Statistical mechanics of broadcast channels using low-density parity-check codes

Kazutaka Nakamura* and Yoshiyuki Kabashima†
Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Yokohama 226-8502, Japan

Robert Morelos-Zaragoza‡
Electrical Engineering Department, San José State University, One Washington Square, San José, California 95192-0084

David Saad§
The Neural Computing Research Group, School of Engineering and Applied Science, Aston University, Birmingham B4 7ET, United Kingdom

(Received 26 August 2002; published 28 March 2003)

We investigate the use of Gallager's low-density parity-check (LDPC) codes in a degraded broadcast channel, one of the fundamental models in network information theory. Combining linear codes is a standard technique in practical network communication schemes and is known to provide better performance than simple time sharing methods when algebraic codes are used. The statistical physics based analysis shows that the practical performance of the suggested method, achieved by employing the belief propagation algorithm, is superior to that of LDPC based time sharing codes while the best performance, when received transmissions are optimally decoded, is bounded by the time sharing limit.

DOI: 10.1103/PhysRevE.67.036703 PACS number(s): 02.70.−c, 89.70.+c, 05.50.+q, 05.70.Fh

I. INTRODUCTION

Progress in digital communication technologies has dramatically increased the information flow in both wired and wireless channels. This makes the role of generic coding techniques, such as error-correcting codes and data compression, more important. As most existing codes are constructed for simple point-to-point communication, they do not necessarily provide optimal performance in multiterminal communication such as the internet, mobile phones, and satellite communication. Therefore, designing improved codes that utilize characteristic properties of these media is a promising direction for enhancing the performance of multiterminal communication.

The broadcast channel is a standard multiterminal communication channel composed of a single sender and multiple receivers, and is characteristic of television (TV) and radio broadcasting. This implies that constructing a jointly optimal code with respect to the multiple channels may provide improved performance (i.e., higher capacity) compared to that of the time sharing scheme, whereby separate optimally designed codes are used for each channel. It has been shown, under the assumption of *degraded channels*, that jointly optimized codes can have a larger capacity region, where error-free communication is possible, than that of time sharing codes [1–4]. However, this proof is nonconstructive and the search for better practical codes for broadcast channels is still an important topic in information theory.

The purpose of this paper is to devise and analyze an improved practical code for a degraded broadcast channel by linearly combining low-density parity-check (LDPC) codes, which have been shown to provide nearly optimal performance for single channels [5–7]. For Reed-Solomon and BCH codes, which are standard suboptimal codes, it has been reported that combining codes linearly results in superior performance with respect to a time sharing transmission [9,8]. This provides the motivation for the current study, investigating the performance of linearly combined LDPC codes.

Generally, one can define two different performance measures for evaluating LDPC codes. The first is the *practical* performance achievable in feasible time scales that grow polynomially with the systems size; while the other is the *optimal* theoretically achievable performance, for which the required computation typicallly increases exponentially with respect to the message length. Utilizing the similarity between LDPC codes and Ising spin systems, statistical physics provides a scheme for evaluating both performance measures within the same framework [10–12]; the current standard method used in the information theory community [13] can only provide an estimate of the practical performance, and practically reduces to the one used within the statistical physics framework. In this paper, we show that the statistical physics based analysis points to a superior practical performance of the suggested method with respect to LDPC based time sharing codes (achieved by employing the belief propagation algorithm); while its optimal performance is bounded by the time sharing limit, which cannot be saturated by known practical methods.

This paper is organized as follows. In the following section, we introduce the general framework for broadcast channels. Unlike simple communication channels, the optimal communication performance is still unknown for most

*Electronic address: knakamur@fe.dis.titech.ac.jp
†Electronic address: kaba@dis.titech.ac.jp
‡Electronic address: rmorelos@email.sjsu.edu
§Electronic address: d.saad@aston.ac.uk

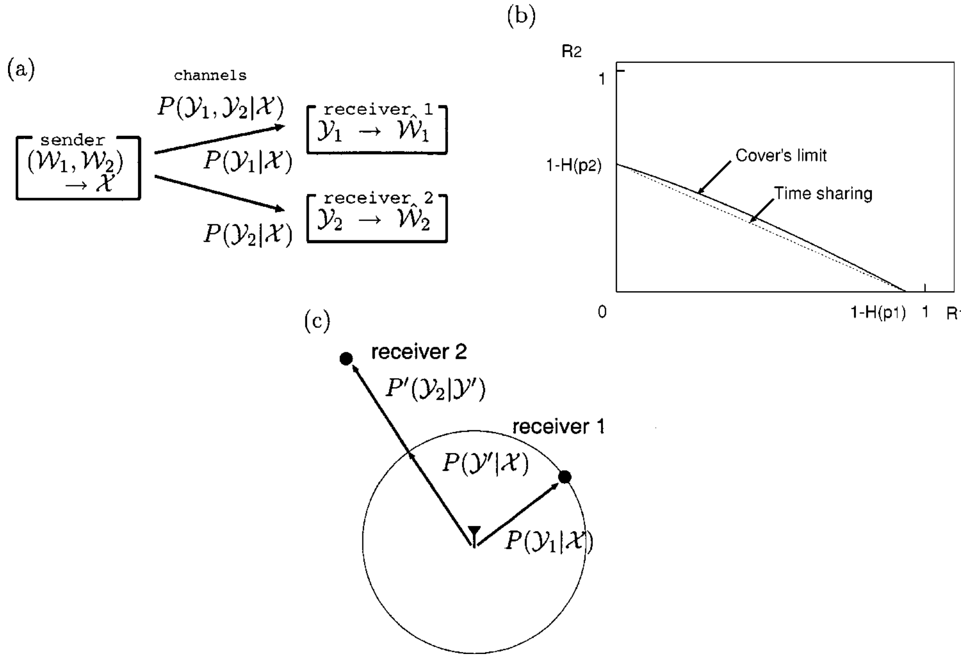

FIG. 1. (a) A single sender and two receivers broadcast channel. (b) The capacity region in the case of binary symmetric channels. The solid curve and the dotted line denote Cover's and time sharing limits, respectively. (c) When the corruption rate increases proportionally to a distance from a broadcast station (sender), the functional form of the conditional probability $P(\mathcal{Y}|\mathcal{X})$ becomes identical on a circular arc of a fixed radius centered at the station. This implies that the conditional probability for the second receiver can be expressed as $P(\mathcal{Y}_2|\mathcal{X}) = \Sigma_{\mathcal{Y}'} P'(\mathcal{Y}_2|\mathcal{Y}')P(\mathcal{Y}'|\mathcal{X}) = \Sigma_{\mathcal{Y}_1} P'(\mathcal{Y}_2|\mathcal{Y}_1)P(\mathcal{Y}_1|\mathcal{X})$, where $\mathcal{Y}'$ is the received code word at the closest point to the second receiver on the circle and $P'(\mathcal{Y}_2|\mathcal{Y}')$ models the corruption process between the second receiver and its closest point on the circle, which seems as if the code word were conveyed to the second receiver in a relay scheme via the first receiver.

broadcast channels, which would make it difficult to evaluate the performance of the proposed scheme. Therefore, we focus here on the degraded channel, for which the capacity region has already been obtained. In Sec. III, an LDPC code based construction for degraded channels is introduced, and is subsequently analyzed in Sec. IV using methods of statistical physics. In Sec. V, the performance of the proposed scheme is evaluated by solving numerically equations that emerge from the analysis. Section VI is devoted to a summary and conclusion.

