Statistical mechanics of the data compression theorem

2002 J. Phys. A: Math. Gen. 35 L95

(http://iopscience.iop.org/0305-4470/35/8/101)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **35** (2002) L95–L100 PII: S0305-4470(02)29410-X

LETTER TO THE EDITOR

Statistical mechanics of the data compression theorem

Tatsuto Murayama

Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Yokohama 2268502, Japan

Received 2 October 2001
Published 15 February 2002
Online at stacks.iop.org/JPhysA/35/L95

Abstract

We analyse the performance of a linear code used for data compression of a Slepian–Wolf type. In our framework, two correlated data are separately compressed into codewords employing Gallager-type codes and cast into a communication network through two independent input terminals. At the output terminal, the received codewords are jointly decoded by a practical algorithm based on the Thouless–Anderson–Palmer approach. Our analysis shows that the achievable rate region presented in the data compression theorem is described as first-order phase transitions among several phases. The typical performance of the practical decoder is also well evaluated by the replica method.

PACS numbers: 05.20.-y, 05.70.Ph, 64.60.Cn, 89.70.+c

Data compression, or source coding, is a scheme to reduce the size of data in information representation. In his seminal paper [9], Shannon showed that for an information source represented by a distribution $\mathcal{P}(\boldsymbol{\xi})$ of an N-dimensional Boolean (binary) vector $\boldsymbol{\xi}$, one can employ another representation in which the message length N is reduced to M ($\leqslant N$) without any distortion, if the code rate $R = M/N$ satisfies $R \geqslant H_2(\boldsymbol{\xi})$ in the limit $N, M \to \infty$. Here, $H_2(\boldsymbol{\xi}) = -(1/N)\mathrm{Tr}_{\boldsymbol{\xi}}\mathcal{P}(\boldsymbol{\xi}) \log_2 \mathcal{P}(\boldsymbol{\xi})$ represents the binary entropy per bit in the original representation $\boldsymbol{\xi}$ indicating the optimal compression rate. Unfortunately, Shannon's theorem itself is non-constructive and does not provide explicit rules for devising the optimal codes. Therefore, it is surprising that a practical code proposed by Lempel and Ziv (LZ) in 1973 [14] saturates Shannon's optimal compression limit in the case of a single-user interface, when lossless compression scenarios are considered. However, it should be emphasized here that generalization of the LZ codes to advanced data compression suitable for multi-user interface is difficult, although the importance of the network is rapidly increasing.

The purpose of this letter is to employ recent developments of the research on error-correcting codes (ECC), to construct a simple model of a data compression scheme and to present a physical picture of it. More specifically, we will investigate the efficacy and the limitation of a linear compression scheme inspired by Gallager's codes [2], which has been actively

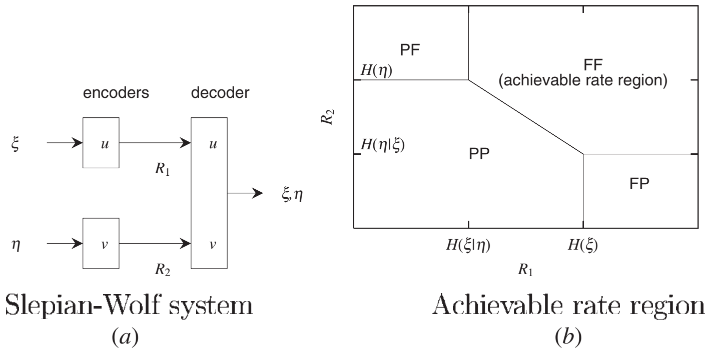

Figure 1. (*a*) SW system: a network introduced in the data compression theorem. Separate coding is assumed in the distributed system. (*b*) Achievable rate region: code rates are classified into four categories according to whether the two compressed data are decodable or not. The parameter regime where the both data are decodable without any distortion is termed the *achievable rate region*.

investigated in both the information theory and physics communities [4–6,8], when it is applied to the data compression problem introduced by Slepian and Wolf (SW) in their research on multi-terminal information theory [1, 10]. Unlike the existing argument in information theory, our approach based on statistical mechanics makes it possible not only to assess the theoretical bounds of the achievable performance but also to provide practical encoding/decoding methods that can be performed in linear timescales with respect to the data length.

