

Statistical mechanics of an error correcting code using monotonic and nonmonotonic treelike multilayer perceptrons

Florent Cousseau*

Graduate School of Frontier Sciences, University of Tokyo, Chiba 277-5861, Japan

Kazushi Mimura†

Faculty of Information Sciences, Hiroshima City University, Hiroshima 731-3194, Japan

Masato Okada

*Graduate School of Frontier Sciences, University of Tokyo, Chiba 277-5861, Japan
and Brain Science Institute, RIKEN, Saitama 351-0198, Japan*

(Received 28 September 2009; published 1 February 2010)

An error correcting code using a treelike multilayer perceptron is proposed. An original message s^0 is encoded into a codeword y_0 using a treelike committee machine (committee tree) or a treelike parity machine (parity tree). Based on these architectures, several schemes featuring monotonic or nonmonotonic units are introduced. The codeword y_0 is then transmitted via a binary asymmetric channel where it is corrupted by noise. The analytical performance of these schemes is investigated using the replica method of statistical mechanics. Under some specific conditions, some of the proposed schemes are shown to saturate the Shannon bound at the infinite codeword length limit. The influence of the monotonicity of the units on the performance is also discussed.

DOI: [10.1103/PhysRevE.81.021104](https://doi.org/10.1103/PhysRevE.81.021104)

PACS number(s): 02.50.-r, 05.50.+q, 89.70.-a

I. INTRODUCTION

Reliability in communication has always been a major concern when dealing with digital data. Especially in today's information-dependent society, it is vital to design efficient ways of preventing data corruption when transmitting information. Error correcting codes have been developed for this purpose since the birth of the information theory field following the work of Shannon [1].

In 1989, Sourlas derived a set of error correcting codes, the so-called Sourlas codes, which theoretically saturate the Shannon bound [2]. Although these codes turned out to be impractical, the main point of interest of this paper was the parallel made between physical spin-glass systems and information theory.

Following this paper, the tools of statistical mechanics have been successfully applied to a wide range of problems of information theory in recent years. In the field of error correcting codes itself [3–5], as well as in spreading codes [6,7], and compression codes [8–13], statistical mechanical techniques have shown great potential.

The present paper uses similar techniques to investigate an error correcting code scheme where the codeword is encoded using treelike multilayer perceptron neural networks. It is known that there exists a natural duality between lossy compression codes and error correcting codes. Indeed, a lossy compression code can be regarded as a standard error correcting code, but one where the codeword is generated using the original decoder of the error correcting code scheme and where the decompressed message is obtained

using the original encoder of the scheme (cf. [14] for details).

Recently, a lossy compression scheme based on a simple perceptron decoder was investigated by Hosaka *et al.* [10]. In their paper, they used statistical mechanical techniques to investigate the theoretical performance of their scheme at the infinite codeword length limit. The perceptron they defined in their model uses a special hat-shaped nonmonotonic transfer function. This rather uncommon feature enables the scheme to deal with biased messages and it is known that this type of function maximizes the storage capacity of the simple perceptron [15,16]. They found that their scheme can theoretically yield Shannon optimal performance. Subsequently, Shinzato *et al.* [17] investigated the same model but in the framework of error correcting code. They found that their model can theoretically yield Shannon optimal performance.

Based on these studies, Mimura *et al.* [12] proposed a treelike multilayer perceptron network for lossy compression purposes but used only the standard sign function as the transfer function of their model. They showed that the parity tree (PT) model can theoretically yield Shannon optimal performance but only when considering unbiased messages. In contrast, they showed that the committee tree (CT) model cannot yield optimal performance even for unbiased messages. However, the advantage of using a multilayer structure is improved replica symmetric (RS) solution stability and an increased number of codewords sharing the same distortion properties [18]. In a recent study, Cousseau *et al.* [13] investigated the same treelike multilayer perceptron model but used the hat-shaped nonmonotonic transfer function introduced by Hosaka *et al.* [10], thus combining both advantages of [10,12]. By doing so, they were able to show that both parity tree and committee tree structures can then theoretically yield Shannon optimal performance even for biased messages under some specific conditions.

*florent@mns.k.u-tokyo.ac.jp

†mimura@hiroshima-cu.ac.jp

The purpose of the present paper is to discuss the performance of the same treelike perceptron models but in the error correcting code framework, thus completing the topic of perceptron type network applications in coding theory. In this paper, we make use of the binary asymmetric channel (BAC). Indeed, the use of the nonmonotonic hat-shaped transfer function introduced by Hosaka *et al.* [10] enables us to control the bias of the codeword sequence and enables the relevant schemes to deal with such an asymmetric channel (the BAC was also used by Shinzato *et al.* [17]). On the other hand, we expect the schemes which use the standard monotonic sign function to be able to deal only with the binary symmetric channel (BSC), which corresponds to a particular case of the BAC. The majority of popular error correcting codes such as turbo codes [19] and low-density parity check codes (LDPCs) [20,21], which provide near Shannon performance in practical time frames, have been widely studied but this was generally restricted to symmetric channels. On the other hand, apart from a few studies [22,23], little is known when dealing with asymmetric channels.

Multilayer perceptrons have been widely studied over the years by the machine learning community and a wide range of problems have been considered (storage capacity, learning rules, etc.). These works revealed nontrivial behaviors of even simple models such as the simple perceptron network, for example. Many of these previous results are summarized in Ref. [24]. The present analysis gives us an opportunity to discuss the difficulty of decoding for densely connected systems (or dense systems as opposed to sparsely connected systems such as LDPC codes for example) using a systematic manner in the context of multilayer networks. There has been relatively little discussion of dense systems mainly because of the computational cost which is obviously higher than for sparse systems. However, because of their rich randomness, dense systems can possibly be regarded as pseudo-random codes such as the dense limit of LDPC codes.

In this paper we mainly focus on the necessary conditions to get Shannon optimal performance. To discuss practical decoders, it is first necessary to investigate the optimality of our schemes. This includes discussion of the optimal parameters for the transfer function since we need to know these parameters to discuss the optimal decoder. In other words, we need a theoretical analysis of the performance before we can study the decoding problem.

The paper is organized as follows. Section II introduces the framework of error correcting codes. Section III describes our model. Section IV deals with the BAC capacity. Section V presents the mathematical tools used to evaluate the performance of the present scheme. Section VI states the results and elucidates the location of the phase transition, which characterizes the best achievable performance of the model. Section VII is devoted to the conclusion and discussion.