II. DEGRADED BROADCAST CHANNEL

In the general framework of broadcast channels, a single sender (station) broadcasts a code word composed of different messages to multiple receivers. For simplicity, we here restrict our attention to the case of two receivers (Fig. 1), where one code word $\mathcal{X}$ (N bits), comprising of two messages $\mathcal{W}_1(R_1N$ bits) and $\mathcal{W}_2(R_2N$ bits), is sent to two receivers. As each channel is noisy, receivers 1 and 2 obtain two corrupted code words $\mathcal{Y}_1$ and $\mathcal{Y}_2$, respectively; this is modeled by a conditional probability $P(\mathcal{Y}_1,\mathcal{Y}_2|\mathcal{X})$. The received corrupted code words $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are decoded by the respective receivers to retrieve only the message addressed to each of them.

Analogously to the case of single channels, error-free communication becomes possible if the corresponding code rate vector (R_1,R_2) lies within a certain convex region, termed the *capacity region*, determined for a given broadcast channel $P(\mathcal{Y}_1,\mathcal{Y}_2|\mathcal{X})$ using an infinite code length N [3]. Evaluation of the capacity region is one of the fundamental problems in multi-user information theory; the problem is difficult and has not yet been solved in general except for a few special cases.

A broadcast channel $P(\mathcal{Y}_1,\mathcal{Y}_2|\mathcal{X})$ is termed *degraded* if there exists a distribution $P'(\mathcal{Y}_2|\mathcal{Y}_1)$ such that

$$P(\mathcal{Y}_2|\mathcal{X}) = \sum_{\{\mathcal{Y}_1\}} P'(\mathcal{Y}_2|\mathcal{Y}_1)P(\mathcal{Y}_1|\mathcal{X}), \tag{1}$$

and is termed *physically degrated* if $P(Y_2|\mathcal{X}) = P(\mathcal{Y}_2|\mathcal{Y}_1)P(\mathcal{Y}_1|\mathcal{X})$. The stochastically and physically degraded channel models are commonly used in the literature [3] and merely indicate that the corruption *process can be viewed* as a two stage process, where the more corrupted received codeword $\mathcal{Y}_2$ can be regarded as a further degradation of the less corrupted codeword $\mathcal{Y}_1$. It does not imply that any actual communication between the two receivers is required. Furthermore, this assumption can also represent realistic scenarios; for instance, where the channel noise rate depends on the distance from the broadcaster, a natural assumption for both wired and wireless communication. In this case, the channel model for the second receiver $[P(\mathcal{Y}_2|\mathcal{X})]$ can be *formally* expressed as if the message were conveyed via the first receiver, who is closer to the broadcaster [Eq. (1)] although the two receivers do not actually communicate. A pictorial explanation is also provided in Fig. 1(c).

The degraded channel is exceptional in the sense that its capacity region can be analytically obtained as the convex hull of the closure of all points (R_1,R_2) that satisfy

$$R_2<I(\mathcal{U};\mathcal{Y}_2),$$
$$R_1<I(\mathcal{X};\mathcal{Y}_1|\mathcal{U}) \tag{2}$$

for a certain joint distribution $P(\mathcal{U})P(\mathcal{X}|\mathcal{U})P(\mathcal{Y}_1,\mathcal{Y}_2|\mathcal{X})$; where the auxiliary random variable $\mathcal{U}$ has a cardinality bounded by $|\mathcal{U}|\leqslant\min\{|\mathcal{X}|,|\mathcal{Y}_1|,|\mathcal{Y}_2|\}$. This region is often called Cover's capacity [1] region. Unfortunately, the derivation of Cover's capacity is nonconstructive and offers little clue to the design of efficient practical codes. Thus, practical codes for the degraded broadcast channel have been actively investigated in the network information theory literature [4].

In the case of binary symmetric channels characterized by flip probabilities p_1 and p_2, condition (1) reduces to an inequality $p_2>p_1$. Then, the expression of Cover's capacity is simplified to

$$R_2<1-H_2(\delta * p_2)$$
$$R_1<H_2(\delta * p_1)-H_2(p_1), \tag{3}$$

where a parameter $0<\delta<1$ specifies the optimal ratio between R_1 and R_2; $\delta * p=\delta(1-p)+(1-\delta)p$ and $H_2(p)$ is Shannon's entropy $H_2(p)=-p\log_2 p-(1-p)\log_2(1-p)$.

The solid convex curve in Fig. 1(b) shows Cover's limit, i.e., the boundary of Cover's capacity for the binary symmetric channels. The straight dotted line corresponds to the time sharing capacity, i.e., the achievable capacity by concatenating two independent codes words optimized for each channel separately. This is realized by using $N(1-\alpha)$ and $N\alpha$ bits of the code word $\mathcal{X}$ for encoding messages $\mathcal{W}_1$ and $\mathcal{W}_2$, respectively. Here, $0<\alpha<1$ is the code length ratio between the two messages. This simple concatenation and the limit achievable by this scheme are often termed the *time sharing* and the *time sharing limit*, respectively. The difference between Cover's and the time sharing limits indicates the capacity gain obtained by optimizing a code for the complete broadcasting system optimizing each of the channels separately.

We have to emphasize that achieving the time sharing limit *in practice* is never trivial as there is no known practical coding scheme that saturates Shannon's limit even for a single channel. Therefore, the design of improved practical codes for broadcasting, by combining existing codes, devised for single channels, is an important research topic in coding theory [4].

III. LINEARLY COMBINED CODES

A linearly combined code is a well-known strategy for designing high performance communication schemes for broadcast channels using multiple linear error-correcting codes [9,8]. In this scheme, the first $N(1-\alpha)$ bits of a code word are obtained by linearly mixing two messages $\mathcal{W}_1$ and $\mathcal{W}_2$ while the other $N\alpha$ bits are generated only from $\mathcal{W}_2$ by some linear transformation. In both coding and decoding, all operations are typically carried out in modulo 2. This method has been developed for algebraic codes, such as Reed-Solomon and BCH, which are standard codes designed for relatively short code lengths. For these codes, it is reported that the minimum distance between code words is larger than that achieved in the time sharing scheme, which implies higher robustness against channel noise [9,8].

However, it is unclear whether a similar construction also offers better performance when different code types are used. Furthermore, it is theoretically interesting and important to examine whether a linearly combined code can saturate Cover's limit for infinite code length (N) or not.

Motivated by these questions, we investigate here the ability and limitations of linearly combined LDPC codes in the limit $N\to\infty$.