Let us start by setting up the framework of the SW problem [10]. In a general scenario, two correlated N-dimensional Boolean vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are *independently* compressed to M-dimensional vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, respectively. These compressed data (or codewords) $\boldsymbol{u}$ and $\boldsymbol{v}$ are decoded to retrieve the original data simultaneously by a single decoder. A schematic representation of this system is shown in figure 1(*a*).

The codes are composed of randomly selected sparse matrices A and B of dimensionality $M_1 \times N$ and $M_2 \times N$, respectively. These are constructed similarly to those of Gallager's ECC [4] as characterized by K_1 and K_2 nonzero unit elements per row and C_1 and C_2 nonzero unit elements per column, respectively. The compression rates can be different between the two terminals. Corresponding to matrices A and B, the rates are defined as $R_1 = M_1/N = K_1/C_1$ and $R_2 = M_2/N = K_2/C_2$, respectively. While both matrices are known to the decoder, encoders only need to know their own matrix; that is, encoding is carried out separately in this scheme as $\boldsymbol{u} = A\boldsymbol{\xi}$ and $\boldsymbol{v} = B\boldsymbol{\eta}$, where Boolean arithmetic is employed. After receiving the codewords $\boldsymbol{u}$ and $\boldsymbol{v}$, the pair of equations $\boldsymbol{u} = A\boldsymbol{S}$, $\boldsymbol{v} = B\boldsymbol{\tau}$ should be solved with respect to $\boldsymbol{S}$ and $\boldsymbol{\tau}$ which become the estimates of the original data $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, respectively.

To facilitate the current investigation we first map the problem to that of an Ising model with finite connectivity [11]. We employ the binary representation (+1, −1) of the dynamical variables $\boldsymbol{S}$ and $\boldsymbol{\tau}$ and of the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ rather than the Boolean (0, 1) one; the vector $\boldsymbol{u}$ is generated by taking products of the relevant binary data bits $u_{\langle i_1, i_2, \ldots, i_{K_1} \rangle} = \xi_{i_1} \xi_{i_2} \cdots \xi_{i_{K_1}}$, where the indices $i_1, i_2, \ldots, i_{K_1}$ correspond to the nonzero elements of A, producing a binary version of $\boldsymbol{u}$, and similarly for $\boldsymbol{v}$. Assuming the thermodynamic limit $N, M_1, M_2 \to \infty$, while keeping the code rates $R_1 = M_1/N$ and $R_2 = M_2/N$ finite is quite natural as communication to date generally requires transmitting large data, where finite size corrections are likely to be negligible. To explore the system's capabilities we examine the partition function

$$\mathcal{Z} = \operatorname*{Tr}_{S,\tau} \mathcal{P}(S,\tau) \prod_{\langle i_1,i_2,\ldots,i_{K_1}\rangle} [1 + \tfrac{1}{2}\mathcal{A}_{\langle i_1,i_2,\ldots,i_{K_1}\rangle}(u_{\langle i_1,i_2,\ldots,i_{K_1}\rangle} \cdot S_{i_1}S_{i_2}\cdots S_{i_{K_1}} - 1)]$$
$$\times \prod_{\langle i_1,i_2,\ldots,i_{K_2}\rangle} [1 + \tfrac{1}{2}\mathcal{B}_{\langle i_1,i_2,\ldots,i_{K_2}\rangle}(v_{\langle i_1,i_2,\ldots,i_{K_2}\rangle} \cdot \tau_{i_1}\tau_{i_2}\cdots\tau_{i_{K_2}} - 1)]. \tag{1}$$