II. ERROR CORRECTING CODES

In a general scheme, an original message s^0 of size N is encoded into a codeword y_0 of size M by some encoding

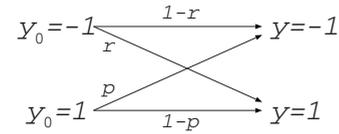


FIG. 1. The binary asymmetric channel (BAC).

device. The aim of this stage is to add redundancy to the original data. Therefore, we necessarily have $M > N$. Based on this redundancy, a proper decoder device should be able to recover the original data even if it were corrupted by noise in the transmission channel. The quantity $R = N/M$ is called the code rate and evaluates the trade-off between redundancy and codeword size. The codeword y_0 is then fed into a channel where the bits are subject to noise. The received noisy message y (which is also M dimensional) is then decoded using its redundancy to infer the original N dimensional message s^0 . In other words, in a Bayesian framework, one tries to maximize the following posterior probability:

$$P(s|y) \propto P(y|s)P(s). \tag{1}$$

As data transmission is costly, generally one wants to be able to ensure error-free transmission while transmitting the fewest possible bits. In other words, one wants to ensure error-free transmission while keeping the code rate as large as possible. For this purpose, the well-known Shannon bound [1] gives a way to compute the best achievable code rate which allows error-free recovery. However, while this gives us the value of such an optimal code rate, it does not give any clue as to how to construct such an optimal code. Therefore, several codes have been proposed over the years in an ongoing quest to find a code which can reach this theoretical bound.

III. ERROR CORRECTING CODES USING MONOTONIC AND NONMONOTONIC MULTILAYER PERCEPTRONS

In this paper, since we make use of techniques derived from statistical mechanics, we will use Ising variables rather than Boolean ones. The Boolean 0 is mapped onto 1 in the Ising framework while the Boolean 1 is mapped to -1 . This mapping can be used without any loss of generality.

We assume that the original message s_0 is generated from the uniform distribution and that all the bits are independently generated so that we have

$$P(s^0) = \frac{1}{2^N}. \tag{2}$$

The channel considered in this study is the BAC where each bit is flipped independently of the others with asymmetric probabilities. If the original bit fed into the channel is 1, then it is flipped with probability p . Conversely, if the original bit is -1 , it is flipped with probability r . Figure 1 shows the BAC properties in details.

The well-known BSC corresponds to the particular case $r = p$.

When the corrupted message y is received at the output of the channel, the goal is then to recover s^0 using y . The state

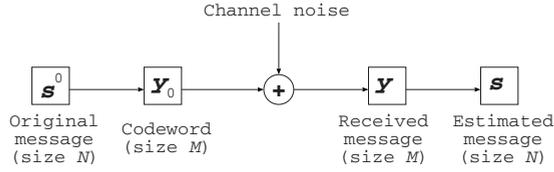


FIG. 2. Layout of the scheme.

of the estimated message is denoted by the vector s . The general outline of the scheme is shown in Fig. 2.

From Fig. 1 we can easily derive the following conditional probability:

$$P(y^\mu | y_0^\mu) = \frac{1}{2} + \frac{y^\mu}{2} [(1-r-p)y_0^\mu + (r-p)], \quad (3)$$

where we make use of the notations $y_0 = (y_0^1, \dots, y_0^\mu, \dots, y_0^M)$ and $y = (y^1, \dots, y^\mu, \dots, y^M)$. Since we assume that the bits are flipped independently, we deduce

$$P(y | y_0) = \prod_{\mu=1}^M \left\{ \frac{1}{2} + \frac{y^\mu}{2} [(1-r-p)y_0^\mu + (r-p)] \right\}. \quad (4)$$

To encode the original message s^0 into a codeword y_0 , we use three nonmonotonic treelike parity machine or committee machine neural networks [(I), (II), and (III)]. In the same way, we also investigate the standard monotonic parity tree and committee tree neural networks [(IV) and (V)].

(I) Multilayer parity tree with nonmonotonic hidden (PTH) units;

$$y_0^\mu(s^0) \equiv \prod_{l=1}^K f_k \left(\sqrt{\frac{K}{N}} s_l^0 \cdot x_l^\mu \right). \quad (5)$$

(II) Multilayer committee tree with nonmonotonic hidden (CTH) units;

$$y_0^\mu(s^0) \equiv \text{sgn} \left(\sum_{l=1}^K f_k \left[\sqrt{\frac{K}{N}} s_l^0 \cdot x_l^\mu \right] \right). \quad (6)$$

Note that in this case, if the number of hidden units K is even, it is possible to get 0 as the argument of the sign function. We avoid this uncertainty by considering only an odd number of hidden units for the committee tree with nonmonotonic hidden units in the sequel.

(III) Multilayer committee tree with a nonmonotonic output (CTO) unit;

$$y_0^\mu(s^0) \equiv f_k \left(\sqrt{\frac{1}{K}} \sum_{l=1}^K \text{sgn} \left[\sqrt{\frac{K}{N}} s_l^0 \cdot x_l^\mu \right] \right). \quad (7)$$

(IV) Multilayer PT;

$$y_0^\mu(s^0) \equiv \prod_{l=1}^K \text{sgn} \left(\sqrt{\frac{K}{N}} s_l^0 \cdot x_l^\mu \right). \quad (8)$$

(V) Multilayer CT;

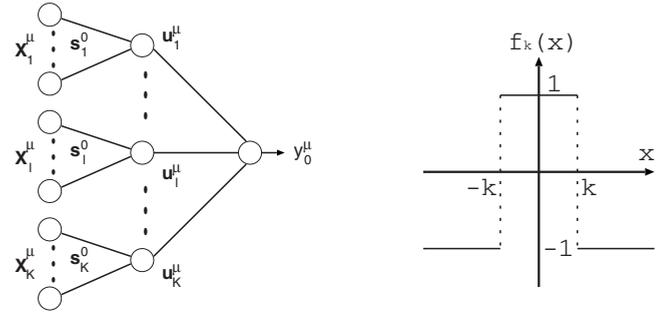


FIG. 3. Left: general architecture of the treelike multilayer perceptrons with N input units and K hidden units. Right: the non-monotonic function f_k .

$$y_0^\mu(s^0) \equiv \text{sgn} \left(\sqrt{\frac{1}{K}} \sum_{l=1}^K \text{sgn} \left[\sqrt{\frac{K}{N}} s_l^0 \cdot x_l^\mu \right] \right). \quad (9)$$

In this case also, if the number of hidden units K is even, it is possible to get 0 as the argument of the sign function. We again avoid this uncertainty by considering only an odd number of hidden units for the committee tree in the sequel.

