Cavity approach to the Sourlas code system

Haiping Huang[1] and Haijun Zhou[1,2]
[1]*Key Laboratory of Frontiers in Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China*
[2]*Kavli Institute for Theoretical Physics China, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China*
(Received 18 May 2009; revised manuscript received 12 August 2009; published 23 November 2009)

The statistical physics properties of regular and irregular Sourlas codes are investigated in this paper by the cavity method. At finite temperatures, the free-energy density of these coding systems is derived and compared with the result obtained by the replica method. In the zero-temperature limit, the Shannon's bound is recovered in the case of infinite-body interactions while the code rate is still finite. However, the decoding performance as obtained by the replica theory has not considered the zero-temperature entropic effect. The cavity approach is able to consider the ground-state entropy. It leads to a set of evanescent cavity fields propagation equations which further improve the decoding performance as confirmed by our numerical simulations on single instances. For the irregular Sourlas code, we find that it takes the trade-off between good dynamical property and high performance of decoding. In agreement with the results found from the algorithmic point of view, the decoding exhibits a first-order phase transition as occurs in the regular code system with three-body interactions. The cavity approach for the Sourlas code system can be extended to consider first-step replica symmetry breaking.

DOI: 10.1103/PhysRevE.80.056113 PACS number(s): 89.90.+n, 02.70.−c, 89.70.−a, 05.50.+q

I. INTRODUCTION

Efficient and reliable transmission of information in noisy environment plays a central role in modern information society. Error-correcting codes, as efficient encoding/decoding mechanisms, find widespread applications ranging from the satellite communication to the storage of information on hard disks. In 1948, Shannon [1] proved that error-free transmission is possible as long as the code rate R (the ratio between the number of bits in the original message and the number of bits in the transmitted message) does not exceed the capacity of the channel (Shannon's bound). More explicitly, for the binary symmetric channel (BSC) where each transmitted bit is flipped independently with flip rate p, the Shannon bound is expressed as $R_c=1-H_2(p)$, where $H_2(p)=-p\log_2 p-(1-p)\log_2(1-p)$ is the binary entropy in the information theory literature [2]. This celebrated channel encoding theorem forms the core of information theory. However, it does not tell us how to construct an optimal code that saturates Shannon's bound. In information science many efforts have been devoted to construct (near) optimal codes [3].

Based on insights gained from the study of disordered systems [4] the Sourlas code was proposed 20 years ago, which relates error-correcting codes to spin-glass (SG) models [5]. In the past decade, the statistical mechanics analysis of Sourlas codes has been successfully generalized to other types of error-correcting codes including low-density parity-check (LDPC) codes, MacKay-Neal codes, Turbo codes, etc. Methods of statistical physics, complementary to those used in information theory, enable one to attain a more complete picture of decoding process by analyzing global properties of the corresponding free-energy landscape. They also allow one to optimize the performances of various codes by changing some construction parameters.

The procedure of constructing a Sourlas code is very simple. To infer which bit is flipped by noise at the receiving end of transmission, one has to introduce redundancy to the original message at the sending end. As for the Sourlas code, the redundancy is introduced by the Boolean sum of randomly selected message bits. Through the transformation $\xi_i=(-1)^{x_i}$, where x_i is the Boolean bit and ξ_i is the Ising spin, the original bit sequence $\{x_i\}$ can be regarded as an Ising spin configuration $\{\xi_i\}$. In this way, the modulo 2 addition is equivalent to spin multiplication; and then the Sourlas code can be mapped to a many-body spin-glass problem [6]. In a general scenario, the original message is an N-dimensional vector $\boldsymbol{\xi}\in\{\pm 1\}^N$, $M(>N)$ sets of interactions are constructed by taking the product of randomly sampled K bits from the sequence of the original message, i.e., $J_a^0=\xi_{a_1}\cdots\xi_{a_K}$ $(a=1,\ldots,M)$. Then they are fed into the noisy channel. At the destination, M corrupted interactions J_a, some of which being different from those at the sending end, are received. The arising problem is how to infer the original bits from the knowledge of channel outputs, statistical properties of the channel and of the source. In the presence of weak noise, searching for the ground state of the corresponding spin-glass model with given outputs $\{J_a\}$ will lead to successful decoding. This decoding scheme is nothing but maximum *a posterior* probability (MAP) decoding. When the noise becomes strong, the finite temperature decoding or marginal *posterior* maximizer (MPM) scheme should be adopted since the ground state would probably contain no information about the original message [6,7].

The fully connected Sourlas code has been studied in Ref. [5]. It was shown that the Shannon's bound is achieved in the limit $R\to 0$. Obviously, its practical potential is greatly limited. The finite rate Sourlas code of greater practical significance has been studied later on (see, e.g., Refs. [8–10]). It turns out that at finite coding rate R the Shannon's bound for the channel capacity can be attained at zero temperature at the limit of $K\to\infty$ [8,10]. However, the Shannon's bound could not be achieved for finite K despite its practical sig-

nificance. All the aforementioned investigations rely upon the replica method developed initially for solving the Sherrington-Kirkpatrick model of spin glass [4,11]. Moreover, they are restricted to the replica symmetry (RS) assumption due to the emerging more complicated saddle-point analysis of replica symmetry breaking. Nevertheless, recent developments in the study of LDPC codes [12] showed that the one-step replica symmetry breaking (1RSB) type algorithm is able to shift the dynamical phase transition [13] to a higher value as compared with RS-type algorithms. Similar results were obtained on the finite connectivity Sourlas code system from the dynamic point of view [14,15]. In this work we study the equilibrium properties of the finite connectivity Sourlas code system by using the cavity method of statistical physics [16–18].

The cavity method has its own advantages over the replica method. The latter is based on a saddle-point analysis of n-dimensional integral in the limit $n \to 0$. This analytic continuation in the number of replicas has not been confirmed to hold generally, neither has the validity of the exchange of the order of two limits ($N \to \infty$ and $n \to 0$). On the other hand, the cavity method adopts a direct probabilistic analysis, which makes it applicable to single problem instances. In this paper, it is expected that the cavity method reproduces results obtained by replica theory. Within the cavity framework, the entropic contribution in the zero-temperature limit can be taken into account by means of first-order corrections in temperature T, which has led to interesting insights on the ground state solution space properties of several disordered systems such as the random vertex cover problem and the random matching problem [19]. Following the same strategy, we derive the evanescent cavity field propagation (ECFP) equation for decoding Sourlas codes and find it outperforms the traditional case where only the hard field or energetic contribution is considered.

The rest of this paper is organized as follows. The model is introduced in Sec. II. In Sec. III, iterative equations for finite temperature decoding and zero-temperature decoding are rederived respectively using the cavity method. Taking into account the entropic contribution, we also propose the ECFP equation. In Sec. IV, regular (with a single K value) and irregular (with several values of K) Sourlas codes are discussed. In this section, it is also observed that the ECFP procedure is able to improve the decoding performance by a significant amount. We conclude this paper in Sec. V and make further discussions there.