The tensor product $\mathcal{A}_{\langle i_1,i_2,\ldots,i_{K_1}\rangle}u_{\langle i_1,i_2,\ldots,i_{K_1}\rangle}$, where $u_{\langle i_1,i_2,\ldots,i_{K_1}\rangle} = \xi_{i_1}\xi_{i_2}\cdots\xi_{i_{K_1}}$, is the binary equivalent of $A\boldsymbol{\xi}$. Elements of the sparse connectivity tensor $\mathcal{A}_{\langle i_1,i_2,\ldots,i_{K_1}\rangle}$ take the value 1 if the corresponding indices of data are chosen (i.e. if all corresponding indices of the matrix A are 1) and 0 otherwise; it has C_1 unit elements per i index representing the system's degree of connectivity. Note that if the product $S_{i_1}S_{i_2}\cdots S_{i_{K_1}}$ is in disagreement with the corresponding element $u_{\langle i_1,i_2,\ldots,i_{K_1}\rangle}$, which implies an error for the parity check, the value of the partition function $\mathcal{Z}$ vanishes. Similar arguments are valid for $\mathcal{B}_{\langle i_1,i_2,\ldots,i_{K_2}\rangle}$ and $v_{\langle i_1,i_2,\ldots,i_{K_2}\rangle}$. The probability $\mathcal{P}(S,\tau)$ represents our prior knowledge of data including the correlation between the sources $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$. Note that the dynamical variables $\boldsymbol{\tau}$, introduced to estimate $\boldsymbol{\eta}$, are irrelevant to the performance measure with respect to the other data $\boldsymbol{\xi}$.

Since the partition function (1) is invariant under the transformations $S_i \to S_i\xi_i$, $\tau_i \to \tau_i\eta_i$, $u_{\langle i_1,i_2,\ldots,i_{K_1}\rangle} \to u_{\langle i_1,i_2,\ldots,i_{K_1}\rangle}\xi_{i_1}\xi_{i_2}\cdots\xi_{i_{K_1}} = 1$ and $v_{\langle i_1,i_2,\ldots,i_{K_2}\rangle} \to v_{\langle i_1,i_2,\ldots,i_{K_2}\rangle}\tau_{i_1}\tau_{i_2}\cdots\tau_{i_{K_2}} = 1$, it is useful to decouple the correlations between the vectors S, τ and $\boldsymbol{\xi}, \boldsymbol{\eta}$. Rewriting equation (1) using this gauge, one obtains a similar expression apart from the first factor which becomes $\mathcal{P}(S\otimes\boldsymbol{\xi}, \tau\otimes\boldsymbol{\eta})$, where $S\otimes\boldsymbol{\xi} = (S_i\xi_i)$ and $\tau\otimes\boldsymbol{\eta} = (\tau_i\eta_i)$ for $i = 1, 2, \ldots, N$.

The random selection of elements in $\mathcal{A}$ and $\mathcal{B}$ introduces disorder to the system; we average the logarithm of the partition function $\mathcal{Z}(\mathcal{A},\mathcal{B},\boldsymbol{u},\boldsymbol{v})$ over the disorder and the statistical properties of both data, using the replica method [13]. In the calculation, a set of order parameters $q_{a_1,a_2,\ldots,a_l} = \frac{1}{N}\sum_{i=1}^N Z_i S_i^{a_1}S_i^{a_2}\cdots S_i^{a_l}$ and $r_{a_1,a_2,\ldots,a_l} = \frac{1}{N}\sum_{i=1}^N Y_i \tau_i^{a_1}\tau_i^{a_2}\cdots\tau_i^{a_l}$ arise, where $a_1, a_2, \ldots, a_l$ $(l = 1, 2, \ldots)$ represent replica indices, and the variables Z_i and Y_i come from enforcing the restriction of C_1 and C_2 connections per index, respectively, as in [6].