An LDPC code is characterized by a parity-check matrix. To devise a linearly combined coding scheme for LDPC codes, we define a parity-check matrix in an upper triangular form

$$A=\begin{pmatrix} A_1 & A_2 \\ 0 & A_3 \end{pmatrix}, \tag{4}$$

where the sizes of the submatrices A_1, A_2, A_3 are $[(1-\alpha)N-R_1N]\times(1-\alpha)N$, $[(1-\alpha)N-R_1N]\times\alpha N$, and $[\alpha N-R_2N]\times\alpha N$, respectively. Furthermore, we assume that A_1,A_2,A_3 have K_1,K_2,K_3 and C_1,C_2,C_3 nonzero elements per row and column, respectively. Based on the parity-check matrix, the generator matrix G^T is constructed as

$$G^T=\begin{pmatrix} G_1^T & G_2^T \\ 0 & G_3^T \end{pmatrix}, \tag{5}$$

where G_i^T $(i=1,3)$ are constructed systematically to satisfy the constraints $A_iG_i^T=0$ (modulo 2) and G_2^T is defined as $-A_1^T[A_1A_1^T]^{-1}[A_2G_3^T]$. The sizes of these matrices are $(1-\alpha)N\times R_1N$, $(1-\alpha)N\times R_2N$, and $\alpha N\times R_2N$, respectively.

The sender produces a code word $\mathcal{X}$ by taking a product of the generator matrix G^T and the original messages $(\mathcal{W}_1,\mathcal{W}_2)^T$. Receiving a possibly corrupted code word, each receiver evaluates the syndrome vectors $\boldsymbol{J}_i=A\mathcal{Y}_i (i=1,2)$, which yield the parity-check equation $\boldsymbol{J}_i=A\boldsymbol{\xi}_i$. The noise vector $\boldsymbol{\xi}_i$ can be thought of as having two separate segments denoted by u (up) and d (down) later on. The parameter α controls the error-correction ability for the second message; the transmitted information redundancy increases with α. The decoding problem for each receiver is to find the most probable noise vectors , $\boldsymbol{s}_i$ and $\boldsymbol{\sigma}_i$, such that the parity-check equation

$$\boldsymbol{J}_i=\begin{pmatrix} A_1 & A_2 \\ 0 & A_3 \end{pmatrix}\begin{pmatrix} \boldsymbol{s}_i \\ \boldsymbol{\sigma}_i \end{pmatrix} \quad (i=1,2) \tag{6}$$

is obeyed, and using prior knowledge about the two noise vectors characterized by the two different channels.

The second receiver has to estimate only the lower part of the noise vector $\boldsymbol{\xi}^{\mathrm{d}}$, which can be carried out using only the lower part of Eq. (6). However, we assume here that both receivers independently solve Eq. (6) using prior knowledge on their own channels since one can show that solving the whole equation provides optimal estimation performance for both receivers. As Eq. (6) has the same form for receivers 1 and 2, we hereafter omit the subscript $i=1,2$.

It should be emphasized here that the upper triangular architecture in the parity-check matrix A is suitable for providing a higher-error correction ability to the second message $\mathcal{W}_2$ which is more degraded according to Eq. (1) than the other message $\mathcal{W}_1$; consequentially, $\boldsymbol{\xi}^{\mathrm{d}}$ can be estimated independent of $\boldsymbol{\xi}^{\mathrm{u}}$, while estimation of $\boldsymbol{\xi}^{\mathrm{u}}$ fails unless $\boldsymbol{\xi}^{\mathrm{d}}$ is correctly retrieved.

For bitwise minimization of the error probability, the optimal estimation is given by maximizing the posterior marginal (MPM)

$$\hat{\xi}_i^{\mathrm{u}}=\mathrm{argmax}_{s_i\in\{0/1\}}P(s_i|\boldsymbol{J}),\quad \hat{\xi}_j^{\mathrm{d}}=\mathrm{argmax}_{\sigma_j\in\{0/1\}}P(\sigma_j|\boldsymbol{J}). \tag{7}$$

An exact evaluation of Eq. (7) is generally hard; therefore, the belief propagation (BP) approximation scheme is widely used as a practical decoding algorithm. The latter has been shown to be identical to the Thouless-Anderson-Palmer (TAP) approach in the current case [14–19].

IV. STATISTICAL MECHANICS

A. Macroscopic analysis—performance evaluation

In order to evaluate the typical error-correction ability of these codes in the limit $N\to\infty$, we investigate the behavior of the MPM decoder using the established methods of statistical mechanics. We first map the current system to an Ising spin model with finite connectivity by employing the binary representation $\{+1,-1,\times\}$ for the alphabet and operator instead of the Boolean one $\{0,1,+\}$. This implies that the posterior probability $P(\boldsymbol{s},\boldsymbol{\sigma}|\boldsymbol{J})$ can be expressed as a Boltzmann distribution at the inverse temperature $\beta=1$ using a Hamiltonian

$$\begin{aligned}\mathcal{H}(\boldsymbol{s},\boldsymbol{\sigma}|\boldsymbol{J})=\lim_{\gamma\to\infty}\Bigg\{&\gamma\sum_{\{\mathcal{I}(K_1),\mathcal{J}(K_2)\}}D^{1,2}_{\mathcal{I}(K_1),\mathcal{J}(K_2)}\\&\times\delta\Bigg(-J^{\mathrm{u}}_{\mathcal{I}(K_1),\mathcal{J}(K_2)};\prod_{i\in\mathcal{I}(K_1)}s_i\prod_{j\in\mathcal{J}(K_2)}\sigma_j\Bigg)\\&+\gamma\sum_{\{\mathcal{J}(K_3)\}}D^3_{\mathcal{J}(K_3)}\delta\Bigg(-J^{\mathrm{d}}_{\mathcal{J}(K_3)};\prod_{j\in\mathcal{J}(K_3)}\sigma_j\Bigg)\Bigg\}\\&-F\sum_{i=1}^{(1-\alpha)N}s_i-F\sum_{j=1}^{\alpha N}\sigma_j,\end{aligned} \tag{8}$$

where $\mathcal{I}(K)=\langle i_1,i_2,\ldots,i_K\rangle$ denotes the combination of the K subscripts chosen from the $i=1,2,\ldots,(1-\alpha)N$ possibilities without duplication (the order is ignored), and $\mathcal{J}(K)=\langle j_1,j_2,\ldots,j_K\rangle$ is the K subscripts combination from $j=1,2,\ldots,\alpha N$ chosen similarly. The tensor $D^{1,2}_{\mathcal{I}(K_1),\mathcal{J}(K_2)}$ becomes 1 when its subscripts agree with the positions of nonzero elements in the parity-check matrices A_1 and A_2, and 0 otherwise. The tensor $D^3_{\mathcal{J}(K_3)}$ similarly corresponds to A_3. The first and second terms in the Hamiltonian (8) correspond to Eq. (6), while the third and fourth terms are provided by the prior distribution of the noise. The field F represents the channel noise level; it is set to $\frac{1}{2}\ln(1-p_1)/p_1$ and $\frac{1}{2}\ln(1-p_2)/p_2$ for the first and the second receivers, respectively.