II. MODEL

Hereafter, we adopt the Ising spin representation of the Boolean numbers. In the Sourlas code scenario, the original binary message $\boldsymbol{\xi} \in \{\pm 1\}^N$ of length N is encoded into a transmitted binary message $\boldsymbol{J}^0 = \{J_1^0, J_2^0, \ldots, J_M^0\}$ of length M, with the ath bit J_a^0 being the product of a subset ∂a of the original message bits, $J_a^0 = \Pi_{i \in \partial a} \xi_i$ [see Fig. 1(a) for a pictorial description, in which a parity check a is represented by a square and a message bit is represented by a circle]. If each parity check involves K bits and each bit is constrained by C parity checks, then the coding rate is $R \equiv \frac{N}{M} = \frac{K}{C}$. The Hamiltonian of the system reads as

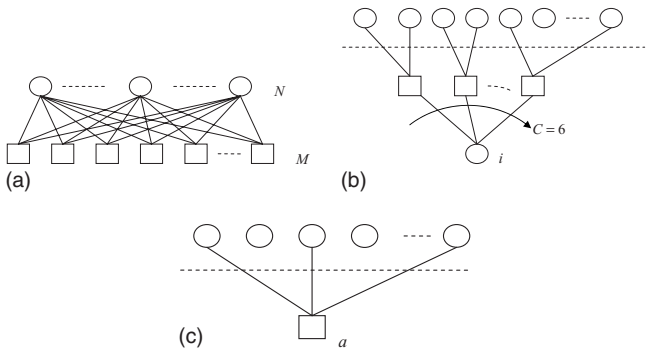

FIG. 1. Factor graph representation of a random construction of (a) Sourlas codes and [(b) and (c)] the cavity method. (a) There are totally N bits (circles) and M parity checks (squares) in the factor graph. Each bit (variable node) is connected to exactly six parity checks (function nodes), and each parity check involves three bits. (b) A single new bit i together with six parity checks is added to the original system denoted by the part above the dashed line. (c) A new function node a connected to three randomly selected bits is added to the original system.

$$\mathcal{H} = -\sum_{a=1}^{M} J_a \prod_{i \in \partial a} \sigma_i, \tag{1}$$

where $\{\sigma_i\}$ are referred to as dynamical spin variables for decoding and $\{J_a\}^M$ is the received message. Due to the noise in the transmission channel, the received message may not be identical to the transmitted one $\{J_a^0\}$. We assume memoryless binary symmetric channel, i.e.,

$$P(J_a|J_a^0) = p\,\delta(J_a + J_a^0) + (1-p)\,\delta(J_a - J_a^0), \tag{2}$$

where p is the flip rate.

Introducing an inverse temperature β as a control parameter, the spin configuration $\boldsymbol{\sigma}$ is sampled with probability

$$P(\boldsymbol{\sigma}|\boldsymbol{J}) = \frac{\exp[-\beta \mathcal{H}(\boldsymbol{\sigma})]}{Z}, \tag{3}$$

where Z is the partition function. On the other hand, Hamiltonian (1) is invariant under the gauge transformation $\sigma_i \to \sigma_i \xi_i$, $J_a \to J_a \Pi_{i \in \partial a} \xi_i$. Therefore, any general message can be mapped onto a ferromagnetic configuration $\{\xi_i = +1\}$. Under this transformation, Eq. (2) can be rewritten as

$$P(J_a) = p\,\delta(J_a + 1) + (1-p)\,\delta(J_a - 1). \tag{4}$$

In this sense, the Sourlas code is actually a multispin ferromagnetically biased $\pm J$ spin-glass model.

The aim of the statistical inference problem is to estimate the marginal posterior $P(\sigma_i|\boldsymbol{J})$. We adopt the MPM estimator $\hat{\xi}_i = \mathrm{sgn}[P(\sigma_i = 1|\boldsymbol{J}) - P(\sigma_i = -1|\boldsymbol{J})] = \mathrm{sgn}\langle \sigma_i \rangle_\beta$. To measure the performance of decoding, one usually defines the overlap between the estimated bits $\{\hat{\xi}_i\}$ and the original message $\{\xi_i\}$ as

$$m(\beta,p) = \frac{1}{N}\sum_{i=1}^{N} \xi_i\hat{\xi}_i = \frac{1}{N}\sum_{i=1}^{N} \xi_i \,\mathrm{sgn}\langle\sigma_i\rangle_\beta, \tag{5}$$

where $\mathrm{sgn}(x)=x/|x|$ for $x\neq 0$. Evaluating $\langle\sigma_i\rangle_\beta$ directly is computationally expensive, however, it can be well approximated using the cavity method presented in the next section. If we focus on typical value of the decoding overlap, Eq. (5) should be averaged over the quenched disorder, i.e.,

$$m(\beta,p) = \frac{1}{N}\left\langle \sum_{i=1}^{N} \xi_i \langle \mathrm{sgn}\langle\sigma_i\rangle_\beta\rangle_{\mathfrak{C},P(\boldsymbol{J}|\boldsymbol{\xi})} \right\rangle_{\boldsymbol{\xi}} \tag{6}$$

where $\mathfrak{C}$ represents the average over random constructions of codes with fixed bit's degree C. The other two types of quenched disorder come from the corruption process $(P(\boldsymbol{J}|\boldsymbol{\xi}))$ and the distribution of the original message bits $P(\boldsymbol{\xi})$. For simplicity, we concentrate on typical properties of the system with unbiased original message and memoryless binary symmetric channel. In the long message limit $(N\to\infty)$, it is believed that the macroscopic observables for a given instance are independent of the particular realization of the disorder [4,10].

III. CAVITY METHOD

Using the replica method, one is forced to work directly with the disorder average from the start, whereas the cavity method admits of taking the average over the quenched disorder after the computation. In this section, we derive the free energy at finite temperature as well as zero temperature for the finite connectivity Sourlas code system using the cavity method and then extend the result to the irregular Sourlas code case. Within the cavity framework, the entropic contribution is considered in the zero-temperature limit and the ECFP equation is proposed as well.

A. Finite temperature decoding

Because of the random construction of Sourlas codes, it is reasonable to assume that the correlation between randomly sampled bits vanishes in the long message limit. We assume all the calculations below are within the RS ansatz (single-state cavity method). The results are straightforward to be generalized to 1RSB case.