Assuming a replica symmetric ansatz, that is, $q_{a_1,a_2,\ldots,a_l} = \int \mathrm{d}x\,\pi(x)x^l$ and $r_{a_1,a_2,\ldots,a_l} = \int \mathrm{d}y\,\rho(y)y^l$ [6], we obtain the following free energy per spin:

$$\begin{aligned}
\mathcal{F} &= -\frac{1}{N}\langle \ln \mathcal{Z}\rangle_{\mathcal{A},\mathcal{B},\mathcal{P}} \\
&= -\operatorname*{Extr}_{\pi,\hat{\pi},\rho,\hat{\rho}} \Bigg\{ \frac{C_1}{K_1}\left\langle \ln\left(\frac{1+\prod_{i=1}^{K_1}x_i}{2}\right)\right\rangle_{\pi} + \frac{C_2}{K_2}\left\langle \ln\left(\frac{1+\prod_{i=1}^{K_2}y_i}{2}\right)\right\rangle_{\rho} \\
&\quad - C_1\left\langle \ln\left(\frac{1+x\hat{x}}{2}\right)\right\rangle_{\pi,\hat{\pi}} - C_2\left\langle \ln\left(\frac{1+y\hat{y}}{2}\right)\right\rangle_{\rho,\hat{\rho}} \\
&\quad + \frac{1}{N}\Bigg\langle \ln\Bigg[\operatorname*{Tr}_{S,\tau}\prod_{i=1}^{N}\prod_{\mu=1}^{C_1}\left(\frac{1+\hat{x}_{\mu i}S_i}{2}\right) \\
&\quad \times \prod_{i=1}^{N}\prod_{\mu=1}^{C_2}\left(\frac{1+\hat{y}_{\mu i}\tau_i}{2}\right)\mathcal{P}(S\otimes\boldsymbol{\xi},\tau\otimes\boldsymbol{\eta})\Bigg]\Bigg\rangle_{\hat{\pi},\hat{\rho},\mathcal{P}}\Bigg\}
\end{aligned} \tag{2}$$

where the brackets with the subscripts π and $\hat{\pi}$ represent averages over the distribution $\pi(x)$ and its conjugate $\pi(\hat{x})$ with respect to variables denoted by $x \in [-1, 1]$ and $\hat{x} \in [-1, 1]$ with and without subscripts, respectively. Similar notations are also used for ρ and $\hat{\rho}$. The bracket with the subscript $\mathcal{P}$ denotes the average with respect to $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ following the data distribution $\mathcal{P}(\boldsymbol{\xi},\boldsymbol{\eta})$.

Taking the functional derivative with respect to the distributions π, $\hat{\pi}$, ρ and $\hat{\rho}$, we obtain the saddle point equations (SPE) with the effective fields $F_i(\cdots)$ such that

$$\frac{\exp\left(F_i(\hat{x}_{\mu j\in\mathcal{L}(\mu)\backslash i},\ \hat{y}_{\mu i};\ \boldsymbol{\xi},\boldsymbol{\eta})\xi_i S_i\right)}{2\cosh F_i(\hat{x}_{\mu j\in\mathcal{L}(\mu)\backslash i},\hat{y}_{\mu i};\boldsymbol{\xi},\boldsymbol{\eta})}$$
$$= \frac{\underset{\boldsymbol{S}\backslash S_i,\boldsymbol{\tau}}{\mathrm{Tr}} \prod_{j\in\mathcal{L}(\mu)\backslash i}\prod_{\mu=1}^{C_1}\left(\frac{1+\hat{x}_{\mu j}S_j}{2}\right)\prod_{i=1}^{N}\prod_{\mu=1}^{C_2}\left(\frac{1+\hat{y}_{\mu i}\tau_i}{2}\right)\mathcal{P}(\boldsymbol{S}\otimes\boldsymbol{\xi},\boldsymbol{\tau}\otimes\boldsymbol{\eta})}{\underset{\boldsymbol{S},\boldsymbol{\tau}}{\mathrm{Tr}}\prod_{i}^{N}\prod_{\mu=1}^{C_1}\left(\frac{1+\hat{x}_{\mu i}S_i}{2}\right)\prod_{i=1}^{N}\prod_{\mu=1}^{C_2}\left(\frac{1+\hat{y}_{\mu i}\tau_i}{2}\right)\mathcal{P}(\boldsymbol{S}\otimes\boldsymbol{\xi},\boldsymbol{\tau}\otimes\boldsymbol{\eta})} \tag{3}$$

