H_{\boldsymbol{q}\hat{\boldsymbol{q}}} & H_{\hat{\boldsymbol{q}}\hat{\boldsymbol{q}}} & & \\ O & & O & H_{\boldsymbol{r}\hat{\boldsymbol{r}}} \\ & & H_{\boldsymbol{r}\hat{\boldsymbol{r}}} & H_{\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}} \end{pmatrix}, \tag{A1}$$

where

$$\begin{aligned} \left(H_{\boldsymbol{qq}}\right)_{AB} &= \begin{cases} 0 & (K \geq 3) \\ -\dfrac{C}{q_0^2}\left(\dfrac{r_A}{r_0}\right)^L \delta_{AB} & (K = 2) \end{cases} \\ \left(H_{\boldsymbol{q}\hat{\boldsymbol{q}}}\right)_{AB} &= \delta_{AB} \\ \left(H_{\hat{\boldsymbol{q}}\hat{\boldsymbol{q}}}\right)_{AB} &= -\frac{C(C-1)}{\hat{q}_0^2}\delta_{AB} \\ \left(H_{\boldsymbol{r}\hat{\boldsymbol{r}}}\right)_{AB} &= \frac{M}{N}\delta_{AB} \\ \left(H_{\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}}\right)_{AB} &= -\frac{M}{N}\frac{L(L-1)}{\hat{r}_0^2}\delta_{AB} \end{aligned} \tag{A2}$$

The block-diagonal structure of the Hessian allows us to decompose the stability problem into two, one with respect to $\boldsymbol{q}$, and another with respect to $\boldsymbol{r}$.

Following the argument in the appendix of [20], one can say that the system is stable with respect to $\boldsymbol{q}$ if the matrix $H_c \equiv H_{\boldsymbol{qq}} - H_{\boldsymbol{q}\hat{\boldsymbol{q}}}\left(H_{\hat{\boldsymbol{q}}\hat{\boldsymbol{q}}}\right)^{-1} H_{\boldsymbol{q}\hat{\boldsymbol{q}}}$ is positive definite. This condition takes into account the fact that $\hat{\boldsymbol{q}}$ depends on $\boldsymbol{q}$. A corresponding statement holds for the stability with respect to $\boldsymbol{r}$.

The stability with respect to $\boldsymbol{r}$ is straightforward, by noting that the matrix $H_{\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}}$ is negative definite, which means that $H_c = -(M/N)^2\left(H_{\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}}\right)^{-1}$ is positive definite.

We consider the stability with respect to $\boldsymbol{q}$. For $K \geq 3$, we have $H_c = [\hat{q}_0^2/C(C-1)]I$, where I is the identity matrix, so that the stability immediately follows, irrespective of the noise level of the channel. For $K = 2$, the matrix H_c is diagonal, and its A-th element is

$$\left(H_c\right)_{AA} = -\frac{C}{q_0^2}\left(\frac{r_A}{r_0}\right)^L + \frac{\hat{q}_0^2}{C(C-1)}. \tag{A3}$$

Using the equality which holds under the RS ansatz,

$$\frac{r_A}{r_0} = \int_{-1}^{1} v^m \rho(v)\, dv, \tag{A4}$$

where $A = \alpha_1 \cdots \alpha_m$, we have, as the stability condition,

$$E_m \equiv \int_{-1}^{1} v^m \rho(v)\, dv < (C-1)^{-1/L}. \tag{A5}$$

for $m = 1, \ldots, n$. Since it can be shown that $E_{2m-1} = E_{2m}$ and $E_{2m} \geq E_{2m+2}$, the critical condition determining the stability is

$$E_1 < (C-1)^{-1/L}. \tag{A6}$$

[1] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948).
[2] Y. Kabashima, T. Murayama, and D. Saad, Phys. Rev. Lett. **84**, 1355 (2000).
[3] T. Murayama, Y. Kabashima, D. Saad, and R. Vicente, Phys. Rev. E **62**, 1577 (2000).
[4] R. Vicente, D. Saad and Y. Kabashima, Europhys. Lett. **51** 698 (2000).
[5] J. van Mourik, D. Saad, and Y. Kabashima (unpublished).
[6] A. Montanari, Eur. Phys. J. B, **23** 121 (2001).
[7] N. Sourlas, Nature (London) **339**, 693 (1989).
[8] P. Ruján, Phys. Rev. Lett. **70**, 2968 (1993).
[9] H. Nishimori and K. Y. Michael Wong, Phys. Rev. E **60**,

132 (1999).
[10] R. Vicente, D. Saad, and Y. Kabashima, Phys. Rev. E **60**, 5352 (1999).
[11] D. J. C. MacKay, IEEE Trans. Inform. Theory **45**, 399 (1999).
[12] R. G. Gallager, IRE Trans. Inform. Theory **IT-8**, 21 (1962).
[13] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, IEEE Trans. Inform. Theory **47**, 619 (2001).
[14] D. J. C. MacKay and R. M. Neal, in *Cryptography and Coding, 5th IMA Conference*, Lecture Notes in Computer Science 1025, edited by C. Boyd (Springer, 1995), p. 100.
[15] I. Kanter and D. Saad, Phys. Rev. Lett. **83**, 2660 (1999).
[16] H. Nishimori, J. Phys. Soc. Jpn. **62**, 2973 (1993).
[17] Y. Iba, J. Phys. A **32** 3875 (1999).
[18] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing* (Oxford University Press, Oxford, UK, 2001).
[19] T. J. Richardson and R. L. Urbanke, IEEE Trans. Inform. Theory **47**, 599 (2001).
[20] T. Tanaka (unpublished).