As shown in Fig. 1(b), if we add one variable node to the original system, C function nodes should be added simultaneously. Then the partition function for the enlarged system is

$$\begin{aligned} Z^{new} &= \sum_{\sigma_i}\sum_{\vec{\sigma}} \exp\left(\sum_{a=1}^{M} \beta J_a \prod_{k\in\partial a}\sigma_k + \beta\sum_{b=1}^{C} J_b\sigma_i \prod_{j\in\partial b\backslash i}\sigma_j\right) \\ &= Z^{old}\sum_{\sigma_i}\prod_{b}\sum_{\{\sigma_j\}:j\in\partial b\backslash i}\prod_{j\in\partial b\backslash i} \\ &\quad\times\left[\frac{e^{\beta h_{j\to b}\sigma_j}}{2\cosh\beta h_{j\to b}}\right]\cdot e^{\beta J_b\sigma_i\Pi_{j\in\partial b/i}\sigma_j} \\ &= Z^{old}\left\{\prod_b\left[\cosh\beta J_b\left(1+\tanh\beta J_b\prod_{j\in\partial b\backslash i}\tanh\beta h_{j\to b}\right)\right]\right. \\ &\quad\left.+\prod_b\left[\cosh\beta J_b\left(1-\tanh\beta J_b\prod_{j\in\partial b\backslash i}\tanh\beta h_{j\to b}\right)\right]\right\}, \end{aligned} \tag{7}$$

where σ_i is the newly added spin, $Z^{old}=\Sigma_{\vec{\sigma}}\exp(\Sigma_{a=1}^{M}\beta J_a\Pi_{i\in\partial a}\sigma_i)$ is the partition function of the old system, $h_{j\to b}$ is the cavity field of variable node j when function node b is removed from the graph, $j\in\partial b\backslash i$ denotes the set of bits involved in function node b but i is excluded from this set. To derive the second equality in Eq. (7), we have made use of the absence of strong correlation between randomly chosen spins, since for one random construction of Sourlas codes depicted in Fig. 1(a), the typical loop size in the corresponding factor graph is of order $\log N$ which diverges in $N\to\infty$. In this sense, the joint probability of a few randomly selected spins $P(\boldsymbol{\sigma}_{\partial a})$ is factorized as $P(\boldsymbol{\sigma}_{\partial a})\approx\Pi_{i\in\partial a}P(\sigma_i)$, where we write single node belief $P(\sigma_i)$ as $P(\sigma_i)=e^{\beta h_i\sigma_i}/2\cosh\beta h_i$ in terms of the local field h_i acting on the spin σ_i.

Upon defining the magnetization $m_{i\to b}\equiv\tanh\beta h_{i\to b}$ and the conjugate magnetization $\hat{m}_{b\to i}\equiv\tanh\beta u_{b\to i}\equiv\tanh\beta J_b\Pi_{j\in\partial b\backslash i}\tanh\beta h_{j\to b}$, where $u_{b\to i}$ is termed the cavity bias, one gets the free-energy shift due to one variable node addition:

$$\begin{aligned} -\beta\Delta F_i &= \log\frac{Z^{new}}{Z^{old}} \\ &= \log\left\{\prod_{b\in\partial i}\left[\cosh\beta J_b(1+\hat{m}_{b\to i})\right]\right. \\ &\quad\left.+\prod_{b\in\partial i}\left[\cosh\beta J_b(1-\hat{m}_{b\to i})\right]\right\}. \end{aligned} \tag{8}$$

As the second step, one function node addition is performed [cf. Fig. 1(c)]. Likewise, the new partition function reads

$$\begin{aligned} Z^{new} &= \sum_{\vec{\sigma}}\exp\left(\beta\sum_{a=1}^{M}J_a\prod_{k\in\partial a}\sigma_k+\beta J_a\prod_{i\in\partial a}\sigma_i\right) \\ &= \sum_{\vec{\sigma}} e^{\beta\Sigma_{a=1}^{M}J_a\Pi_{k\in\partial a}\sigma_k}\sum_{\vec{\sigma}}\frac{e^{\beta\Sigma_{a=1}^{M}J_a\Pi_{k\in\partial a}\sigma_k}}{\Sigma_{\vec{\sigma}}e^{\beta\Sigma_{a=1}^{M}J_a\Pi_{k\in\partial a}\sigma_k}}e^{\beta J_a\Pi_{i\in\partial a}\sigma_i} \\ &= Z^{old}\sum_{\vec{\sigma}}P(\vec{\sigma})e^{\beta J_a\Pi_{i\in\partial a}\sigma_i} \\ &= Z^{old}\sum_{\{\sigma_i\}:i\in\partial a}\prod_{i\in\partial a}\left[\frac{e^{\beta h_{i\to a}\sigma_i}}{2\cosh\beta h_{i\to a}}\right]e^{\beta J_a\Pi_{i\in\partial a}\sigma_i} \\ &= Z^{old}\cdot\cosh\beta J_a\left(1+\tanh\beta J_a\prod_{i\in\partial a}m_{i\to a}\right). \end{aligned} \tag{9}$$

The corresponding free-energy shift is $-\beta\Delta F_a=\log[\cosh\beta J_a(1+\tanh\beta J_a\Pi_{i\in\partial a}m_{i\to a})]$. Finally the total free-energy density is given by [17]

$$f = \frac{1}{N}\sum_i \Delta F_i - \frac{1}{N}\sum_a (|\partial a| - 1)\Delta F_a$$
$$= \langle \Delta F_i \rangle_{pop} - \frac{K-1}{K} C \langle \Delta F_a \rangle_{pop}, \quad (10)$$

where $\langle \cdots \rangle_{pop}$ means the average over populations of $\{m_{i\to a}, \hat{m}_{a\to i}\}$ when the population dynamics recipe [17] is adopted. The second term in the final expression of Eq. (10) can be understood as follows: when one variable node is added, the number of overgenerated function nodes is $\frac{K-1}{K}C$ on average; the contribution of these nodes should be eliminated from the total free energy. Following the same line mentioned above, one can write $m_{i\to a}$ as a function of $\{\hat{m}_{b\to i}\}_{b\in\partial i\backslash a}$, then obtain a closed set of equations in the form of distribution:

$$P(m_{i\to a}) = \int \left[\prod_{b\in\partial i\backslash a} Q(\hat{m}_{b\to i}) d\hat{m}_{b\to i}\right] \times \delta\left(m_{i\to a} - \frac{\prod_{b\in\partial i\backslash a}(1+\hat{m}_{b\to i}) - \prod_{b\in\partial i\backslash a}(1-\hat{m}_{b\to i})}{\prod_{b\in\partial i\backslash a}(1+\hat{m}_{b\to i}) + \prod_{b\in\partial i\backslash a}(1-\hat{m}_{b\to i})}\right), \quad (11a)$$

$$Q(\hat{m}_{b\to i}) = \int \left[\prod_{j\in\partial b\backslash i} P(m_{j\to b}) dm_{j\to b}\right] \times \delta\left(\hat{m}_{b\to i} - \tanh \beta J_b \prod_{j\in\partial b\backslash i} m_{j\to b}\right). \quad (11b)$$