and similarly for the other field. Notice that the notation $\boldsymbol{S}\backslash S_i$ represents the set of all dynamical variables $\boldsymbol{S}$ except S_i. On the other hand, $\mathcal{L}_1(\mu)$ and $\mathcal{L}_2(\mu)$ denote the set of all indices of nonzero components in the μth row of A and B, respectively. The notation $\mathcal{L}_1(\mu)\backslash i$ represents the set of all indices belonging to $\mathcal{L}_1(\mu)$ except i.

After solving these equations, the expectation of the overlap $\mathsf{m}_1 = \frac{1}{N}\langle\sum_{i=1}^{N}\xi_i \,\mathrm{sign}\,\langle S_i\rangle\rangle_{\mathcal{A},\mathcal{P}}$ can be theoretically evaluated, and similarly for m_2 of the overlap between $\boldsymbol{\eta}$ and its estimator. The performance of the current compression method can be measured by the vector $\mathbf{m} = (\mathsf{m}_1, \mathsf{m}_2)$. Hereafter, we use the term 'ferromagnetic' to specify the perfect retrieval, that is, $\mathsf{m}_1 = 1$ (or $\mathsf{m}_2 = 1$), while the term 'paramagnetic' implies the distortion, that is, $\mathsf{m}_1 < 1$ (or $\mathsf{m}_2 < 1$). For instance, a term such as 'ferromagnetic–paramagnetic (FP) phase' denotes the phase characterized by the performance vector $\mathbf{m} \in \{(\mathsf{m}_1,\mathsf{m}_2)|\mathsf{m}_1 = 1, \mathsf{m}_2 < 1\}$, and so on.

One can show that the ferromagnetic–ferromagnetic state (FF), described by the solutions $\pi(x) = \delta(x-1), \hat{\pi}(\hat{x}) = \delta(\hat{x}-1), \rho(y) = \delta(y-1)$ and $\hat{\rho}(\hat{y}) = \delta(\hat{y}-1)$, always satisfies the SPE. In addition, in the limit of $C_1, C_2 \to \infty$, four solutions describing the paramagnetic–paramagnetic state (PP), $\pi(x) = \delta(x), \hat{\pi}(\hat{x}) = \delta(\hat{x}), \rho(y) = \delta(y)$ and $\hat{\rho}(\hat{y}) = \delta(\hat{y})$, the paramagnetic–ferromagnetic (PF) phase, $\pi(x) = \delta(x), \hat{\pi}(\hat{x}) = \delta(\hat{x}), \rho(y) = \delta(y-1)$ and $\hat{\rho}(\hat{y}) = \delta(\hat{y}-1)$ and the FP state, $\pi(x) = \delta(x-1), \hat{\pi}(\hat{x}) = \delta(\hat{x}-1), \rho(y) = \delta(y)$ and $\hat{\rho}(\hat{y}) = \delta(\hat{y})$, are also analytically obtained for an *arbitrary* joint distribution $\mathcal{P}(\boldsymbol{\xi},\boldsymbol{\eta})$. Free energies corresponding to these solutions are provided from equation (2) as $\mathcal{F}_{\mathrm{FF}} = -\frac{1}{N}\mathrm{Tr}_{\boldsymbol{\xi},\boldsymbol{\eta}}\mathcal{P}(\boldsymbol{\xi},\boldsymbol{\eta})\ln\mathcal{P}(\boldsymbol{\xi},\boldsymbol{\eta})$, $\mathcal{F}_{\mathrm{PP}} = (R_1+R_2)\ln 2$, $\mathcal{F}_{\mathrm{FP}} = R_2\ln 2 - \frac{1}{N}\mathrm{Tr}_{\boldsymbol{\xi}}\mathcal{P}(\boldsymbol{\xi})\ln\mathcal{P}(\boldsymbol{\xi})$, $\mathcal{F}_{\mathrm{PF}} = R_1\ln 2 - \frac{1}{N}\mathrm{Tr}_{\boldsymbol{\eta}}\mathcal{P}(\boldsymbol{\eta})\ln\mathcal{P}(\boldsymbol{\eta})$, where subscripts stand for corresponding states and $\mathcal{P}(\boldsymbol{\xi}) = \mathrm{Tr}_{\boldsymbol{\eta}}\mathcal{P}(\boldsymbol{\xi},\boldsymbol{\eta})$ and $\mathcal{P}(\boldsymbol{\eta}) = \mathrm{Tr}_{\boldsymbol{\xi}}\mathcal{P}(\boldsymbol{\xi},\boldsymbol{\eta})$ represent marginal distributions for the vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, respectively.