The original message s^0 is split into N/K -dimensional K disjoint vectors so that s^0 can be written $s^0 = (s_1^0, \dots, s_K^0)$. In schemes (I), (II), and (III), f_k is a nonmonotonic function of a real parameter k of the form

$$f_k(x) = \begin{cases} 1 & \text{if } |x| \leq k \\ -1 & \text{if } |x| > k, \end{cases} \quad (10)$$

and the vectors x_l^μ are fixed N/K -dimensional independent vectors uniformly distributed on $\{-1, 1\}$. The use of random input vectors is known to maximize the storage capacity of perceptron networks, making such a scheme promising for error correcting tasks. The sgn function denotes the sign function taking 1 for $x \geq 0$ and -1 for $x < 0$. Each of these architectures applies a different nonlinear transformation to the original data s^0 . The general architecture of these perceptron-based encoders and the nonmonotonic function f_k are displayed in Fig. 3.

Note that we can also consider an encoder based on a committee tree where both the hidden units and the output unit are nonmonotonic. However, this introduces an extra parameter (we will have one threshold parameter for the hidden units and one for the output unit) to tune and the performance should not change drastically. For simplicity, we restrict our study to the above three cases.

To keep the notation as general as possible, as long as explicit use of the encoder is not necessary in computations, we will denote the transformation performed on vector s by the respective encoders using the following notation:

$$\mathcal{F}_k \left(\left\{ \sqrt{\frac{K}{N}} s_l \cdot x_l^\mu \right\} \right). \quad (11)$$

\mathcal{F}_k takes a different expression for the five different types of network and k denotes the fact that all the encoders depend on a real threshold parameter k [except for schemes (IV) and (V), where this function does not depend on k ; however for

consistency, we will keep this notation for these schemes]. Furthermore, note that \mathcal{F}_k contains all the terms depending on index l [i.e., $\mathcal{F}_k(\{u_l\})$ contains all the terms $u_1, \dots, u_l, \dots, u_K$].

IV. BAC CAPACITY

In this section, we compute the capacity of the BAC. According to Shannon's channel coding theorem, the optimal code rate is given by the capacity of the channel. Any code rate bigger than the channel capacity will inevitably lead to information loss. The definition of the channel capacity C is

$$C = \max_{\text{input probability}} \{I(X, Y)\}, \quad (12)$$

where I denotes mutual information, X denotes the channel input distribution, and Y denotes the channel output distribution. Computation of the capacity of such a binary channel requires only simple algebra and calculations are straightforward, giving

$$C_{BAC} = H_2(\gamma_C) - \frac{1 + \Omega_C}{2} H_2(p) - \frac{1 - \Omega_C}{2} H_2(r), \quad (13)$$

where

$$H_2(x) = -x \log_2(x) - (1-x) \log_2(1-x), \quad (14)$$

$$\gamma_C = \frac{1}{1 + \Delta_C} = \frac{1}{2} [(1-p)(1 + \Omega_C) + r(1 - \Omega_C)], \quad (15)$$

$$\Delta_C = \left[\frac{r^r(1-r)^{1-r}}{p^p(1-p)^{1-p}} \right]^{1/(1-r-p)}, \quad (16)$$

$$\Omega_C = \frac{2\gamma_C - 1 - r + p}{1 - r - p}. \quad (17)$$

In the special case $r=p$, the capacity simplifies to

$$C_{BSC} = 1 - H_2(p), \quad (18)$$

which corresponds to the capacity of the BSC.

V. ANALYTICAL EVALUATION

As stated in Sec. II, our goal is to maximize the posterior $P(s|y)$. Let us define the following Hamiltonian:

$$\mathcal{H}(y, s) = -\ln[P(s|y)P(s)] = -\ln P(y, s). \quad (19)$$

The ground state of the above Hamiltonian trivially corresponds to the *maximum a posteriori* (MAP) estimator of the

posterior $P(s|y)$. Then, let us compute the joint probability of y and s . We have

$$P(y, s) = P(y|s)P(s). \quad (20)$$

Since the relation between an arbitrary message s and the codeword fed into the channel is deterministic, for any s , we can write

$$\begin{aligned} P(y|s) &= P\left(y | \mathcal{F}_k\left(\left\{\sqrt{\frac{K}{N}} s_l \cdot \mathbf{x}_l^\mu\right\}\right)\right), \\ &= \prod_{\mu=1}^M \left\{ \frac{1}{2} + \frac{y^\mu}{2} \left[(1-r-p) \mathcal{F}_k\left(\left\{\sqrt{\frac{K}{N}} s_l \cdot \mathbf{x}_l^\mu\right\}\right) \right. \right. \\ &\quad \left. \left. + (r-p) \right] \right\}. \end{aligned} \quad (21)$$

We finally get the explicit expression of the Hamiltonian,

$$\begin{aligned} \mathcal{H}(y, s) &= -\ln P(y, s) \\ &= -\ln \left[\frac{1}{2^N} \prod_{\mu=1}^M \left\{ \frac{1}{2} + \frac{y^\mu}{2} \left[(1-r-p) \right. \right. \right. \\ &\quad \left. \left. \left. \times \mathcal{F}_k\left(\left\{\sqrt{\frac{K}{N}} s_l \cdot \mathbf{x}_l^\mu\right\}\right) + (r-p) \right] \right\} \right]. \end{aligned} \quad (22)$$

Using this Hamiltonian, we can define the following partition function:

$$Z(\beta, y, \mathbf{x}) = \sum_s \exp[-\beta \mathcal{H}(y, s)], \quad (23)$$

where the sum over s represents the sum over all possible states for vector s , and β is the inverse temperature parameter. Such a partition function can be identified with the partition function of a spin-glass system with dynamical variables s and quenched variables \mathbf{x} . The average of this partition function over y and \mathbf{x} naturally contains all the interesting typical properties of the scheme, such as the free energy. However, it is hard to evaluate this average and we need some techniques to investigate it. In this paper, we use the so-called *replica method* to calculate the average of the partition function. Once the free energy is obtained, one can compute the critical code rate at which a phase transition occurs between the ferromagnetic phase (error recovery possible) and the paramagnetic phase (decoding impossible). This gives us the best code rate the scheme can achieve. A code rate exceeding this critical value will make decoding impossible. The calculations to obtain the average of the partition function $\langle Z(\beta, y, \mathbf{x}) \rangle_{y, \mathbf{x}}$ are detailed in Appendix.