Equation (11) is nothing but the belief propagation equation when applied to a single instance (one particular realization of Sourlas codes) [9]. Population dynamics recipe is applied to solve the recursive equations above. When the iteration reaches a steady state, the free energy can be computed and the marginal posterior can be well approximated by $P(\sigma_i) = \frac{(1+m_i\sigma_i)}{2}$ for the sparse random graph. According to Eq. (5), the performance of decoding is evaluated via $m = \frac{1}{N}\Sigma_i \xi_i \hat{\xi}_i = \int dm_i P(m_i)\mathrm{sgn}(m_i)$, where the gauge transformation has been performed and the magnetization m_i obeys the distribution

$$P(m_i) = \int \left[\prod_{b\in\partial i} Q(\hat{m}_{b\to i}) d\hat{m}_{b\to i}\right] \times \delta\left(m_i - \frac{\prod_{b\in\partial i}(1+\hat{m}_{b\to i}) - \prod_{b\in\partial i}(1-\hat{m}_{b\to i})}{\prod_{b\in\partial i}(1+\hat{m}_{b\to i}) + \prod_{b\in\partial i}(1-\hat{m}_{b\to i})}\right). \quad (12)$$

B. Zero-temperature decoding

The finite temperature decoding is facilitated through Eq. (11). However, searching for the ground state of the system requires performing zero-temperature decoding, and the equations derived above can be further simplified. Taking the limit $\beta\to\infty$, one obtains the recursive equations for cavity fields and biases:

$$P(h_{i\to a}) = \int \left[\prod_{b\in\partial i\backslash a} du_{b\to i} Q(u_{b\to i})\right] \delta\left(h_{i\to a} - \sum_{b\in\partial i\backslash a} u_{b\to i}\right), \quad (13a)$$

$$Q(u_{b\to i}) = \int \left[\prod_{j\in\partial b\backslash i} dh_{j\to b} P(h_{j\to b})\right] \times \delta\left[u_{b\to i} - \mathrm{sgn}\left(J_b \prod_{j\in\partial b\backslash i} h_{j\to b}\right)\right], \quad (13b)$$

and the free-energy shifts

$$-\Delta F_i = C - \sum_{b\in\partial i} |u_{b\to i}| + \left|\sum_{b\in\partial i} u_{b\to i}\right|, \quad (14a)$$

$$-\Delta F_a = 1 - 2\Theta\left(-J_a \prod_{i\in\partial a} h_{i\to a}\right), \quad (14b)$$

where $\Theta(x)$ is a step function taking values $\Theta(x)=0$ for $x\le 0$ and $\Theta(x)=1$ for $x>0$. In Eq. (13b), we take the convention $\mathrm{sgn}(0)=0$. Similarly, the overlap in the zero-temperature limit reads $m=\int dh P(h)\mathrm{sgn}(h)$ where the field is subject to the distribution

$$P(h) = \int [\Pi_{b\in\partial i} Q(u_{b\to i}) du_{b\to i}] \delta(h - \Sigma_{b\in\partial i} u_{b\to i}),$$

where $Q(u_{b\to i})$ is the distribution of cavity biases according to Eq. (13b).

C. Evanescent cavity fields propagation

In Sec. III B, only the hard field or energetic contribution is considered. We expect that the neglected entropic contribution will provide useful information for improving the decoding performance. To derive the ECFP equation, we rewrite Eq. (11) in terms of cavity fields:

$$\eta_{i\to a} \equiv 2h_{i\to a} = \sum_{b\in\partial i\backslash a} \frac{1}{\beta} \log\left[\frac{1+\hat{m}_{b\to i}}{1-\hat{m}_{b\to i}}\right]. \quad (15)$$

When we consider only the energetic contribution in the zero-temperature limit, the resulting closed set of equations Eq. (13) are called warning propagation (WP) [19]. The limit $\beta\to\infty$ selects the ground state of the system under consideration, therefore WP also corresponds to the MAP estimator. However, as T goes to zero, the local field h_i vanishes linearly in T, consequently contributes to the corresponding local magnetization [18]. That is to say, even if the local field takes value of zero, the nonvanishing evanescent part, defined as the coefficient of first-order correction of cavity field with respect to T, still results in a finite magnetization. Therefore, these evanescent fields are expected to provide useful information for improving the decoding performance. Expanding the cavity field $h_{i\to a}$ up to the first order in T, i.e.,

$$\eta_{i\to a} = 2I_{i\to a} + \frac{r_{i\to a}}{\beta}, \quad (16)$$

where $I_{i\to a}$ is an integer corresponding to the energetic contribution and $r_{i\to a}$ a finite real value corresponding to the entropic contribution, then substituting Eq. (16) into Eq. (15), one readily gets ECFP equations:

$$I_{i\to a} = \sum_{b\in\partial i\backslash a} \mathrm{sgn}\left(J_b \prod_{j\in\partial b\backslash i} I_{j\to b}\right), \tag{17a}$$

$$r_{i\to a} = \sum_{b\in\partial i\backslash a} \mathbb{I}(I_{j\to b}=0\ \forall\ j\in\partial b\backslash i)\log\left[\frac{\Pi_{j\in\partial b\backslash i}(e^{r_{j\to b}}+1)+J_b\Pi_{j\in\partial b\backslash i}(e^{r_{j\to b}}-1)}{\Pi_{j\in\partial b\backslash i}(e^{r_{j\to b}}+1)-J_b\Pi_{j\in\partial b\backslash i}(e^{r_{j\to b}}-1)}\right]$$

$$+\mathbb{I}(\text{at least one } I_{j\to b}=0,\ \text{at most } (K-2)\ I_{j\to b}=0\ \forall\ j\in\partial b\backslash i)\cdot\log\left[\frac{\Pi'_{j\in\partial b\backslash i}(1+e^{-|r_{j\to b}|})+\tilde{J}_b\Pi'_{j\in\partial b\backslash i}(1-e^{-|r_{j\to b}|})}{\Pi'_{j\in\partial b\backslash i}(1+e^{-|r_{j\to b}|})-\tilde{J}_b\Pi'_{j\in\partial b\backslash i}(1-e^{-|r_{j\to b}|})}\right]$$

$$-\mathbb{I}(I_{j\to b}\neq 0\ \forall\ j\in\partial b\backslash i)\mathrm{sgn}\left(J_b\prod_{j\in\partial b\backslash i} I_{j\to b}\right)\log(1+R_{b\to i}), \tag{17b}$$