Perfect decoding is theoretically possible if $\mathcal{F}_{\mathrm{FF}}$ is the lowest among the above four. The corresponding parameter regime termed *achievable rate region* is shown in figure 1(*b*) as an intersection of the inequalities

$$R_1 + R_2 \geqslant H_2(\boldsymbol{\xi},\boldsymbol{\eta}) \qquad R_1 \geqslant H_2(\boldsymbol{\xi}|\boldsymbol{\eta}) \qquad R_2 \geqslant H_2(\boldsymbol{\eta}|\boldsymbol{\xi}) \tag{4}$$

where $H_2(\boldsymbol{\xi},\boldsymbol{\eta}) = -\frac{1}{N}\mathrm{Tr}_{\boldsymbol{\xi},\boldsymbol{\eta}}\mathcal{P}(\boldsymbol{\xi},\boldsymbol{\eta})\ln\mathcal{P}(\boldsymbol{\xi},\boldsymbol{\eta})$, $H_2(\boldsymbol{\xi}|\boldsymbol{\eta}) = H_2(\boldsymbol{\xi},\boldsymbol{\eta}) - H_2(\boldsymbol{\eta})$ and $H_2(\boldsymbol{\eta}|\boldsymbol{\xi}) = H_2(\boldsymbol{\xi},\boldsymbol{\eta}) - H_2(\boldsymbol{\xi})$. It is worth noticing that this coincides with the achievable rate region saturated by the *optimal* data compression in the current framework previously shown by SW [10]. Namely, in the limit $C_1, C_2 \to \infty$, the current compression codes provide the optimal performance for *arbitrary* information sources.

For finite C_1 and C_2, the SPE can be solved numerically, but the properties of the system highly depend on the source distribution $\mathcal{P}(\boldsymbol{\xi},\boldsymbol{\eta})$, which makes it difficult to go further without any assumption on the distribution. As a simple but non-trivial example, we will focus here on a component-wise correlated joint distribution $\mathcal{P}(\boldsymbol{S},\boldsymbol{\tau}) = \prod_{i=1}^{N}(1 + m_1 S_i + m_2\tau_i + qS_i\tau_i)/4$, where a set of parameters m_1, m_2 and q characterize the data sources. Notice that q represents the overlap between the data. To make it a distribution, these parameters must satisfy four

