An Algebraic Construction of Codes for Slepian-Wolf Source Networks*

Tomohiko UYEMATSU, Member IEEE

May 21, 2001

*This work was supported in part by Japan Society for the Promotion of Science (JSPS) under Grant JSPS-RFTF97P00601. The material in this paper was presented in part at IEEE International Symposium on Information Theory, Ulm Germany 1997 and in part at IEEE International Symposium on Information Theory, MIT, Cambridge, MA 1998.

Tomohiko Uyematsu is with the Department of Communications and Integrated Systems, Tokyo Institute of Technology, Ookamaya 2-12-1, Meguro-ku, Tokyo 152-8552, Japan.

E-mail: uematsu@it.ss.titech.ac.jp

Abstract

This paper proposes an explicit construction of fixed length codes for Slepian-Wolf source networks. The proposed code is linear, and has two-step encoding and decoding procedures similar to the concatenated code used for channel coding. Encoding and decoding of the code can be done in a polynomial order of the block length. The proposed code can achieve arbitrary small probability of error for ergodic sources with finite alphabets, if the pair of encoding rates is in the achievable region. Further, if the sources are memoryless, the proposed code can be modified to become universal and the probability of error vanishes exponentially as the block length tends to infinity.

key words: ergodic process, fixed length code, Slepian-Wolf coding, source coding.

I. Introduction

The separate coding problem for correlated sources has been first investigated by Slepian and Wolf [1], and the coding problem for Slepian-Wolf (SW) source network is the most fundamental source coding problem in the areas of multi-terminal information theory. After the original proof of coding theorem by Slepian and Wolf, Cover [2] extended the proof of the coding theorem for ergodic sources by using bin coding, while Ahlswede and Körner [3] gave an alternative proof by using constructive algorithm. However, these proofs do not suggest any explicit construction of codes, since they require some searches for code constructions. Further, the complexity of encoding/decoding for these codes increases exponentially as the block length tends to infinity. Hence, the problem of explicitly constructing codes for SW network with low cost of encoding and decoding still remains open. The construction of such codes is not only interesting in its own right but also very important from the standpoint of practical communications, since the coding problem for SW network is strongly related to many kinds of problems in multi-terminal information theory. For example, a multiple access code can be derived from a code constructed for SW network [4, Proof of Theorem 3.2.3]. Further, a code for SW network can also be applied to the common randomness problem, i.e. to generate a common random key for two users by using correlated sources and communication over public channel without letting an eavesdropper obtain information about the generated key [5, Proof of Proposition 1].

On the other hand, in the areas of channel coding, Delsarte and Piret [6] succeeded to explicitly construct a channel code whose decoding error probability decreases exponentially with the code length for a class of regular channels. They apply Justesen's idea of "variable concatenation" [7] to provide an algebraic construction of encoders. In this paper, we apply the idea of Delsarte and Piret to construct explicit codes for the SW source network. Our proposed code has two-step encoding and decoding procedures similar to the concatenated code for channel coding [8]. In the first step of encoding, the block of source symbols is divided into subblocks, and each subblock is encoded by a distinct linear code. Here, a set of linear codes has the property that the average probability of error over the set vanishes as the block length tends to infinity. In the second step of encoding, the total block is again encoded into a syndrome of an algebraic geometry code which is a class of linear error correcting codes. In the decoding, linear codes are decoded first by using typical set decoding [2]. Then, by using linearity of the algebraic geometry code, we can obtain the syndrome which corresponds to the errors caused by the codes in the first step. These errors can be easily corrected by using the procedure for error correction of the algebraic geometry code. The proposed code can achieve arbitrary small probability of error for ergodic sources, if the pair of encoding rates is in the achievable region. Lastly, we deal with the case where the sources are memoryless. In such a case, if we employ the minimum entropy decoding instead of the typical set decoding, the proposed code becomes universal and its decoding as well as encoding do not depend on the generic distribution of the sources. Further,

we clarify that the probability of error vanishes exponentially as the block length tends to infinity.

It should be noted that Csiszár [9] showed that an inner point of the achievable region can be achieved by a linear code. However, since he employed random coding technique to construct linear codes, his construction requires the computational complexity of exponential order of the block length in order to obtain a good linear code. Further, the decoding complexity of his code is of exponential order of the block length. On the contrary, the construction of our code as well as its decoding can be done within the polynomial order of the block length. Hence, our code is much more efficient than that proposed by Csiszár.

II. Preliminaries

a. Slepian-Wolf coding problem

Following Slepian and Wolf [1], we consider the problem of separate encoding and joint decoding of two ergodic sources (X, Y) with finites alphabets $\mathcal{X}$ and $\mathcal{Y}$. Suppose that the joint distribution of the sources is described by Q.