where $\Pi'_{j\in\partial b\backslash i}\equiv\Pi_{\substack{j\in\partial b\backslash i\\ \{I_{j\to b}=0\}}}$, $\tilde{J}_b=J_b\ \mathrm{sgn}(\Pi_{\substack{k\in\partial b\backslash i\\ \{I_{k\to b}\neq 0\}}} I_{k\to b})\times\mathrm{sgn}(\Pi'_{j\in\partial b\backslash i} r_{j\to b})$, $\mathbb{I}(\cdot)$ is the indicator function of an event, and $R_{b\to i}=\Sigma_{j\in\partial b\backslash i}\tilde{R}_{j\to b}$, where $\tilde{R}_{j\to b}=\exp[-\mathrm{sgn}(I_{j\to b})r_{j\to b}]$ if $|I_{j\to b}|=1$ and 0 otherwise. In the summation of Eq. (17b), the first term corresponds to the case where $I_{j\to b}=0$ for all $j\in\partial b\backslash i$, the second term the case where at least one $I_{j\to b}=0$, at most $(K-2)I_{j\to b}=0$ and the last term the case where $I_{j\to b}\neq 0$ for all $j\in\partial b\backslash i$. Then the decoding can be easily performed via $m=\int dI_i dr_i P(I_i)Q(r_i)[\mathrm{sgn}(I_i)+\mathbb{I}(I_i=0)\mathrm{sgn}(r_i)]$, where P, Q represent the distributions for the hard fields $\{I_i\}$ and evanescent fields $\{r_i\}$, respectively, when the population dynamics technique is used to solve the ECFP equations. Actually, in the zero-temperature limit, the estimated message bit $\xi_i=\mathrm{sgn}(m_i)=\mathrm{sgn}(\tanh\beta h_i)=\mathrm{sgn}(I_i)$ if $I_i\neq 0$ and $\mathrm{sgn}(r_i)$ otherwise.

D. Decoding irregular Sourlas codes

All the aforementioned computations are limited to the regular case where the check's degree K takes a single value. It is worthwhile to study the irregular case. The irregular Sourlas code is defined as the code with various values of K. We assume the check's degree K follows a distribution with two delta peaks

$$P(K)=\gamma\delta(K-2)+(1-\gamma)\delta(K-3). \tag{18}$$

We adopt this form of distribution for two reasons. One is the Sourlas code has perfect dynamical properties for $K=2$ and high decoding performance for $K=3$. The other is the result can be compared with that obtained for the cascading Sourlas code [15,20]. The formula for the total free-energy density of the combined system is given by

$$f=\frac{1}{N}\sum_i \Delta F_i-\frac{1}{N}\sum_a(|\partial a|-1)\Delta F_a$$

$$=\langle\Delta F_i\rangle_{pop}-\frac{M}{N}\sum_K P(K)(K-1)\langle\Delta F_a\rangle_{pop}$$

$$=\langle\Delta F_i\rangle_{pop}-\frac{C}{\bar{K}}[\gamma\langle\Delta F_{K=2}\rangle_{pop}+2(1-\gamma)\langle\Delta F_{K=3}\rangle_{pop}], \tag{19}$$

where $\bar{K}=3-\gamma$ and the code rate $R=\frac{3-\gamma}{C}$. The recursive equations are of the form

$$P(m_j)=\int\left[\prod_{b=1}^{C-1} d\hat{m}_b Q(\hat{m}_b)\right]\times\delta\left(m_j-\frac{\Pi_{b=1}^{C-1}(1+\hat{m}_b)-\Pi_{b=1}^{C-1}(1-\hat{m}_b)}{\Pi_{b=1}^{C-1}(1+\hat{m}_b)+\Pi_{b=1}^{C-1}(1-\hat{m}_b)}\right), \tag{20a}$$

$$Q(\hat{m}_b)=\sum_K\frac{P(K)K}{\bar{K}}\int\left[\prod_{j=1}^{K-1}P(m_j)dm_j\right]\times\delta\left(\hat{m}_b-\tanh\beta J_b\prod_{j=1}^{K-1}m_j\right). \tag{20b}$$

Equation (20b) can be understood as follows: since γ represents the fraction of function nodes with two-spin interaction, for one randomly chosen bit, it is connected to a parity check involving two bits with probability $P_2=\frac{2\gamma}{3-\gamma}$ and to that involving three bits with probability $P_3=\frac{3(1-\gamma)}{3-\gamma}$. Obviously, $P_2+P_3=1$. The formula for zero-temperature decoding of irregular codes can be derived similarly. In the next section, we will discuss the performance of decoding for regular and irregular codes, respectively.

IV. RESULTS AND DISCUSSIONS

A. Regular Sourlas codes

Properties of regular Sourlas codes have been studied using replica theory [8,10]. In this section, we reproduce

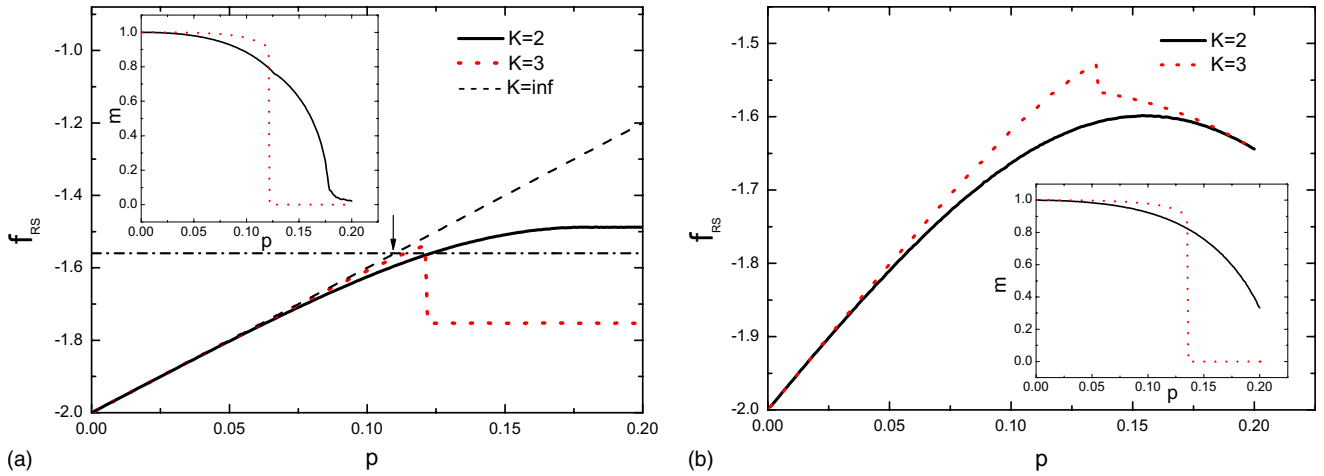

FIG. 2. (Color online) The decoding performance for regular Sourlas codes with $R=0.5$. The calculated mean values are shown and the corresponding variances are smaller than the symbol size. (a) The replica symmetry free-energy density versus flip rate when zero-temperature decoding is performed. The solid line corresponds to $K=2$ case while the dotted line $K=3$ case. The dashed line represents the case of $K\to\infty$, and the dashed-dotted line corresponds to the frozen spin solution. The arrow indicates the critical noise level where the Shannon's bound is achieved. (b) The replica symmetry free-energy density versus flip rate when finite temperature decoding is performed. The decoding temperature is chosen to be Nishimori temperature $\beta_p=\frac{1}{2}\log\frac{1-p}{p}$. Insets: the overlap versus flip rate for zero-temperature and finite temperature decoding, respectively.

results obtained on regular Sourlas codes on the basis of the cavity method.