In order to simplify the calculation, we first employ the gauge transformation $s_i\to s_i\xi_i^{\mathrm{u}},\sigma_j\to\sigma_j\xi_j^{\mathrm{d}}$, $J^{\mathrm{u}}_{\ldots}\to 1$, and $J^{\mathrm{d}}_{\ldots}\to 1$, which reduces complicated couplings expressed in the first and second terms in Hamiltonian (8) to simple ferromagnetic interactions.

As the parity-check matrices and noise vectors are generated randomly, we have to perform averages over these variables for extracting typical properties of the code. This can be carried out by the replica method $-\beta\mathcal{F}=1/N\langle\ln\mathcal{Z}\rangle_{A,\boldsymbol{\xi}^{\mathrm{u}},\boldsymbol{\xi}^{\mathrm{d}}}=\lim_{n\to 0}(1/nN)\ln\langle\mathcal{Z}^n-1\rangle_{A,\boldsymbol{\xi}^{\mathrm{u}},\boldsymbol{\xi}^{\mathrm{d}}}$, where Z is the partition function and $\langle\cdots\rangle_{A,\boldsymbol{\xi}^{\mathrm{u}},\boldsymbol{\xi}^{\mathrm{d}}}$ represents an average over the parity-check matrix A and the noise vectors $\boldsymbol{\xi}^{\mathrm{u}}$ and $\boldsymbol{\xi}^{\mathrm{d}}$ (i.e., the quenched variables). This gives rise to three sets of order parameters

$$\begin{aligned}q_{\{a_1,a_2,\ldots,a_m\}}&=\frac{1}{N}\sum_{i=1}^{(1-\alpha)N}X_is_i^{a_1}\ldots s_i^{a_m},\\ r_{\{a_1,a_2,\ldots,a_m\}}&=\frac{1}{N}\sum_{j=1}^{\alpha N}Y_j\sigma_j^{a_1}\ldots\sigma_j^{a_m},\\ t_{\{a_1,a_2,\ldots,a_m\}}&=\frac{1}{N}\sum_{j=1}^{\alpha N}Z_j\sigma_j^{a_1}\ldots\sigma_j^{a_m},\end{aligned} \tag{9}$$

where $a_1,a_2,\ldots,a_m$ denote the replica indices running from 1 to n, and their conjugates $\hat{q}_{\{a_1,a_2,\ldots,a_m\}}$, $\hat{r}_{\{a_1,a_2,\ldots,a_m\}}$, $\hat{t}_{\{a_1,a_2,\ldots,a_m\}}$. The variables Z_j are introduced to express the constraint of the parity-check matrix A_3 as

$$\begin{aligned}&\delta\Bigg(\sum_{\{\mathcal{J}(K_3)\setminus j\}}D^3_{\langle j,\mathcal{J}(K_3),\setminus j\rangle}-C_3\Bigg)\\&\quad=\oint\frac{dZ_j}{2\pi}Z_j^{\sum_{\{\mathcal{J}(K_3)\setminus j\}}D^3_{\langle j,\mathcal{J}(K_3),\setminus j\rangle}-(C_3+1)}.\end{aligned} \tag{10}$$

The variables X_i and Y_j are similarly introduced for A_1 and A_2.

In order to proceed further, one has to make an assumption about the symmetry of replica indices. Here we employ the simplest replica symmetric (RS) ansatz, expressed in the current case by $q_{\{a_1,\ldots,a_m\}}=q_0\int dx\,\pi(x)x^m$, $r_{\{a_1,\ldots,a_m\}}=r_0\int dy\,\rho(y)y^m$, $t_{\{a_1,\ldots,a_m\}}=t_0\int dz\,\phi(z)z^m$, where q_0, r_0, and t_0 are the normalization constants to make $\pi(x)$, $\rho(y)$, and $\phi(z)$ proper probability distributions over the interval $[-1,1]$, respectively. Unspecified integrals are performed over $[-1,1]$. We also assume a similar ansatz for the conjugate variables. A further complicated assumption about the order parameter symmetry is generally required in most disordered systems [20,21]. However, the validity of the RS ansatz in the current system is strongly supported by a recent report on the absence of the replica symmetry breaking (in the dominant state) in gauged systems where Nishimori's temperature is used [22]. The latter corresponds to using the correct priors in decoding [23], as performed in the current analysis.

Under these assumptions, one obtains the free energy

$$
\begin{aligned}
\mathcal{F}=&(1-R_1-R_2)\ln 2-(1-\alpha-R_1)\\
&\times\left\langle \ln\left(1+\prod_{l=1}^{K_1}x_l\prod_{l'=1}^{K_2}y_{l'}\right)\right\rangle_{\pi^{K_1},\rho^{K_2}}-(\alpha-R_2)\\
&\times\left\langle \ln\left(1+\prod_{l=1}^{K_3}z_l\right)\right\rangle_{\phi^{K_3}}+(1-\alpha)C_1\langle\ln(1+x\hat{x})\rangle_{\pi,\hat{\pi}}\\
&+\alpha C_2\langle\ln(1+y\hat{y})\rangle_{\rho,\hat{\rho}}+\alpha C_3\langle\ln(1+z\hat{z})\rangle_{\phi,\hat{\phi}}+(1-\alpha)\\
&\times\left\langle \ln\left[\mathrm{Tr}_s e^{s\xi^{\mathrm{u}}F}\prod_{l=1}^{C_1}(1+s\hat{x}_l)\right]\right\rangle_{\xi,\hat{\pi}^{C_1}}\\
&+\alpha\left\langle \ln\left[\mathrm{Tr}_\sigma e^{\sigma\xi^{\mathrm{d}}F}\prod_{l=1}^{C_2}(1+\sigma\hat{y}_l)\right.\right.\\
&\left.\left.\times\prod_{l'=1}^{C_3}(1+\sigma\hat{z}_{l'})\right]\right\rangle_{\xi,\hat{\rho}^{C_2},\hat{\phi}^{C_3}},
\end{aligned}
\tag{11}
$$

where $\langle\cdots\rangle_{P^K}$ denotes an integral of the form $\int\prod_{k=1}^{K}dx_kP(x_k)(\cdots)$ and $\langle f(\xi)\rangle_\xi=(1-p)f(+1)+pf(-1)$.