Definition 1 (Codes for Slepian-Wolf networks): A code of block length n for the Slepian-Wolf (SW) networks is defined by a triple of mappings (f, g, φ) where the encoder f maps $\mathcal{X}^n$ into a set of codewords $\mathcal{M}_1$ and the encoder g maps $\mathcal{Y}^n$ into another set of codewords $\mathcal{M}_2$, while the decoder φ maps $\mathcal{M}_1 \times \mathcal{M}_2$ into $\mathcal{X}^n \times \mathcal{Y}^n$. Further, let $R_1 \stackrel{\triangle}{=} \log|\mathcal{M}_1|/n$ and $R_2 \stackrel{\triangle}{=} \log|\mathcal{M}_2|/n$ where $|\cdot|$ denotes the cardinality

of the set. Then (R_1, R_2) is called the *rate pair* of the code.

Definition 2 (Probability of Error): The *probability of error* for a code (f, g, φ) of block length n for a SW network is defined as

$$\begin{aligned} p_e &= p_e(f, g, \varphi) \\ &\triangleq Q^n(\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \varphi(f(\boldsymbol{x}), g(\boldsymbol{y})) \neq (\boldsymbol{x}, \boldsymbol{y})\}). \end{aligned} \tag{1}$$

Definition 3 (Achievable Rate Region): A rate pair (R_1, R_2) is said to be *achievable* for a SW network, if there exists a sequence of codes $(f^{(n)}, g^{(n)}, \varphi^{(n)})$ with increasing block length n such that the rate pair of the code is (R_1, R_2) and the probability of error $p_e^{(n)} \to 0$ as $n \to \infty$. The *achievable rate region* is the closure of the set of achievable rates.

Theorem 1 [2]: For a SW network with the jointly ergodic source (X, Y), the achievable rate region is given by

$$\begin{aligned} R_1 &\geq H(X|Y), \\ R_2 &\geq H(Y|X), \\ R_1 + R_2 &\geq H(X, Y), \end{aligned}$$

where $H(Y|X)$ and $H(X|Y)$ denote the conditional entropy rates [2] and $H(X, Y)$ denotes the entropy rate [2] for the joint ergodic source (X, Y).

b. Typical sequences and typical set decoder

Let (X, Y) be a jointly ergodic source of discrete random variables with a fixed joint distribution Q. Then, we define typical sequences as follows:

Definition 4 (Typical Sequences): The set A_ϵ^n of *ϵ-typical n-sequences* $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ is defined by

$$\begin{aligned} A_\epsilon^n \triangleq \Bigg\{ & (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \\ & \left| -\frac{1}{n} \log Q^n(\boldsymbol{x}, \boldsymbol{y}) - H(X,Y) \right| \leq \epsilon, \\ & \left| -\frac{1}{n} \log Q^n(\boldsymbol{x}) - H(X) \right| \leq \epsilon, \\ & \left| -\frac{1}{n} \log Q^n(\boldsymbol{y}) - H(Y) \right| \leq \epsilon \Bigg\}, \end{aligned} \tag{2}$$

where $H(\cdot)$ denotes the entropy rate.

By the asymptotic equipartition property, we have the following Lemma (e.g. see [2]).

Lemma 1: For any $\epsilon > 0$ and sufficiently large n, the set A_ϵ^n satisfies

$$Q^n(A_\epsilon^n) \triangleq Q^n\{(\boldsymbol{x}, \boldsymbol{y}) \in A_\epsilon^n\} \geq 1 - \epsilon, \tag{3}$$

$$|A_\epsilon^n| \leq \exp\{n(H(X,Y) + \epsilon)\}. \tag{4}$$

Further, for each $\boldsymbol{y} \in \mathcal{Y}^n$, let the set of $\boldsymbol{x}$ which is jointly typical with $\boldsymbol{y}$ be defined by $A_\epsilon^n(\boldsymbol{y}) = \{\boldsymbol{x} \in \mathcal{X}^n : (\boldsymbol{x}, \boldsymbol{y}) \in A_\epsilon^n\}$. Then, we have

$$|A_\epsilon^n(\boldsymbol{y})| \leq \exp\{n(H(X|Y) + \epsilon)\}. \tag{5}$$

By using the set A_ϵ^n of typical sequences, we consider the following typical set decoder [2].

Definition 5 (Typical Set Decoder): For a given pair of encoders (f, g) with block length n, let $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$ be the outputs of two encoders. Then, for a given $\epsilon > 0$, the typical set decoder outputs a pair $(\boldsymbol{x}, \boldsymbol{y}) \in A_\epsilon^n$ if there is one and only one pair of $(\boldsymbol{x}, \boldsymbol{y})$ such that $f(\boldsymbol{x}) = m_1$ and $g(\boldsymbol{y}) = m_2$. Otherwise, the decoder declares an error.