For regular Sourlas codes, we consider the case of $K=2$ and $K=3$ with the same code rate $R=0.5$. In particular, the other cases ($K>3$) show the same qualitative behavior as the $K=3$ case. It is worthwhile to mention that Eq. (11) yields a paramagnetic solution, i.e., $P(m_{i\to a})=\delta(m_{i\to a}), Q(\hat{m}_{b\to i})=\delta(\hat{m}_{b\to i})$. Following Eq. (10), one readily acquires the paramagnetic free energy $f_{para}=-\frac{1}{\beta}(\log 2+\frac{1}{R}\log\cosh\beta)$ and the entropy $s_{para}=\frac{1}{R}(\log\cosh\beta-\beta\tanh\beta)+\log 2$. In zero-temperature limit, $s_{para}=(1-\frac{1}{R})\log 2$. Since the entropy becomes negative when $R<1$, the paramagnetic solution is irrelevant for the error-correcting purpose. On the other hand the RS spin-glass solution reduces to the paramagnetic one, therefore, the replica symmetry should be broken for low enough temperature, and a simple assumption (frozen spins assumption) is adopted to avoid the negative entropy [8,10], i.e., for low enough temperature, the system settles in a completely frozen glassy phase. On the transition boundary, both the frozen glassy phase and paramagnetic phase share the identical free energy, and the transition temperature is determined by $s_{para}(\beta_g)=0$. When $T<T_g$, the spin-glass phase takes over, and the corresponding free-energy density can be written as $f_{sg}=f_{para}(\beta_g)$, independent of the temperature. Besides the paramagnetic and spin-glass solution, there exists a ferromagnetic solution ($m=1$). This solution is possible only in the case of $K\to\infty$ (note that R is kept finite). The related ferromagnetic free energy with vanishing entropy could be derived according to Eq. (10), i.e., $f_{ferro}=-\frac{1}{R}(1-2p)$, independent of the temperature as well. By identifying f_{ferro} with f_{sg}, one can recover the Shannon's bound as predicted by Shannon's channel encoding theorem, implying $p_c\simeq 0.110\,028$ when $R=0.5$ [cf. Fig. 2(a), the arrow indicates this critical noise level]. We report the phase diagram for the $K\to\infty$ code in Fig. 3, note that the code rate is still kept to be finite. It is important to remark that when the finite connectivity is considered, modest loss in the final decoding quality should be paid, i.e., the decoding overlap will be smaller than unity. To illustrate the phase transition in the finite connectivity case, we refer to the phase with finite decoding overlap as the ferromagnetic phase until the glassy phase dominates. The ferromagnetic-spin-glass transition is determined by identifying the ferromagnetic free energy with frozen glassy free energy, then the glassy phase ($m=0$) sets in to replace the ferromagnetic phase (finite m). The correspond-

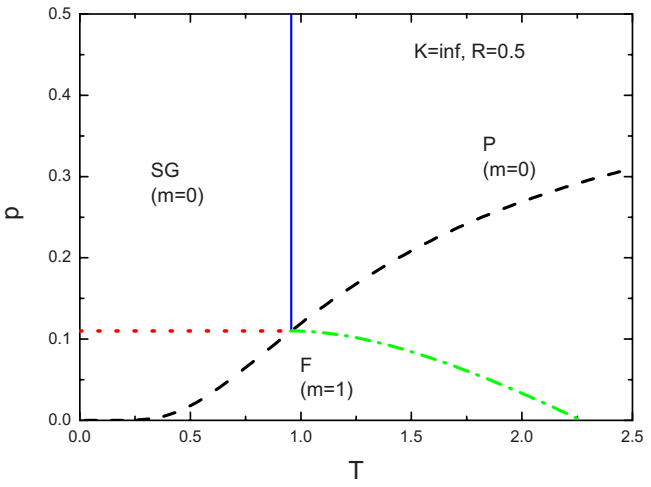

FIG. 3. (Color online) The phase diagram for regular Sourlas codes with $K\to\infty$ keeping $R=0.5$. The dashed line indicates the Nishimori line, the dotted line the boundary between spin-glass (SG) phase and ferromagnetic (F) phase, the dashed-dotted line the boundary between F and paramagnetic (P) phase and the solid line the boundary between SG and P.

ing critical noise level is obviously smaller than the point where the magnetization (more precisely the decoding overlap) drops to zero.

To solve Eq. (11), population dynamics technique introduced in Ref. [17] is applied. The size of population is taken to be of order 10^4. Results are reported in Fig. 2. For $K=2$, no prior knowledge of the original message is required for decoding, and the phase transition is of second order. As shown in Fig. 2(a), the critical noise level is determined by the point where the frozen spin free energy coincides with the RS free energy. After the transition, the spin-glass phase dominates and the corresponding free energy is fixed to be f_{sg}. Conversely, the phase transition is of first order for $K=3$, and there is a remarkable drop in the free-energy profile. However, the computed free energy, which seems to be lower than the frozen spin one, is unphysical after the phase transition because of its corresponding negative entropy. Therefore the RS assumption is incorrect and many states assumption should be adopted. The performance of finite temperature decoding is also shown in Fig. 2(b). The decoding temperature is chosen to be the optimal one, $\beta_p = \frac{1}{2}\log\frac{1-p}{p}$ named Nishimori temperature [6,7]. In this case, the thermal temperature is identical to the noise temperature, and it is observed that the performance is better than that of zero-temperature decoding. Actually, the average spin alignment m of decoding at Nishimori temperature sets an upper bound for all achievable alignments [7]. As our numerical simulation has shown, only the Nishimori temperature survives to get high overlap when the critical noise level is approached. In contrast to the $K=2$ case, the case of $K=3$ improves the decoding performance significantly. However, the basin of attraction (BOA) shrinks dramatically. We have to assume initial bias $m_I=0.8$ for finite temperature decoding and $m_I=0.75$ for zero-temperature decoding. The compromise between good dynamical properties on one side ($K=2$) and good performance on the other side ($K=3$) motivates the following investigation of the typical properties of the combined system with various K.