After long calculations, the RS free energy is obtained,

$$\begin{aligned} -f_{RS}(q, \hat{q}, m, \hat{m}) &= \text{extr}_{q, \hat{q}, m, \hat{m}} \left\{ \sum_{y=\pm 1} \int_{-\infty}^{\infty} \left(\prod_{l=1}^K DR_l \right) \int_{-\infty}^{\infty} \left(\prod_{l=1}^K Dt_l \right) \ln[I(y, R_l, t_l, m, q)] \left(\frac{1}{2} + \frac{y}{2} \{ (1-r-p) \mathcal{F}_k(\{R_l\}) + (r-p) \} \right) \right. \\ &\quad \left. + R \int_{-\infty}^{\infty} DU \ln[2 \cosh(\sqrt{\hat{q}}U + \hat{m})] - R \ln 2 - Rm\hat{m} - R \frac{\hat{q}(1-q)}{2} \right\}, \end{aligned} \quad (24)$$

where

$$I(y, R_l, t_l, m, q) = \int_{-\infty}^{\infty} \left[\prod_{l=1}^K D z_l \right] \times \left\{ \frac{1}{2} + \frac{y}{2}(r-p) + \frac{y}{2}(1-r-p) \mathcal{F}_k(\{\sqrt{1-qz_l} + \sqrt{q-m^2 t_l + m R_l}\}) \right\}, \quad (25)$$

$$Dx = \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx. \quad (26)$$

and where extr denotes extremization. The sum denotes the sum over all possible states for the variable y , that is, ± 1 .

Note also that we set $\beta=1$. This choice of finite temperature decoding (in contrast to $\beta \rightarrow \infty$ which corresponds to the zero temperature limit) corresponds to the *maximizer of posterior marginals* (MPMs) estimator, while the zero temperature decoding corresponds to the MAP estimator [25,28]. The MPM estimator is known to be optimal for the purpose of decoding [26–28]. On top of that, in this paper we suppose that all the channel properties [i.e., the true values of (p, r)] are known to the decoder which implies that the system's state we consider is located on the Nishimori line [26,27].

To retrieve the free energy one has to extremize Eq. (24) with respect to the order parameters q, \hat{q}, m, \hat{m} . This is done by solving the following saddle-point equations:

$$\frac{\partial f_{RS}}{\partial q} = 0 \Leftrightarrow \hat{q} = -2R^{-1} \sum_{y=\pm 1} \int_{-\infty}^{\infty} \left[\prod_{l=1}^K D R_l \right] \int_{-\infty}^{\infty} \left[\prod_{l=1}^K D t_l \right] \times \frac{I'_q(y, R_l, t_l, m, q)}{I(y, R_l, t_l, m, q)} \times \left(\frac{1}{2} + \frac{y}{2} \{(1-r-p) \mathcal{F}_k(\{R_l\}) + (r-p)\} \right), \quad (27)$$

$$\frac{\partial f_{RS}}{\partial m} = 0 \Leftrightarrow \hat{m} = R^{-1} \sum_{y=\pm 1} \int_{-\infty}^{\infty} \left[\prod_{l=1}^K D R_l \right] \int_{-\infty}^{\infty} \left[\prod_{l=1}^K D t_l \right] \times \frac{I'_m(y, R_l, t_l, m, q)}{I(y, R_l, t_l, m, q)} \times \left(\frac{1}{2} + \frac{y}{2} \{(1-r-p) \mathcal{F}_k(\{R_l\}) + (r-p)\} \right), \quad (28)$$

$$\frac{\partial f_{RS}}{\partial \hat{q}} = 0 \Leftrightarrow q = \int_{-\infty}^{\infty} D U \tanh^2(\sqrt{\hat{q}} U + \hat{m}), \quad (29)$$

$$\frac{\partial f_{RS}}{\partial \hat{m}} = 0 \Leftrightarrow m = \int_{-\infty}^{\infty} D U \tanh(\sqrt{\hat{q}} U + \hat{m}), \quad (30)$$

where

$$I'_q(y, R_l, t_l, m, q) = \frac{\partial I(y, R_l, t_l, m, q)}{\partial q}, \quad (31)$$

$$I'_m(y, R_l, t_l, m, q) = \frac{\partial I(y, R_l, t_l, m, q)}{\partial m}. \quad (32)$$

An error correcting code scheme typically admits two solutions: one where $m=q=1$, called the ferromagnetic solution and one where $m=q=0$, called the paramagnetic solution. As the names indicate, these solutions come from the physical ferromagnet state and correspond to the case where the spins are all ordered ($m=q=1$) or to the case where the spins take completely random states ($m=q=0$). As we can deduce from Eqs. (A3) and (A6), the ferromagnetic solution corresponds to decoding success since $m=1$ implies perfect overlap. Conversely, the paramagnetic phase implies failure in the decoding process (overlap m is 0).

A. Replica symmetric solution using a parity tree with nonmonotonic hidden units

Using a parity tree with nonmonotonic hidden units [Eq. (5)], the encoder function becomes

$$\mathcal{F}_k(\{u_l\}) = \prod_{l=1}^K f_k(u_l). \quad (33)$$

Using this encoder function and substituting $m=q=0$ in the saddle-point equations, one can find a consistent solution where $q=m=\hat{q}=\hat{m}=0$. This corresponds to the paramagnetic solution, where decoding of the received message fails. Using these conditions in Eq. (24), one can retrieve the free energy of the paramagnetic phase,

$$-f_{para} = -H_2 \left\{ \frac{1}{2} [(1-p)(1 + \Omega_{PTH}) + r(1 - \Omega_{PTH})] \right\} \times \ln 2, \quad (34)$$

where

$$\Omega_{PTH} = \prod_{l=1}^K \int_{-\infty}^{+\infty} D z_l f_k(z_l). \quad (35)$$