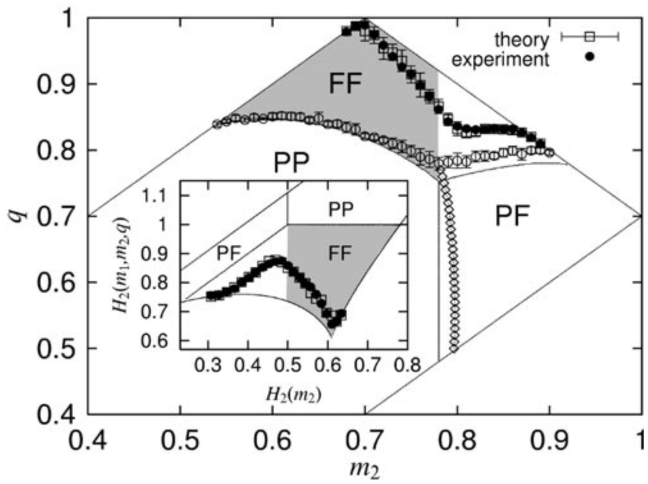

Figure 2. Phase diagram for $K_1 = K_2 = 6$, $C_1 = C_2 = 3$ code in the case of component-wise correlated source. Figure shows that the feasible region in the m_2–q plane for $m_1 = 0.7$ is classified into three states. Phase boundaries obtained by numerical methods are indicated by ○ with errorbars (FF/PP and FF/PF) and ◇ (PF/PP). These are close to those for $K_1 = K_2 \to \infty$, $C_2 = C_2 \to \infty$ (curves and the vertical line). Practically decodable limits of the TAP/BP algorithm obtained for $N = 10^4$ systems are indicated as ●. These are well evaluated by the spinodal points of non-FF solutions (□ with errorbars). Inset: the practical limits are represented by the sizes of transmitted information. Horizontal and vertical axes show the entropy of the second source τ and the joint entropy, respectively. The shaded region cannot be achieved without the simultaneous decoding.

inequalities: $1 + m_1 + m_2 + q \geqslant 0$, $1 - m_1 + m_2 - q \geqslant 0$, $1 + m_1 - m_2 - q \geqslant 0$ and $1 - m_1 - m_2 + q \geqslant 0$.

Solving equations rigorously for decoding is computationally hard in general cases. However, one can construct a practical decoding algorithm based on the belief propagation (BP) [7] or the Thouless–Anderson–Palmer (TAP) approach [12]. It has recently been shown that these two frameworks provide the same algorithm in the case of ECC [3]. This is also the case under the current context. For this distribution, the algorithm derived from the BP/TAP frameworks becomes

$$\begin{aligned} m^1_{\mu i} &= \frac{a_{\mu i} + m_1 + m_2 a_{\mu i} b_i + q b_i}{1 + m_1 a_{\mu i} + m_2 b_i + q a_{\mu i} b_i} & m^2_{\mu i} &= \frac{b_{\mu i} + m_2 + m_1 a_i b_{\mu i} + q a_i}{1 + m_1 a_i + m_2 b_{\mu i} + q a_i b_{\mu i}} \\ \hat{m}^1_{\mu i} &= u_\mu \prod_{j \in \mathcal{L}_1(\mu)\backslash i} m^1_{\mu j} & \hat{m}^2_{\mu i} &= v_\mu \prod_{j \in \mathcal{L}_2(\mu)\backslash i} m^2_{\mu j} \end{aligned} \tag{5}$$

where we denote $a_{\mu i} \equiv \tanh \sum_{\nu \in \mathcal{M}_1(i)\backslash \mu} \tanh^{-1} \hat{m}^1_{\nu i}$ and $a_i \equiv \tanh \sum_{\mu \in \mathcal{M}_1(i)} \tanh^{-1} \hat{m}^1_{\mu i}$, and similarly for b. Here, $\mathcal{M}_1(i)$ and $\mathcal{M}_2(i)$ indicate the set of all indices of nonzero components in the ith column of the sparse matrices A and B, respectively. Equation (5) can be solved iteratively from the appropriate initial conditions. After obtaining a solution, approximated posterior means can be calculated for $i = 1, 2, \ldots, N$ as