To further improve the decoding performance in the limit when the temperature goes to zero, we have proposed the ECFP equation in Sec. III C. The decoding overlap is plotted against the flip rate in Fig. 4. Results obtained by WP are also shown for comparison. Apparently, the decoding performance is improved within an intermediate range of flip rate. In the presence of weak noise, most of the propagating cavity fields take values larger than 2 and the energetic contribution plays a dominant role. Thus both methods lead to identical performance. Once the noise becomes no more small, the decoding performance achieved by ECFP starts to deteriorate due to the divergence of some of the evanescent fields. If we set a cutoff (e.g., 4.0), to our surprise, the problem mentioned above can be successfully circumvented. As shown in Fig. 4, the result indeed outperforms that obtained by WP which neglects the entropic effects. This can be understood as follows, as flip rate becomes high enough, the relevant cavity fields with $|I_{i\to a}|=1$ or $I_{i\to a}=0$, emerge and contribute to the entropic effects [19]. These information, omitted by WP, is correctly extracted by ECFP procedure, and the decoding performance is finally boosted. According to our numerical simulations, the value (e.g., 3.0, 4.0, 5.0) we choose for the

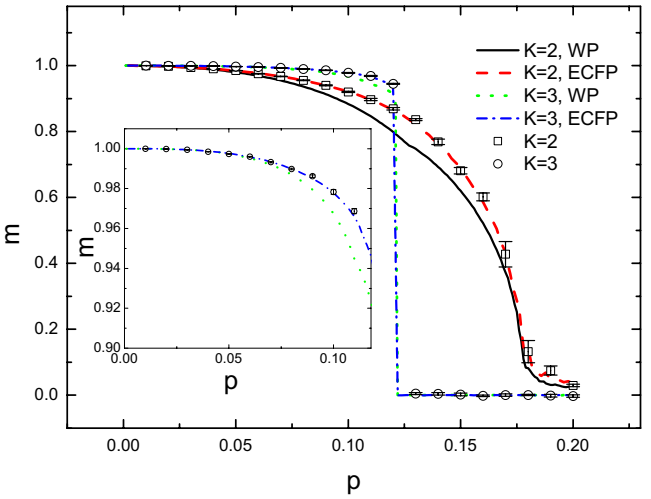

FIG. 4. (Color online) The decoding overlap m as a function of flip rate p for regular Sourlas codes with $R=0.5$. The solid or dashed line corresponds to zero-temperature decoding for the $K=2$ case while the dotted or dashed dotted line the $K=3$ case. The solid or dotted one represents results obtained by the conventional warning propagation (WP) while the dashed or dashed dotted one evanescent cavity fields propagation (ECFP). The cutoff takes the value 4.0. Numerical simulations on a single graph by ECFP are consistent with the mean-field results. The size of the graph is $N=10\,000$. The decoding result on single graph is averaged over ten individual simulations for each flip rate p. The error bars indicate the standard deviations. Inset: a detailed view of the significant improvement using ECFP for the $K=3$ case.

cutoff does not affect the decoding results. When zero-temperature decoding is concerned, we have observed that ECFP is able to do a better job than WP since the entropic effects have been incorporated. However, its decoding performance still lies beneath that achieved by the optimal decoding (MPM) where the decoding temperature is chosen to be the Nishimori type. However, for MPM, one has to have a prior knowledge of the channel noise, i.e., the flip rate of the noise. We present the comparison between these two different kinds of decoding in Fig. 5. In order to validate the mean-field results, we run the ECFP decoding algorithm on a single instance. The comparison is shown in Fig. 4. The size of the graph is set to be $N=10\,000$ and the code rate $R=0.5$. For one iteration step, messages sent from each bit on the graph are updated one time on average. We also set the maximal number of iteration steps $\mathcal{T}$ to be 500. The decoding result on single graph is averaged over ten individual simulations for each flip rate p. As observed in our simulations, the number of iteration steps, around $p=0.12$, exceeds the preset value on most of the presented instances, which manifests the ECFP starts to lose the convergence on a single graph. However, the agreement with the mean-field result is excellent.

B. Irregular Sourlas codes

As defined above, the irregular Sourlas code is a combined system with various values of K. It takes well the trade-off between excellent convergence property of low-K

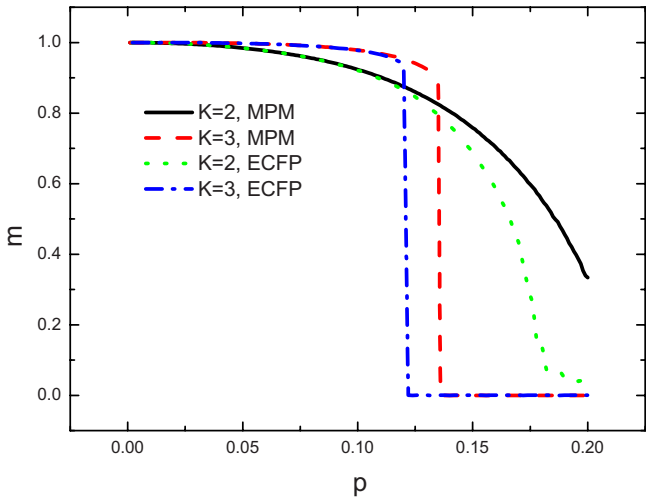

FIG. 5. (Color online) The decoding overlap m as a function of flip rate p for regular Sourlas codes with $R=0.5$. The calculated mean values are shown and the corresponding variances are smaller than the symbol size. The solid and dashed lines correspond to finite temperature decoding (MPM) for the $K=2$ case and $K=3$ case, respectively, while the dotted and dashed dotted lines the ECFP decoding for which the cutoff takes the value 4.0.

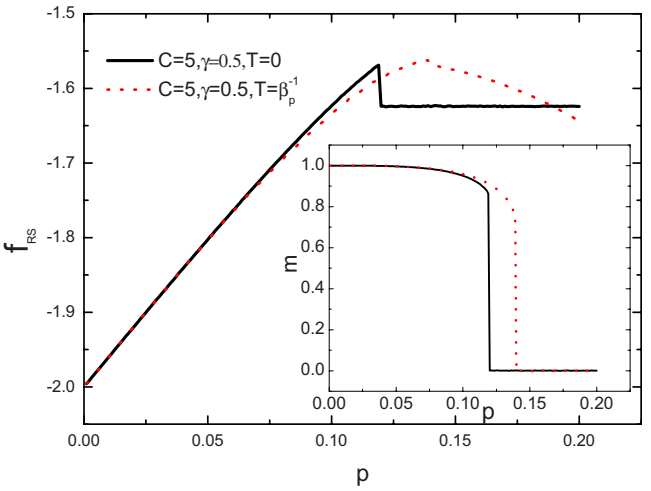

FIG. 6. (Color online) The decoding performance for irregular Sourlas codes with $R=0.5$. The calculated mean values are shown and the corresponding variances are smaller than the symbol size. The solid line corresponds to zero-temperature decoding while the dotted line finite temperature decoding. Inset: the overlap versus flip rate.

codes and high decoding performance of high-K codes. From the algorithmic point of view, the irregular code is also termed cascading code put forward in Ref. [20] and further studied in Ref. [15]. In this section, we report results on typical properties of the combined system based on the cavity analysis presented in Sec. III D.