In the same way, substituting $m=q=1$ in the saddle-point equations, one can find a consistent solution. However, the ferromagnetic solution cannot be computed analytically. So we proceed numerically by simply checking the integrand of Eqs. (27) and (28). We did that extensively for values of $K=1$, $K=2$, and $K=3$. In each case we found that the integrand diverges so that when $(q, m) \rightarrow (1, 1)$, we have both $\hat{q} \rightarrow \infty$ and $\hat{m} \rightarrow \infty$. Substituting $\hat{q} \rightarrow \infty$ and $\hat{m} \rightarrow \infty$ into Eqs. (29) and (30) clearly yields $q=m=1$. So $q=m=1$, $\hat{q} \rightarrow \infty$, and $\hat{m} \rightarrow \infty$ is a consistent solution of the saddle-point equations which corresponds to the ferromagnetic solution, where decoding of the received message succeeds. We also checked higher values of K (up to $K=5$) and did not find any other consistent solution. We conjecture that this result holds for any finite value of K . Finally, substituting $m=q=1$, $\hat{m} \rightarrow \infty$, and $\hat{q} \rightarrow \infty$ into Eq. (24), one can get the free energy of the ferromagnetic phase,

$$-f_{ferro} = -\frac{\ln 2}{2}[(1 + \Omega_{PTH})H_2(p) + (1 - \Omega_{PTH})H_2(r)] - R \ln 2. \quad (36)$$

Note that when $K=1$, the present scheme corresponds to the case of Shinzato *et al.* [17]. The result we obtained when $K=1$ is indeed equivalent to what they found.

B. Replica symmetric solution using a committee tree with nonmonotonic hidden units

When a committee tree with nonmonotonic hidden units [Eq. (6)] is used, the encoder function becomes

$$\mathcal{F}_k(\{u_l\}) = \text{sgn} \left[\sum_{l=1}^K f_k(u_l) \right]. \quad (37)$$

Using this encoder function and substituting $m=q=0$ in the saddle-point equations, one can find a consistent solution where $q=m=\hat{q}=\hat{m}=0$. This corresponds to the paramagnetic solution, where decoding of the received message fails. Using these conditions in Eq. (24), one can retrieve the free energy of the paramagnetic phase,

$$-f_{para} = -H_2 \left\{ \frac{1}{2} [(1-p)(1 + \Omega_{CTH}) + r(1 - \Omega_{CTH})] \right\} \times \ln 2, \quad (38)$$

where

$$\Omega_{CTH} = \int_{-\infty}^{+\infty} \left[\prod_{l=1}^K D z_l \right] \times \text{sgn} \left[\sum_{l=1}^K f_k(z_l) \right]. \quad (39)$$

In the same way, by substituting $m=q=1$ in the saddle-point equations one can find a consistent solution. However, the ferromagnetic solution cannot be computed analytically, so we proceed numerically by simply checking the integrand of Eqs. (27) and (28). We did that extensively for $K=3$ (we consider only odd values of K for this scheme and when $K=1$ the present scheme is equivalent to the parity tree case). We found that the integrand diverges so that when $(q, m) \rightarrow (1, 1)$, we have both $\hat{q} \rightarrow \infty$ and $\hat{m} \rightarrow \infty$. We also checked higher values of K (up to $K=5$) and did not find any other consistent solution. We conjecture that this result holds for any finite value of K . Finally, substituting $m=q=1$, $\hat{m} \rightarrow \infty$, and $\hat{q} \rightarrow \infty$ into Eq. (24), one can get the free energy of the ferromagnetic phase,

$$-f_{ferro} = -\frac{\ln 2}{2}[(1 + \Omega_{CTH})H_2(p) + (1 - \Omega_{CTH})H_2(r)] - R \ln 2. \quad (40)$$

C. Replica symmetric solution using a committee tree with a nonmonotonic output unit

When a committee tree with a nonmonotonic output unit [Eq. (7)] is used, the encoder function becomes

$$\mathcal{F}_k(\{u_l\}) = f_k \left[\sqrt{\frac{1}{K}} \sum_{l=1}^K \text{sgn}(u_l) \right]. \quad (41)$$

Using this encoder function and substituting $m=q=0$ in the saddle-point equations do not imply $\hat{m}=\hat{q}=0$ and a nontrivial solution is found, which makes the free energy too complex to be investigated. This scheme is likely to give nonoptimal performance in such a case and will not be considered in what follows.

Note that the limit where $K \rightarrow \infty$ was not studied because the saddle-point equations take a nontrivial form that is difficult to investigate (in the lossy compression case, this study is still tractable). The techniques to investigate the free energy in the $K \rightarrow \infty$ limit described in Ref. [24] cannot be easily applied here. However, based on the previous results of Cousseau *et al.* [13], it is probable that in the $K \rightarrow \infty$ limit, the committee tree with a nonmonotonic output unit saturates the Shannon bound in the general BAC case.

D. Replica symmetric solution using a parity tree

Using a parity tree [Eq. (8)], the encoder function becomes

$$\mathcal{F}_k(\{u_l\}) = \prod_{l=1}^K \text{sgn}(u_l). \quad (42)$$

Using this encoder function and substituting $m=q=0$ in the saddle-point equations, one can find a consistent solution where $q=m=\hat{q}=\hat{m}=0$ but only when $K > 1$. This corresponds to the paramagnetic solution, where decoding of the received message fails. Using these conditions in Eq. (24), one can retrieve the free energy of the paramagnetic phase,

$$-f_{para} = -H_2 \left\{ \frac{1}{2} [(1-p)(1 + \Omega_{PT}) + r(1 - \Omega_{PT})] \right\} \times \ln 2, \quad (43)$$

where

$$\Omega_{PT} = \prod_{l=1}^K \int_{-\infty}^{+\infty} D z_l \times \text{sgn}(z_l). \quad (44)$$

When $K=1$ is considered, $m=q=0$ does not imply $\hat{m}=\hat{q}=0$ and a nontrivial solution is found that makes the free energy too complex to be investigated. The scheme is likely to give nonoptimal performance in such a case and will not be considered in what follows.