Varying Eq. (11), one obtains a set of saddle-point equations,

$$
\pi(x)=\left\langle \delta\left(x-\tanh\left[\sum_{l=1}^{C_1-1}\tanh^{-1}\hat{x}_l+\xi^{\mathrm{u}}F\right]\right)\right\rangle_{\xi,\hat{\pi}^{C_1-1}},
$$

$$
\begin{aligned}
\rho(y)=&\left\langle \delta\left(y-\tanh\left[\sum_{l=1}^{C_2-1}\tanh^{-1}\hat{y}_l\right.\right.\right.\\
&\left.\left.\left.+\sum_{l'=1}^{C_3}\tanh^{-1}\hat{z}_{l'}+\xi^{\mathrm{d}}F\right]\right)\right\rangle_{\xi,\hat{\rho}^{C_2-1},\hat{\phi}^{C_3}},
\end{aligned}
$$

$$
\begin{aligned}
\phi(z)=&\left\langle \delta\left(z-\tanh\left[\sum_{l=1}^{C_2}\tanh^{-1}\hat{y}_l\right.\right.\right.\\
&\left.\left.\left.+\sum_{l'=1}^{C_3-1}\tanh^{-1}\hat{z}_{l'}+\xi^{\mathrm{d}}F\right]\right)\right\rangle_{\xi,\hat{\rho}^{C_2},\hat{\phi}^{C_3-1}},
\end{aligned}
$$

$$
\hat{\pi}(x)=\left\langle \delta\left(\hat{x}-\prod_{l=1}^{K_1-1}x_l\prod_{l'=1}^{K_2}y_{l'}\right)\right\rangle_{\pi^{K_1-1},\rho^{K_2}},
$$

$$
\hat{\rho}(y)=\left\langle \delta\left(\hat{y}-\prod_{l=1}^{K_1}x_l\prod_{l'=1}^{K_2-1}y_{l'}\right)\right\rangle_{\pi^{K_1},\rho^{K_2-1}},
$$

$$
\hat{\phi}(z)=\left\langle \delta\left(\hat{z}-\prod_{l=1}^{K_3-1}z_l\right)\right\rangle_{\phi^{K_3-1}}. \tag{12}
$$

The overlaps $M_{\mathrm{u}}=1/(1-\alpha)N\Sigma_i\hat{s}_i\xi_i^{\mathrm{u}}$ and $M_{\mathrm{d}}=(1/\alpha N)\Sigma_j\hat{\sigma}_j\xi_j^{\mathrm{d}}$ serve as performance measures for the error-correcting ability. After solving the saddle-point equations (12), these can be calculated as

$$
M_{\mathrm{u}}=\int dh\,h_{\mathrm{eff}}^{\mathrm{u}}(h)\,\mathrm{sgn}(h),\quad M_{\mathrm{d}}=\int dh\,h_{\mathrm{eff}}^{\mathrm{d}}(h)\,\mathrm{sgn}(h), \tag{13}
$$

where distributions of effective fields $h_{\mathrm{eff}}(h)$ are evaluated as

$$
h_{\mathrm{eff}}^{\mathrm{u}}(h)=\left\langle \delta\left(h-\tanh\left[\sum_{l=1}^{C_1}\tanh^{-1}\hat{x}_l+\xi F\right]\right)\right\rangle_{\xi,\hat{\pi}^{C_1}}
$$

$$
\begin{aligned}
h_{\mathrm{eff}}^{\mathrm{d}}(h)=&\left\langle \delta\left(h-\tanh\left[\sum_{l=1}^{C_2}\tanh^{-1}\hat{y}_l\right.\right.\right.\\
&\left.\left.\left.+\sum_{l'=1}^{C_3}\tanh^{-1}\hat{z}_{l'}+\xi F\right]\right)\right\rangle_{\xi,\hat{\rho}^{C_2},\hat{\phi}^{C_3}}.
\end{aligned}
\tag{14}
$$

B. Microscopic analysis—practical decoding

As already mentioned, it is computationally hard to perform MPM decoding (7) exactly. Instead, the BP algorithm [14] is widely used for a practical decoding in LDPC codes. Belief propagation has recently been shown to be equivalent to the Bethe method [15,16], in general, and to provide an equivalent result of the TAP approach [17], in particular, for spin glass models [18,19]. Since the current system is somewhat similar to spin glass models, we use a term BP-TAP for referring to this scheme from now on.

The BP-TAP approach offers an iterative algorithm to approximately evaluate marginal posterior distributions based on local dependencies between syndrome and variables. These local dependencies can be uniquely identified with conditional probabilities. In the current system, these become $q_{\mu l}^n=P(n_l=n|\{\boldsymbol{J}/J_\mu\})$ and $\hat{q}_{\mu l}^n\propto P(J_\mu|n_l=n,\{\boldsymbol{J}/J_\mu\})$, where n_l and J_μ represent components of spin variables $\boldsymbol{s}$, $\boldsymbol{\sigma}$, and syndrome $\boldsymbol{J}$, respectively; $\{\boldsymbol{J}/J_\mu\}$ denotes the set of syndrome bits excluding μth component. As most syndrome and spin variables are not directly related, we assign the conditional probabilities only to pairs μl that have nonzero elements in the parity-check matrix A.

Evaluating the two types of conditional probabilities using directly connected components, the BP-TAP algorithm can be generally expressed as

$$
q_{\mu l}^n=\alpha_{\mu l}e^{Fn}\prod_{\nu\in\mathcal{M}(l)\setminus\mu}\hat{q}_{\nu l}^n, \tag{15}
$$

$$
\hat{q}_{\mu l}^n=\hat{\alpha}_{\mu l}\sum_{\{n_{j\in\mathcal{L}(\mu)\setminus l}\}}\delta\left(J_\mu;n\prod_{j\in\mathcal{L}(\mu)\setminus j}n_j\right)\prod_{j\in\mathcal{L}(\mu)\setminus l}q_{\mu j}^{n_j}, \tag{16}
$$

where $\mathcal{M}(l)$ and $\mathcal{L}(\mu)$ denote the sets of syndrome and spin variable indices that are directly linked to spin and syndrome indices l and μ, respectively; $\mathcal{M}(l)\setminus\mu$ represents the set of indices $\nu\in\mathcal{M}(l)$ excluding μ and similarly for $\mathcal{L}(\mu)\setminus l$ and other sets. Normalization constants, $\alpha_{\mu l}$ and $\hat{\alpha}_{\mu l}$, are introduced to make $q^n_{\mu l}$ and $\hat{q}^n_{\mu l}$ probability distributions of the spin variable n. A field F is introduced to represent the prior probability.

Since spin variable n takes only two values ± 1, it is convenient to express the BP-TAP algorithm using spin averages $\Sigma_{n=\pm 1} n q^n_{\mu l}$ and $\Sigma_{n=\pm 1} n\hat{q}^n_{\mu l}$ rather than the distributions $q^n_{\mu l}$ and $\hat{q}^n_{\mu l}$ themselves. As the parity-check matrix A is structured, it may be useful to assign a different notation to the spin averages according to the submatrix, to which the pair of indices μl belongs to. We use $x_{\mu l}, y_{\mu l}$, and z_μ to denote $\Sigma_{n=\pm 1} n q^n_{\mu l}$ when the pair of indices μl belongs to A_1, A_2, and A_3, respectively. Similar notations $\hat{x}_{\mu l}, \hat{y}_{\mu l}$, and $\hat{z}_{\mu l}$ are used for $\Sigma_{n=\pm 1} n\hat{q}^n_{\mu l}$. Then, the BP-TAP algorithms (15) and (16), which are expressed as a set of functional equations, are reduced to a couple of nonlinear equations

$$x_{\mu l}=\tanh\left[\sum_{\nu\in A_1^{\mathrm{col}}(l)/\mu}\tanh^{-1}\hat{x}_{\nu l}+F\right],$$