To retain the same code rate $R=0.5$, we choose $C=5$, $\gamma=0.5$ as code construction parameters. Population dynamics recipe is used to solve Eq. (20), and the size of population is assumed to be of order 10^4. As shown in Fig. 6, the combined system exhibits a first-order phase transition as the $K=3$ case, which was also observed in Ref. [15] where the cascaded encoding/decoding scheme was employed. After this transition, the free energy crosses over to a lower value. However, as the flip rate p increases to a high enough value, the RS entropy will be negative and the RS assumption is then incorrect, indicating replica symmetry should be broken. As observed in Fig. 6, the finite temperature (Nishimori's temperature) decoding is superior to the zero temperature one when the noise level becomes no longer low. Compared with the regular code of $K=3$, the BOA for the combined system becomes larger thus we only need to take the initial bias $m_I=0.6$. Additionally, the overlap of decoding for the combined system is higher than that of $K=2$. Therefore, results demonstrated in Fig. 6 provide us an opportunity to construct an optimal code. As has been stated in Refs. [15,20], one can use multiple values of K in the interactions. As a first step, belief propagation or 1RSB algorithm is run on a partial system with only low $K(K=2)$ interactions since the low-K code has perfect convergence properties. The end overlap at the first stage is expected to be well within the BOA of the combined system. Once higher body (e.g., $K=3$) interactions are invoked, an end overlap higher than the one obtained by the initial step will be resulted in.

V. CONCLUSIONS AND FUTURE PERSPECTIVES

In this work, we have studied the finite connectivity Sourlas code based on the cavity method. Conventional replica results on the regular code are cross-checked. Moreover, this cavity analysis is extended to the irregular case. Typical properties of the combined system are investigated. It is shown that the decoding for the combined system exhibits a first-order phase transition as occurs in the regular case ($K=3$). The combined system is of two striking features, one is the initial bias required for convergence is degraded, the other is the final performance is enhanced. Actually, this does mean that the good dynamical properties (large BOA) and high decoding performance should be compromised in the algorithmic implementation. Thus introducing gradually higher K interactions seems to be an effective way to take advantage of this trade-off.

As for the regular codes system, the evanescent cavity fields propagation equation is proposed. And it is capable of extracting the entropic information in the zero-temperature limit, thus the decoding performance is considerably enhanced compared with the traditional case where only the hard field is taken into account. Numerical simulations on single instances are compatible with the mean-field results.

The cavity methodology, applied in our work, is very promising. Unlike replica trick, it formulates assumptions in a more explicit manner, even opens the way to algorithmic implementations on one single instance. In this work, we also discovered that the system shows negative entropy in the presence of low enough decoding temperature and high enough flip rate, therefore 1RSB is needed for further investigation on the finite connectivity Sourlas code. Fortunately, the cavity method can be easily generalized to 1RSB case. Meanwhile, the frozen spin-glass scheme we have adopted in Sec. IV A could be also cross-checked. On the other hand,

further study is required for the combined system to elucidate under what conditions the channel capacity is achieved [21]. Finally, the methodology is expected to be applied to more practical codes like LDPC codes. These lines of research are currently under way and these further investigations are anticipated to provide deeper insights into a variety of codes with low-density nature of constructions.

ACKNOWLEDGMENTS

We thank Pan Zhang and Jie Zhou for stimulating discussions, and we are grateful to anonymous referees for many helpful comments. The present work was in part supported by the National Science Foundation of China (Grant No. 10774150) and by the National Basic Research Program (973-Program) of China (Grant No. 2007CB935903).

[1] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948).
[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
[3] R. G. Gallager, *Low-Density Parity-Check Codes* (MIT Press, Cambridge, 1963); D. J. C. MacKay and R. M. Neal, Electron. Lett. **32**, 1645 (1996); D. J. C. MacKay, IEEE Trans. Inf. Theory **45**, 399 (1999); A. Montanari and N. Sourlas, Eur. Phys. J. B **18**, 107 (2000).
[4] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
[5] N. Sourlas, Nature (London) **339**, 693 (1989); Europhys. Lett. **25**, 159 (1994).
[6] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford University Press, Oxford, 2001).
[7] P. Ruján, Phys. Rev. Lett. **70**, 2968 (1993); H. Nishimori, J. Phys. Soc. Jpn. **62**, 2973 (1993).
[8] Y. Kabashima and D. Saad, Europhys. Lett. **45**, 97 (1999).
[9] Y. Kabashima and D. Saad, Europhys. Lett. **44**, 668 (1998).
[10] R. Vicente, D. Saad, and Y. Kabashima, Phys. Rev. E **60**, 5352 (1999).
[11] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **35**, 1792 (1975).
[12] G. Migliorini and D. Saad, Phys. Rev. E **73**, 026122 (2006).
[13] S. Franz, M. Leone, A. Montanari, and F. Ricci-Tersenghi, Phys. Rev. E **66**, 046120 (2002).
[14] B. Wemmenhove and H. J. Kappen, J. Phys. A **39**, 1265 (2006).
[15] J. P. L. Hatchett and Y. Kabashima, J. Phys. A **39**, 10659 (2006).
[16] M. Mézard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, Oxford, 2009).
[17] M. Mézard and G. Parisi, Eur. Phys. J. B **20**, 217 (2001); J. Stat. Phys. **111**, 1 (2003); H. Zhou, Fron. Phys. China **2**, 238 (2007).
[18] M. Mézard, G. Parisi, and R. Zecchina, Science **297**, 812 (2002); M. Mézard and R. Zecchina, Phys. Rev. E **66**, 056126 (2002).
[19] L. Zdeborová and M. Mézard, J. Stat. Mech.: Theory Exp. (2006) P05003; L. Zdeborová and F. Krząkała, Phys. Rev. E **76**, 031131 (2007); J. Zhou and H. Zhou, *ibid.* **79**, 020103(R) (2009); P. Zhang, Y. Zeng, and H. Zhou, *ibid.* **80**, 021122 (2009).
[20] I. Kanter and D. Saad, Phys. Rev. E **61**, 2137 (2000).
[21] R. Vicente, D. Saad, and Y. Kabashima, in *Advances in Imaging and Electron Physics*, edited by P. Hawkes (Academic Press, New York, 2002), Vol. 125, pp. 232–353.