In the same way, substituting $m=q=1$ in the saddle-point equations, one can find a consistent solution but only when $K > 1$. However, the ferromagnetic solution cannot be computed analytically, so we proceed numerically by simply checking the integrand of Eqs. (27) and (28). We did that extensively for values of $K=2$ and $K=3$. In each case, we found that the integrand diverges so that when $(q, m) \rightarrow (1, 1)$ we have both $\hat{q} \rightarrow \infty$ and $\hat{m} \rightarrow \infty$. We also checked higher values of K (up to $K=5$) and did not find any other

consistent solution. We conjecture that this result holds for any finite value of $K > 1$. Finally, substituting $m=q=1$, $\hat{m} \rightarrow \infty$, and $\hat{q} \rightarrow \infty$ into Eq. (24), one can get the free energy of the ferromagnetic phase,

$$-f_{ferro} = -\frac{\ln 2}{2}[(1 + \Omega_{PT})H_2(p) + (1 - \Omega_{PT})H_2(r)] - R \ln 2. \quad (45)$$

E. Replica symmetric solution using a committee tree

Using a committee tree [Eq. (9)], the encoder function becomes

$$\mathcal{F}_k(\{u_i\}) = \text{sgn} \left[\sqrt{\frac{1}{K} \sum_{l=1}^K \text{sgn}(u_l)} \right]. \quad (46)$$

Using this encoder function and substituting $m=q=0$ in the saddle-point equations do not imply $\hat{m}=\hat{q}=0$ and a nontrivial solution is found that makes the free energy too complex to be investigated. This scheme is likely to give nonoptimal performance in such a case and will not be considered in what follows. As in the lossy compression case [12], the committee tree is unable to yield Shannon optimal performance.

Note that the limit where $K \rightarrow \infty$ was not studied because the saddle-point equations take a nontrivial form that is difficult to investigate (in the lossy compression case, this study is still tractable). The techniques to investigate the free energy in the $K \rightarrow \infty$ limit described in reference [24] cannot be easily applied here. However, based on the previous results of Mimura *et al.* [12], it is probable that in the $K \rightarrow \infty$ limit the committee tree still fails to saturate the Shannon bound even in the BSC case.

VI. PHASE TRANSITION

For the parity and committee tree with nonmonotonic hidden units and for the standard parity tree, we found a paramagnetic and a ferromagnetic solution of the following form:

$$-f_{para} = -H_2 \left\{ \frac{1}{2}[(1-p)(1+\Omega) + r(1-\Omega)] \right\} \times \ln 2, \quad (47)$$

$$-f_{ferro} = -\frac{\ln 2}{2}[(1+\Omega)H_2(p) + (1-\Omega)H_2(r)] - R \ln 2, \quad (48)$$

where Ω is given by Ω_{PTH} , Ω_{CTH} , or Ω_{PT} depending on the encoder considered.

It then becomes possible to calculate the critical value of the code rate R at which a sharp phase transition occurs between the ferromagnetic and the paramagnetic phase. This indicates the boundary between possible decoding (ferromagnetic phase) and impossible decoding (paramagnetic

phase). In other words, this enables us to calculate the optimal code rate for each scheme. At the phase transition point, we have

$$f_{para} = f_{ferro}. \quad (49)$$

Simple algebra leads to

$$R = H_2(\gamma) - \frac{1+\Omega}{2}H_2(p) - \frac{1-\Omega}{2}H_2(r), \quad (50)$$

where

$$\gamma = \frac{1}{2}[(1-p)(1+\Omega) + r(1-\Omega)] \quad (51)$$

and where Ω is given by the encoder considered (Ω_{PTH} , Ω_{CTH} , or Ω_{PT}). This equation has exactly the same form as the BAC capacity Eq. (13) and in fact is equivalent to the BAC capacity if and only if $\Omega = \Omega_C$. Since Ω depends on the encoder, we will treat each case in the following subsections.

A. Tuning of the parity tree with nonmonotonic hidden units

In the parity tree with nonmonotonic hidden units case, we have

$$\Omega \equiv \Omega_{PTH} = \prod_{l=1}^K \int_{-\infty}^{+\infty} D z_l f_k(z_l). \quad (52)$$

The parity tree with nonmonotonic hidden units is optimal if and only if

$$\Omega_{PTH} = \Omega_C \Leftrightarrow H(k) = \frac{1}{4}(1 - \sqrt[k]{\Omega_C}), \quad (53)$$

where

$$H(x) = \int_x^{+\infty} D z. \quad (54)$$

This gives us a condition on the threshold parameter k of the nonmonotonic transfer function f_k . If the threshold k is tuned to satisfy Eq. (53), the scheme achieves the Shannon limit. The only remaining issue is whether such an optimal threshold k exists.

We solved Eq. (53) numerically with parameters $(p, r) \in \{0, 1\}^2$ and always found an optimal threshold parameter k up to $K=11$. Note that Ω_C can be negative, which causes problems for the K th root when considering an even number of hidden units K . However a simple permutation of the probability p and r changes the sign of Ω_C . Since the original messages are drawn from the uniform distribution, this permutation can be done without any loss of generality. Instead of using s_0 , one uses $-s_0$. We did not check higher values of K , but we conjecture that the same result holds. This means that the parity tree with nonmonotonic hidden units saturates the Shannon bound in the large codeword length limit for any number of hidden units K .

B. Tuning of the committee tree with nonmonotonic hidden units

In the committee tree with non monotonic hidden units case, we have

$$\Omega \equiv \Omega_{CTH} = \int_{-\infty}^{+\infty} \left[\prod_{l=1}^K Dz_l \right] \times \text{sgn} \left[\sum_{l=1}^K f_k(z_l) \right]. \quad (55)$$

The committee tree with nonmonotonic hidden units is optimal if and only if

$$\Omega_{CTH} = \Omega_C \Leftrightarrow \Omega_C = \sum_{l=0}^{\frac{K-1}{2}} \binom{K}{l} \{ [2H(k)]^l [1 - 2H(k)]^{K-l} - [2H(k)]^{K-l} [1 - 2H(k)]^l \}, \quad (56)$$

where $\binom{x}{y}$ denotes the binomial coefficient. This gives us a condition on the threshold parameter k of the nonmonotonic transfer function f_k . If the threshold k is tuned to satisfy Eq. (56), the scheme achieves the Shannon limit. Thus, we should check if such an optimal threshold k exists.

We solved Eq. (56) numerically with parameters $(p, r) \in \{]0, 1[\}^2$ and always found an optimal threshold parameter k up to $K=11$. We did not check higher values of K , but we conjecture that the same result holds. Note that as mentioned in the definition of this encoder, we considered only an odd number of hidden units K . Therefore, these results mean that the committee tree with nonmonotonic hidden units saturates the Shannon bound in the large codeword length limit for any odd number of hidden units K .