$$y_{\mu l}=\tanh\left[\sum_{\nu\in A_2^{\mathrm{col}}(l)/\mu}\tanh^{-1}\hat{y}_{\nu l}+\sum_{\nu\in A_3^{\mathrm{col}}(l)}\tanh^{-1}\hat{z}_{\nu l}+F\right],$$

$$z_{\mu l}=\tanh\left[\sum_{\nu\in A_2^{\mathrm{col}}(l)}\tanh^{-1}\hat{y}_{\nu l}+\sum_{\nu\in A_3^{\mathrm{col}}(l)/\mu}\tanh^{-1}\hat{z}_{\nu l}+F\right],$$

$$\hat{x}_{\mu l}=\mathrm{sgn}(J_\mu)\prod_{i\in A_1^{\mathrm{row}}(\mu)/l}x_{\mu i}\prod_{j\in A_2^{\mathrm{row}(\mu)}}y_{\mu j},$$

$$\hat{y}_{\mu l}=\mathrm{sgn}(J_\mu)\prod_{i\in A_1^{\mathrm{row}}(\mu)}x_{\mu i}\prod_{j\in A_2^{\mathrm{row}(\mu)}/l}y_{\mu j},$$

$$\hat{z}_{\mu l}=\mathrm{sgn}(J_\mu)\prod_{j\in A_3^{\mathrm{row}}(\mu)/l}z_{\mu j}, \tag{17}$$

where $A^{\mathrm{row}}(\mu)$ and $A^{\mathrm{col}}(l)$ denote sets of nonzero elements in the μth row and ith column of matrix A, respectively.

Equations (17) can be solved iteratively from appropriate initial conditions (prior means are usually chosen as initial states). Less than 50 iterations are typically sufficient for convergence. After obtaining the solutions, approximated posterior means can be calculated

$$\langle s_i\rangle=\tanh\left[\sum_{\nu\in A_1^{\mathrm{col}}(i)}\tanh^{-1}\hat{x}_{\nu i}+F\right],$$

$$\langle\sigma_j\rangle=\tanh\left[\sum_{\nu\in A_2^{\mathrm{col}}(j)}\tanh^{-1}\hat{y}_{\nu j}+\sum_{\nu\in A_3^{\mathrm{col}}(j)}\tanh^{-1}\hat{z}_{\nu j}+F\right], \tag{18}$$

which provides the MPM estimators $\hat{s}_i=\mathrm{sgn}(\langle s_i\rangle)$ and $\hat{\sigma}_j=\mathrm{sgn}(\langle\sigma_j\rangle)$.

It can be shown that the BP-TAP framework provides an exact result when the global structure of the connectivities is graphically expressed by a tree [14]. Unfortunately, it is still unclear how good are the approximations obtained when a given system does not admit a tree architecture.

The graphical architecture of LDPC codes generally has many loops, which implies that the BP-TAP framework does not necessarily offer a good approximation. However, it is conjectured, and partially confirmed, that a nearly exact result can be obtained, as long as no other locally stable solution exists, when the parity-check matrix A is randomly constructed and in the limit $N\to\infty$; this is due to the fact that the typical loop length scales as $O(\ln N)$ for randomly constructed matrices, which implies that LDPC codes can be locally treated as trees ignoring the effect of loops [24].

Neglecting the effect of loops naturally leads to a macroscopic description of the BP-TAP algorithm (17) utilizing density functions of messages $x_{\mu l}$, $y_{\mu l}$, $z_{\mu l}$, $\hat{x}_{\mu l}$, $\hat{y}_{\mu l}$, and $\hat{z}_{\mu l}$, which becomes identical to the simple iteration of the saddle-point equation (12) [24]. Surprisingly, the celebrated method known as the *density evolution* (DE) [13], recently discovered independently in the information theory community, reduces exactly to the same equation (12). As both the DE and the current analysis reduce to an identical equation (12), the estimates provided by the two frameworks generally coincide for the practical performance. However, as the concept of free energy is missing from the DE framework, it does not provide a way for evaluating the optimal (theoretical) performance, for a given code; this is naturally characterized, in the statistical physics framework by thermodynamical transitions between decoding success and failure phases.

V. RESULTS

In order to theoretically examine the typical performance that can be obtained by the linearly combined coding scheme, we solved the saddle-point equations (12). Since solving the equations analytically is generally difficult, we mainly resorted to numerical methods. The solutions were obtained by iterating the saddle-point equations (12), and approximating the distributions by $O(10^4)$ sample vectors. Less than 50 iterations were typically sufficient for obtaining a solution.

Solving the equations for several parameter sets, assuming $\alpha > R_2/(R_1+R_2)$, we found that the solutions can be classified into three categories depending on whether overlaps M_{u} and M_{d} are 1 or not. The first one is referred to the *ferromagnetic* (FM) solution ($M_{\mathrm{u}}=M_{\mathrm{d}}=1$) corresponding to a perfect retrieval for both messages $\mathcal{W}_1$ and $\mathcal{W}_2$. The *half-ferromagnetic* (HFM) solution which is characterized by $M_{\mathrm{u}}\neq 1$ and $M_{\mathrm{d}}=1$ implies that only the second message $\mathcal{W}_2$ is perfectly retrieved, while $\mathcal{W}_1$ is not. The last category, termed paramagnetic (PM) solution, describes a decoding failure for both messages being characterized by $M_{\mathrm{u}}\neq 1, M_{\mathrm{d}}\neq 1$. The ferromagnetic solution always exists and is locally stable for $C_1\geqslant 3$ and $C_3\geqslant 3$, while one can find other solu-

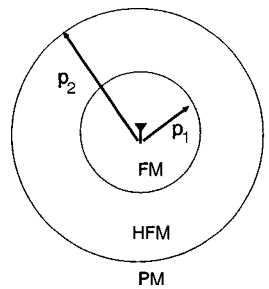

FIG. 2. Areas corresponding to FM, HFM, and PM for a wireless degraded channel assuming that the noise corruption rate grows proportionally to the distance from a broadcast station (sender). For $\alpha > R_2/(R_1+R_2)$ (left) , the area where $\mathcal{W}_2$ can be perfectly retrieved becomes broader than for $\alpha < R_2/(R_1+R_2)$ (right) because of the existence of the HFM solution.

tions only for relatively higher noise levels. As the noise level increases, HFM and PM solutions emerge in this order.