$$m^1_i = \frac{a_i + m_1 + m_2 a_i b_i + q b_i}{1 + m_1 a_i + m_2 b_i + q a_i b_i} \qquad m^2_i = \frac{b_i + m_2 + m_1 a_i b_i + q a_i}{1 + m_1 a_i + m_2 b_i + q a_i b_i} \tag{6}$$

which provide an approximation to the Bayes-optimal estimators as $\xi_i = \mathrm{sign}(m^1_i)$ and $\eta_i = \mathrm{sign}(m^2_i)$, respectively.

In order to investigate the efficacy of the current method for finite C_1 and C_2, we have numerically solved the SPE and (5) for $K_1 = K_2 = 6$ and $C_1 = C_2 = 3$ ($R_1 = R_2 = 1/2$), results of which are summarized in figure 2. Numerical results for the SPE were obtained by an iterative method using 10^4–10^5 bin models for each probability distribution. 10–10^2 updates were sufficient for convergence in most cases. Similarly to the case of $C_1, C_2 \to \infty$, there can be four types of solutions corresponding to combinations of decoding success and failure on the two sources. The obtained phase diagram is quite similar to that for $C_1, C_2 \to \infty$. This implies that the current compression code *theoretically* has a good performance close to the optimal one that is saturated in the limit $C_1, C_2 \to \infty$, although the choice of $C_1 = C_2 = 3$ is far from such a limit.

However, this does not directly mean that the suggested performance can be obtained *in practice*. Since the variables are updated locally in the BP/TAP decoding algorithm (5), it may become difficult to find the thermodynamically dominant state when there appear suboptimal states which have large basins of attraction. This suggests that the practical performance for the perfect decoding is determined by the spinodal points of the suboptimal states, similar to the case of ECC [6]. To confirm this conjecture, we have numerically compared the practical limit of the perfect decoding obtained by the BP/TAP decoding algorithm (5) and the spinodal points of the non-FF solutions. These two results exhibit an excellent consistency supporting our conjecture.

In summary, we have developed an efficient method of data compression in a multi-terminal scenario, taking advantage of the sparse matrix based linear compression codes. We observed several practical properties of codes of this type in the simplest model of data compression. Studying the typical performance of the linear compression codes in a network, which complements the methods used in the information theory literature, is the first step towards understanding typical properties of networks.

The author thanks Y Kabashima and T Ohira for valuable discussions. This work was partly supported by the Japan Society for the Promotion of Science.

References

[1] Cover T M and Thomas J A 1991 *Elements of Information Theory* (New York: Wiley)
[2] Gallager R G 1962 *IRE Trans. Inf. Theory* **8** 21
[3] Kabashima Y and Saad D 1998 *Europhys. Lett.* **44** 668
[4] MacKay D J C 1999 *IEEE Trans. Inf. Theory* **45** 399
[5] Montanari A 2001 *Preprint* cond-mat/0104079
[6] Murayama T, Kabashima Y, Saad D and Vicente R 2000 *Phys. Rev.* E **62** 1577
[7] Pearl J 1988 *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference* (San Francisco, CA: Morgan Kaufmann)
[8] Richardson T, Shokrollahi A and Urbanke R 1999 Design of provably good low density parity check codes *Preprint*
[9] Shannon C E 1948 *Bell Syst. Tech. J.* **27** 379
[10] Slepian D and Wolf J K 1973 *IEEE Trans. Inf. Theory* **19** 471
[11] Sourlas N 1989 *Nature* **339** 693
[12] Thouless D J, Anderson P W and Palmer R G 1977 *Phil. Mag.* **35** 593
[13] Wong K Y M and Sherrington D 1987 *J. Phys. A: Math. Gen.* **20** L793
[14] Ziv J and Lempel A 1977 *IEEE Trans. Inf. Theory* **23** 337