C. Tuning of the parity tree

In the parity tree case, we have

$$\Omega \equiv \Omega_{PT} = \prod_{l=1}^K \int_{-\infty}^{+\infty} Dz_l \times \text{sgn}(z_l). \quad (57)$$

The parity tree is optimal if and only if

$$\Omega_{PT} = \Omega_C \Leftrightarrow \Omega_C = 0. \quad (58)$$

This gives us a strong condition on Ω_C . From the definition (17), it can be easily seen that $\Omega_C=0$ if and only if $r=p$: that is when the BAC channel turns into the particular case of the BSC channel. This means that the standard monotonic parity tree saturates the Shannon bound in the large codeword length limit, but only in the BSC case and for a number of hidden units $K > 1$. This confirms what we expected and is the equivalent of Mimura *et al.* [12] lossy compression case.

VII. CONCLUSION AND DISCUSSION

We investigated an error correcting code scheme for uniformly unbiased Boolean messages using parity tree and committee tree multilayer perceptrons. All the schemes which use the nonmonotonic transfer function f_k in their

hidden layer were shown to saturate the Shannon bound under some specific conditions. The use of f_k enables the relevant schemes to deal with asymmetric channels such as the BAC while monotonic networks using only the standard sign function can deal only with symmetric channels such as the BSC.

Indeed, we confirmed that the standard monotonic parity tree saturates the Shannon bound only in the case of the BSC channel. The standard monotonic committee tree, however, fails to provide optimal performance even in the BSC case.

As a general conclusion, this paper shows that treelike multilayer perceptrons introduced in [10,12,13] within the framework of lossy compression can also be used efficiently in an error correcting code scheme. For each network considered, we provided a theoretical analysis of the typical performance and gave the necessary conditions for obtaining optimal performance. In each case, we were able to derive results similar to the lossy compression results. Finally, in the case of error correcting code, the replica symmetric solution stability [18] was not checked because no replica symmetry breaking is expected on the Nishimori line [29].

This paper discusses only the typical performance of the schemes at the infinite codeword length, however, and does not provide any explicit decoder. Because the present schemes make use of densely connected systems, a formal decoder cannot be implemented as it would require a decoding time which would grow exponentially with the size of the original message. One promising alternative is to use the popular belief propagation (BP) algorithm to calculate an approximation of the marginalized posterior probabilities. The BP algorithm is known for giving good results when working in the ferromagnetic phase, where no frustration is present into the system.

With the previous work done on lossy compression [10,12,13,30] and on error correcting code [17] using perceptron type networks, there is now a sufficient theoretical background to investigate and compare the practical performance (in the finite codeword length limit) of all the schemes with the theoretical performance. In the case of lossy compression with a simple perceptron, the study of the BP algorithm performance has already been done by Hosaka *et al.* [30]. Their work provides a solid base from which to begin investigating the more complicated multilayer structure. The influence of the number of hidden units on the practical performance of the scheme is an interesting issue which will be examined in future work.

ACKNOWLEDGMENTS

This work was partially supported by a Grant-in-Aid for Encouragement of Young Scientists (B) (Grant No. 18700230), Grant-in-Aid for Scientific Research on Priority Areas (Grants No. 18079003 and No. 18020007), Grant-in-Aid for Scientific Research (C) (Grant No. 16500093), and a Grant-in-Aid from JSPS Foundations (Grant No. 06J06774) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

**APPENDIX: ANALYTICAL EVALUATION
USING THE REPLICA METHOD**

The free energy can be evaluated by the replica method,

$$f(\beta, R) = -\frac{1}{\beta N} \lim_{n \rightarrow 0} \frac{\langle Z(\beta, \mathbf{y}, \mathbf{x})^n \rangle_{\mathbf{y}, \mathbf{x}} - 1}{n} \quad (\text{A1})$$

where $Z(\beta, \mathbf{y}, \mathbf{x})^n$ denotes the n -times replicated partition function

$$Z(\beta, \mathbf{y}, \mathbf{x})^n = \sum_{s^1, \dots, s^n} \prod_{a=1}^n \exp\{-\beta \mathcal{H}[\mathbf{y}, \hat{\mathbf{y}}(s^a)]\}. \quad (\text{A2})$$

Vector s^a is given by $s^a = (s_1^a, \dots, s_K^a)$ and superscript a denotes the replica index.

We proceed to the calculation of the replicated partition function (A2). Inserting the following two identities:

$$1 = \prod_{a=1}^n \prod_{l=1}^K \int_{-\infty}^{+\infty} dm_l^a \delta\left(s_l^0 \cdot s_l^a - \frac{N}{K} m_l^a\right) = \left(\frac{1}{2\pi i}\right)^{nK} \int \left(\prod_a \prod_l dm_l^a d\hat{m}_l^a\right) \exp\left[\sum_a \sum_l \hat{m}_l^a \left(s_l^0 \cdot s_l^a - \frac{N}{K} m_l^a\right)\right] \quad (\text{A3})$$

and

$$1 = \prod_{a<b}^n \prod_{l=1}^K \int_{-\infty}^{+\infty} dq_l^{ab} \delta\left(s_l^a \cdot s_l^b - \frac{N}{K} q_l^{ab}\right) = \left(\frac{1}{2\pi i}\right)^{n(n-1)K/2} \int \left(\prod_{a<b} \prod_l dq_l^{ab} d\hat{q}_l^{ab}\right) \exp\left[\sum_{a<b} \sum_l \hat{q}_l^{ab} \left(s_l^a \cdot s_l^b - \frac{N}{K} q_l^{ab}\right)\right] \quad (\text{A4})$$

into Eq. (A2) enables us to separate the relevant order parameters and to calculate the average moment $\langle Z(\beta, \mathbf{y}, \mathbf{x})^n \rangle_{\mathbf{y}, \mathbf{x}}$ for natural numbers n as