The HFM solution may look counterintuitive at first because the corruption process for the second receiver is expressed as Eq. (1), which gives the impression that the code word was relayed to the second receiver via the first one. Retrieval of $\mathcal{W}_2$ would therefore fail unless $\mathcal{W}_1$ has been correctly decoded. However, one should keep in mind that the two receivers carry out two different tasks; the first receiver has to retrieve more information from the slightly corrupted code word, while the second receiver retrieves less information from a more degraded message. A failure of the first receiver in decoding *both* components of the message, $\boldsymbol{\xi}^{\mathrm{u}}$ and $\boldsymbol{\xi}^{\mathrm{d}}$, does not provide any information on its ability (or the ability of receiver 2) to successfully decode part of the message $\boldsymbol{\xi}^{\mathrm{d}}$. In addition, the current code based on an upper triangular parity-check matrix is designed to provide a higher-error-correction ability for $\mathcal{W}_2$ as it has to be transmitted to a farther distance and relatively more resource is assigned for $\mathcal{W}_2$ in construction of a code word $\mathcal{X}$ for $\alpha > R_2/(R_1+R_2)$, makes it possible to produce the nontrivial solution HFM. For $\alpha < R_2/(R_1+R_2)$, on the other hand, we found only two solutions: FM and PM. A pictorial explanation is provided in Fig. 2. The solution that has the lowest free energy among the three becomes thermodynamically dominant. As the noise level p becomes higher (or the field F becomes weaker), the dominant state changes from FM to HFM and PM in this order. Since receivers are required to retrieve only their own messages, the transition point between HFM and PM corresponds to the maximum noise level for error-free communication in the second channel, while maximum noise level for the first channel is given by the transition point between FM and HFM.

However, this does not imply a successful decoding up to the critical points in *practical* time scales. Practical perfect decoding by the BP-TAP algorithm is possible only when no suboptimal solutions exist, which means that the practically achievable limit is given by the *spinodal points* of the HFM and PM solutions for the first and the second channels, respectively; i.e., the point where new suboptimal solutions emerge. A similar phenomenon has been reported before for similar systems [10,11].

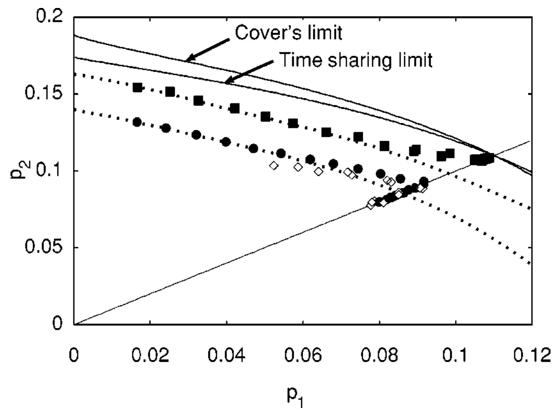

FIG. 3. Optimal and practical performance of the MPM decoder calculated by methods of statistical mechanics for different α values. For the first channel, the optimal performance is given by the thermodynamical transition between FM and HFM solutions, while the transition between HFM and PM solutions marks the optimal performance for the second channel. On the other hand, the practical performance is given by the spinodal points of the HFM and PM solutions for the first and the second channels, respectively. Monte Carlo solutions based on 10^4 sample vectors were employed for solving the saddle-point equation (12). The standard deviation values resulting from ten trials are smaller than the symbol size. The black squares and the black circles denote the optimal and the practical performances for the linearly combined coding scheme, where code parameters are set to $C_1 = C_3 = 3, C_2 = 4$, $R_1 = R_2 = 1/4$. Diamond symbols denote the maximum noise levels for decoding success by the BP-TAP algorithm, determined from 50 experiments. The error bars are smaller than the symbol size. Broken lines denote the optimal and practical performances of the time sharing for corresponding LDPC codes. The two lines in the upper right are Cover's and time sharing capacities calculated in the information theory literature.

Figure 3 shows the maximum noise levels for perfect decoding of the linearly combined coding method obtained for $C_2 = 4$ and 0 fixing $C_1 = C_3 = 3$; $C_2 = 0$ corresponds to the time sharing scheme, for which $A_2 = 0$. One can find that both optimal and practical performances of the MPM decoder are improved by the introduction of the additional submatrix A_2, as anticipated, in spite of the fact that the parameter $C_2 (=4)$ is not optimally tuned. This result may induce the hope that Cover's limit can be saturated by optimally tuning the submatrices. However, our analysis contradicts this conjecture. Solving Eq. (12) in the limit $C_3 \to \infty$ and C_1 or $C_2 \to \infty$ is feasible; it is known that the MPM decoder provides the optimal performance in this limit, while practical BP-TAP decoding becomes difficult. The three solutions correspond to those already mentioned before, but can be analytically expressed as the following.

(1) *FM solution:* Both messages are decodable ($M_{\mathrm{u}} = M_{\mathrm{d}} = 1$). The corresponding solutions and free energy are

$$\begin{gathered} \pi(x) = \delta(x-1), \quad \hat{\pi}(\hat{x}) = \delta(\hat{x}-1), \\ \rho(y) = \delta(y-1), \quad \hat{\rho}(\hat{y}) = \delta(\hat{y}-1), \\ \phi(z) = \delta(z-1), \quad \hat{\phi}(\hat{z}) = \delta(\hat{z}-1), \\ \mathcal{F} = -(1-2p)F. \end{gathered} \tag{19}$$

(2) *HFM solution:* Message $\mathcal{W}_2$ is only decodable ($M_{\mathrm{u}} \neq 1, M_{\mathrm{d}} = 1$).

$$\pi(x) = \langle \delta(x - \tanh \xi F) \rangle_{\xi}, \quad \hat{\pi}(\hat{x}) = \delta(\hat{x}),$$
$$\rho(y) = \delta(y-1), \quad \hat{\rho}(\hat{y}) = \delta(\hat{y}),$$
$$\phi(z) = \delta(z-1), \quad \hat{\phi}(\hat{z}) = \delta(\hat{z}-1), \tag{20}$$
$$\mathcal{F} = (1-\alpha-R_1)\ln 2 - (1-2p)F - (1-\alpha)\ln 2 H_2(p).$$

(3) *PM solution:* Both messages are not decodable ($M_{\mathrm{u}} \neq 1, M_{\mathrm{d}} \neq 1$).

$$\pi(x) = \langle \delta(x - \tanh \xi F) \rangle_{\xi}, \quad \hat{\pi}(\hat{x}) = \delta(\hat{x}),$$
$$\rho(y) = \langle \delta(y - \tanh \xi F) \rangle_{\xi}, \quad \hat{\rho}(\hat{y}) = \delta(\hat{y}),$$
$$\phi(z) = \langle \delta(z - \tanh \xi F) \rangle_{\xi}, \quad \hat{\phi}(\hat{z}) = \delta(\hat{z}), \tag{21}$$
$$\mathcal{F} = (1-R_1-R_2)\ln 2 - (1-2p)F - \ln 2 H_2(p).$$

Examining the critical condition for decoding success in each channel, and comparing the free energy of the solutions, one obtains the capacity region of the linearly combined coding scheme

$$R_2 < \alpha[1 - H(p_2)],$$
$$R_1 < (1-\alpha)[1 - H(p_1)]. \tag{22}$$

This is, unfortunately, identical to the time sharing capacity that can be achieved by a simple concatenation of two independent codes. This result implies that the advantage of the linearly combined coding scheme vanishes as the submatrices become dense and this method cannot saturate Cover's limit.

VI. SUMMARY AND CONCLUSION

In this paper, we have examined the performance of linearly combined LDPC codes for information transmission in a broadcast channel. Our analysis shows that the capacity of the suggested coding scheme is upper bound by the time sharing capacity, in spite of the apparent improvement in both optimal and practical performance with respect to LDPC based time sharing codes characterized by finite connectivity values.

The reason for the failure of linearly combined LDPC codes to saturate Cover's limit may be explained by the code word structure produced by this scheme. In his proof, Cover optimized the code performance by introducing a specific structure termed the *cloud coding*, employing an auxiliary random variable $\mathcal{U}$ as in Eq. (2). In cloud coding, a code word $\mathcal{X}$ is randomly generated according to $P(\mathcal{X}|\mathcal{U})$ around a *cloud center* $\mathcal{U}$ sampled from $P(\mathcal{U})$. Knowing this structure, one can use the cloud center $\mathcal{U}$ and the coset $\mathcal{X}_c = \mathcal{X} - \mathcal{U}$ for encoding $\mathcal{W}_2$ and $\mathcal{W}_1$, respectively.

In the case of binary symmetric channels, the optimal cloud center $\mathcal{U}$ can be obtained by sampling N bit unbiased vectors, for which the entropy per bit can be maximized to 1. On the other hand, one can produce the optimal coset $\mathcal{X}_c$ by independently and randomly generating each bit using a uniform bias $0 < \delta < 1$, which provides an entropy $H_2(\delta)$ per bit.

In an ideal situation, a noise vector $\boldsymbol{\xi}_1$ that is biased with a flip probability p_1 is added to the coset $\mathcal{X}_c$ in the first channel. This implies that the entropy of the received coset becomes $H_2(\delta * p_1)$ per bit, while the entropy of the noise vector is $H_2(p_1)$ per bit. Since one can use the difference between the entropies to convey the information of $\mathcal{W}_1$, the capacity of the first channel becomes $R_1 < H_2(\delta * p_1) - H_2(p_1)$, which is the second inequality of Eq. (2). On the other hand, for the second channel, characterized by a flip rate p_2, the coset $\mathcal{X}_c$ together with a channel noise $\boldsymbol{\xi}_2$ serves as a single noise vector, for which the entropy becomes $H_2(\delta * p_2)$ per bit. As the entropy of the received cloud center can be maximized to 1 per bit, this means that the capacity of the second channel is given by $R_2 < 1 - H_2(\delta * p_2)$, which is the first inequality of Eq. (2).

In linearly combined coding scheme $\binom{G_2^T}{G_3^T}\mathcal{W}_2 + \binom{G_1^T}{0}\mathcal{W}_1$, $\binom{G_3^T}{G_2^T}\mathcal{W}_2$ becomes almost random, which may serve as the optimal cloud center. However, the second part $\binom{G_1^T}{0}\mathcal{W}_1$, which corresponds to the coset, is somewhat structured, differing from the optimal choice of uniformly biased random vectors.

In order to compare the structured coset with the optimal one, let us fix the maximum entropy per bit of $\binom{G_1^T}{0}\mathcal{W}_1$, which equals $1-\alpha$, to that of the optimal coset $H_2(\delta)$. Then, one can show that the entropy of the corrupted coset with flip probability p per bit always increases from $H_2(\delta * p)$ to $(1-\alpha) + \alpha H_2(p) = H_2(\delta) + H_2(p) \geqslant H_2(\delta * p)$. This means that the critical rate of the first channel increases from $H_2(\delta * p_1) - H_2(p_1)$ to $(1-\alpha)[1 - H_2(p_1)]$, while that of the second channel reduces from $1 - H_2(p_2)$ to $\alpha[1 - H_2(p_2)]$. This trade-off between the capacities of the two channels limits the performance of linearly combined coding scheme to the time sharing limit that is always within Cover's capacity region.

In conclusion, while the suggested linearly combined LDPC coding scheme provides an improved performance over LDPC based time sharing codes for finite connectivity constructions, in both theoretical and practical limits, it cannot go beyond the theoretical time sharing limit; for that to happen, different coding schemes should be examined.

ACKNOWLEDGMENTS

Support by Grants-in-aid, MEXT (Grant Nos. 13680400 and 14084206) and JSPS (Y.K.), and The Royal Society, Grant Nos. EPSRC-GR/N63178 (D.S.) is acknowledged.

[1] T.M. Cover, IEEE Trans. Inf. Theory **18**, 2 (1972).
[2] P.P. Bergmans, IEEE Trans. Inf. Theory **19**, 197 (1973).
[3] T.M. Cover and J.A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
[4] T.M. Cover, IEEE Trans. Inf. Theory **44**, 2524 (1998).
[5] R.G. Gallager, IRE Trans. Inf. Theory **8**, 21 (1962).
[6] D.J.C. MacKay and R.M. Neal, *Lecture Notes in Computer Science* (Springer, Berlin, 1995), Vol. 1025, p. 100.
[7] M.C. Davey, *Record-Breaking Error Correction Using Low-Density Parity-Check Codes* (Gonville and Caius College, Cambridge, 1998).
[8] F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error-Correcting Codes* (North-Holland, Amsterdam, 1978).
[9] W. Van Gils, IEEE Trans. Inf. Theory **29**, 866 (1983); **30**, 544 (1984).
[10] Y. Kabashima, T. Murayama, and D. Saad, Phys. Rev. Lett. **84**, 1355 (2000).
[11] Y. Kabashima, T. Murayama, D. Saad, and R. Vicente, in *Advances in Neural Information Processing System*, edited by S. Solla, T. Leen, and K. Müler (MIT Press, Cambridge, MA, 2000), Vol. 12, p. 272.
[12] N. Sourlas, Nature (London) **339**, 693 (1989).
[13] T. Richardson and R. Urbanke, IEEE Trans. Inf. Theory **47**, 599 (2001).
[14] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference* (Morgan Kaufmann, San Francisco, CA, 1988).
[15] H.A. Bethe, Proc. R. Soc. London, Ser. A **151**, 552 (1935).
[16] J.S. Yedidia, W.T. Freeman, and Y. Weiss, in *Advances in Neural Information Processing Systems*, edited by T.K. Leen, T. Dietterich, and V. Tresp (MIT Press, Cambridge, MA, 2001), Vol. 13, p. 689.
[17] D.J. Thouless, P.W. Anderson, and R.G. Palmer, Philos. Mag. **35**, 593 (1977).
[18] Y. Kabashima and D. Saad, Europhys. Lett. **44**, 668 (1998).
[19] Y. Kabashima, e-print cond-mat/0211500.
[20] M. Mézard, G. Parisi, and M.A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
[21] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing* (Oxford University Press, Oxford, UK, 2001).
[22] H. Nishimori and D. Sherrington, in *Disorderd and Complex Systems*, edited by P. Sollich, A.C.C. Coolen, L.P. Hughston, and R.F. Streater (AIP, New York, 2001), p. 67.
[23] Y. Iba, J. Phys. A **32**, 3875 (1999).
[24] R. Vicente, D. Saad, and Y. Kabashima, in *Advances in Neural Information Processing Systems* (Ref. [16]), Vol. 13, p. 322.