$$\begin{aligned} \langle Z(\beta, \mathbf{y}, \mathbf{x})^n \rangle_{\mathbf{y}, \mathbf{x}} &\approx \int \left(\prod_a \prod_l dm_l^a \frac{d\hat{m}_l^a}{2\pi i}\right) \int \left(\prod_{a<b} \prod_l dq_l^{ab} \frac{d\hat{q}_l^{ab}}{2\pi i}\right) \exp\left\{N \left[R^{-1} \ln \left[\sum_y \int \left(\prod_l du_l \frac{d\mathbf{v}_l}{2\pi} dR_l \frac{dW_l}{2\pi}\right) \right. \right. \right. \\ &\quad \times \left. \left. \left. \left(\frac{1}{2} + \frac{y}{2} [(1-r-p)\mathcal{F}_k(\{R_l\}) + (r-p)] \right) \prod_a \left\{ \exp \left[\beta \ln \left(\frac{1}{2} + \frac{y}{2} [(1-r-p)\mathcal{F}_k(\{u_l^a\}) + (r-p)] \right) \right] \right\} \right. \right. \right. \\ &\quad \times \left. \left. \left. \prod_l \left\{ \exp \left[-\frac{1}{2} (W_l)^2 - \frac{1}{2} \mathbf{v}_l \cdot \mathcal{Q}_l \cdot \mathbf{v}_l - W_l \mathcal{M}_l \cdot \mathbf{v}_l + iR_l W_l + i\mathbf{v}_l \cdot \mathbf{u}_l \right] \right\} \right] \right\} \\ &\quad \left. + \frac{1}{K} \ln \left[\sum_{s^a} \exp \left[\sum_{a,l} \hat{m}_l^a s_l^a + \sum_{a<b,l} \hat{q}_l^{ab} s_l^a s_l^b \right] \right] - \frac{1}{K} \sum_{a,l} m_l^a \hat{m}_l^a - \frac{1}{K} \sum_{a<b,l} q_l^{ab} \hat{q}_l^{ab} - \beta \sum_a \ln 2 \right] \right\}, \quad (\text{A5}) \end{aligned}$$

where \mathcal{Q}_l is an $n \times n$ matrix having elements $\{q_l^{ab}\}$ and where \mathcal{M}_l is an n dimensional vector having elements $\{m_l^a\}$. We analyze the scheme at the thermodynamic limit $N, M \rightarrow +\infty$ while the code rate R is kept finite. In this limit, Eq. (A5) can be evaluated using the saddle-point method with respect to $m^a, \hat{m}^a, q_l^{ab}, \hat{q}_l^{ab}$ so that the free energy can be retrieved. To continue the calculation, we have to make some assumptions about the structure of these order parameters. In this paper, we use the so-called replica symmetric (RS) ansatz,

$$m_l^a = m, \quad q_l^{ab} = (1-q)\delta_{ab} + q,$$

$$\hat{m}_l^a = \hat{m}, \quad \hat{q}_l^{ab} = (1-\hat{q})\delta_{ab} + \hat{q}, \quad (\text{A6})$$

where δ_{ab} denotes the Kronecker delta. This ansatz means that all the hidden units are equivalent after averaging over the disorder.

Also note that by definition, order parameter m is equivalent to quantity $s^0 \cdot s/N$, which gives the overlap between the decoded message s and the original message s^0 . An overlap of 1 indicates perfect decoding while an overlap of 0 denotes complete failure.

- [1] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).
 [2] N. Surlas, Nature (London) **339**, 693 (1989).
 [3] Y. Kabashima, T. Murayama, and D. Saad, Phys. Rev. Lett. **84**, 1355 (2000).

- [4] H. Nishimori and K. Y. Michael Wong, Phys. Rev. E **60**, 132 (1999).
 [5] A. Montanari and N. Surlas, Eur. Phys. J. B **18**, 107 (2000).
 [6] T. Tanaka, Europhys. Lett. **54**, 540 (2001).

- [7] T. Tanaka and M. Okada, *IEEE Trans. Inf. Theory* **51**, 700 (2005).
- [8] T. Murayama and M. Okada, *J. Phys. A* **36**, 11123 (2003).
- [9] T. Murayama, *Phys. Rev. E* **69**, 035105(R) (2004).
- [10] T. Hosaka, Y. Kabashima, and H. Nishimori, *Phys. Rev. E* **66**, 066126 (2002).
- [11] T. Hosaka and Y. Kabashima, *J. Phys. Soc. Jpn.* **74**, 488 (2005).
- [12] K. Mimura and M. Okada, *Phys. Rev. E* **74**, 026108 (2006).
- [13] F. Cousseau, K. Mimura, T. Omori, and M. Okada, *Phys. Rev. E* **78**, 021124 (2008).
- [14] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, Cambridge, England, 2003).
- [15] R. Monasson and D. O’Kane, *Europhys. Lett.* **27**, 85 (1994).
- [16] G. J. Bex, R. Serneels, and C. Van den Broeck, *Phys. Rev. E* **51**, 6309 (1995).
- [17] T. Shinzato and Y. Kabashima, *Proceeding in Workshop on Information-Based Induction Sciences (IBIS2006)*, Osaka, Japan (2006) (in Japanese).
- [18] J. R. L. de Almeida and D. J. Thouless, *J. Phys. A* **11**, 983 (1978).
- [19] C. Berrou, A. Glavieux, and P. Thitimajshima, *Proceedings of IEEE International Conference on Communications (ICC)*, 1993 (unpublished), 1064–1070.
- [20] R. Gallager, *IRE Trans. Inf. Theory* **8**, 21 (1962).
- [21] D. MacKay and R. Neal, *Lect. Notes Comput. Sci.* **1025**, 100 (1995).
- [22] C.-C. Wang, S. R. Kulkarni, and H. V. Poor, *IEEE Trans. Inf. Theory* **51**, 4216 (2005).
- [23] I. Neri, N. S. Skantzos, and D. Bolla, *J. Stat. Mech.: Theory Exp.* (2008), 10018.
- [24] A. Engel and C. Van den Broeck, *Statistical Mechanics of Learning* (Cambridge University Press, Cambridge, England, 2001).
- [25] P. Ruján, *Phys. Rev. Lett.* **70**, 2968 (1993).
- [26] H. Nishimori, *J. Phys. Soc. Jpn.* **62**, 2973 (1993).
- [27] N. Sourlas, *Europhys. Lett.* **25**, 159 (1994).
- [28] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing* (Oxford University Press, New York, 2001).
- [29] H. Nishimori and D. Sherrington, in *Disordered and Complex System*, edited by P. Sollich, A. Coolen, L. Hughston, and R. Streater (AIP, New York, 2001), pp. 67–72.
- [30] T. Hosaka and Y. Kabashima, *Physica A* **365**, 113 (2006).