c. Set of linear encoders

In what follows, $|\mathcal{X}|$ and $|\mathcal{Y}|$ are assumed to be powers of two. Unless both sizes of the source alphabets $\mathcal{X}$ and $\mathcal{Y}$ are powers of two, we can add some dummy symbols with probability zero in order to satisfy this condition. This allows us to endow $\mathcal{X}$ and $\mathcal{Y}$ with the structure of Galois fields, and $\mathcal{X}^n$ and $\mathcal{Y}^n$ are considered as vector spaces over these fields. Further, let us endow the vector spaces $\mathcal{X}^k$ and $\mathcal{Y}^k$ with the structure of the extension fields of $\mathcal{X}$ and $\mathcal{Y}$, respectively. The encoder $f : \mathcal{X}^n \to \mathcal{X}^k$ is said to be linear if f is an $\mathcal{X}$-linear mapping from $\mathcal{X}^n$ into $\mathcal{X}^k$. The linearity of the encoder $g : \mathcal{Y}^n \to \mathcal{Y}^k$ can be defined similarly.

We first describe a set of linear encoders for $\mathcal{X}^n$. For any positive integer n and $k(\leq n)$, define the decomposition $n = ku + m$ with $u \stackrel{\triangle}{=} \lceil n/k \rceil - 1$ and $1 \leq m \leq k$, where $\lceil z \rceil$ denotes the minimum integer greater than or equal to z. Then, $\boldsymbol{x} \in \mathcal{X}^n$ can be rewritten as $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_u)$, where $\boldsymbol{x}_0 \in \mathcal{X}^m$ and $\boldsymbol{x}_i \in \mathcal{X}^k$ $(i = 1, 2, \cdots, u)$. To each element $\gamma \in \mathcal{X}^k$, we associate the linear encoder $f : \mathcal{X}^n \to \mathcal{X}^k$

given by

$$f(\boldsymbol{x}) = \psi(\boldsymbol{x}_0) + \sum_{i=1}^{u} \gamma^i \boldsymbol{x}_i, \tag{6}$$

where $\psi(\boldsymbol{z})$ denotes the k-dimensional vector over $\mathcal{X}$ formed by adding $k-m$ zeros components after the components of m-dimensional vector $\boldsymbol{z}$. This encoding operation (6) is interpreted as follows: Since γ and $\boldsymbol{x}_i$ $(i = 1, 2, \cdots, u)$ are elements of the Galois field $\mathcal{X}^k$, $\sum_{i=1}^{u} \gamma^i \boldsymbol{x}_i$ can be computed over the Galois field $\mathcal{X}^k$. Then, by interpreting $\sum_{i=1}^{u} \gamma^i \boldsymbol{x}_i$ as a k-dimensional vector over the Galois field $\mathcal{X}$, the last addition can be computed.

We define $C(n, k, \mathcal{X})$ to be a set of encoders given by (6) for all $\gamma \in \mathcal{X}^k$. Obviously, we have $|C(n, k, \mathcal{X})| = |\mathcal{X}|^k$. In a similar manner, we also define the set $C(n, k, \mathcal{Y})$ of encoders $g : \mathcal{Y}^n \to \mathcal{Y}^k$ with cardinality $|C(n, k, \mathcal{Y})| = |\mathcal{Y}|^k$.

d. Generalized Hermitian codes

We introduce a class of error correcting codes (ECC) used in the following chapters. As for the details of ECC, please refer to e.g. [10].

In 1981, Goppa [11] introduced some methods of algebraic geometry to the construction of codes, and high-performance codes are constructed from various algebraic curves. Among them, Shen [12] showed an explicit construction of high-performance codes C_H from generalized Hermitian curves.

Definition 6 (Codes from Generalized Hermitian Curves): A code $C_H(N, K, D)$ constructed from a generalized Hermitian curve is a linear ECC over $GF(2^{2m})$ with

length N, dimension K and minimum distance D satisfying

$$\left.\begin{array}{l} 0 < N < 2^{m(\ell+1)} \\ K \leq N - g(\ell, 2^m) \\ D \geq N - K + 1 - g(\ell, 2^m) \end{array}\right\}, \tag{7}$$

while ℓ is an integer greater than 1, and the genus $g(\ell, 2^m)$ is given by

$$g(\ell, x) = \frac{1}{2}\left\{\sum_{i=1}^{\ell-1} x^{\ell+1-i}(x+1)^{i-1} - (x+1)^{\ell-1} + 1\right\}. \tag{8}$$

Since this code is constructed from a generalized Hermitian curve, we call $C_H(N, K, D)$ as a generalized Hermitian code.

Since Reed-Solomon code has the parameters (N, K, D) satisfying (7) with $\ell = 1$, the generalized Hermitian code may be regarded as a generalization of Reed-Solomon code. As for details of the code construction of generalized Hermitian codes, please refer to [12]. Generalized Hermitian codes can be efficiently decoded up to the designed minimum distance $D_{des} \stackrel{\triangle}{=} N - K + 1 - g(\ell, 2^m)$ by employing such algorithms as that of Shen and Tzeng [13], and its computational complexity is $O(N^3)$.

III. Main Result

In this chapter, we propose explicit encoding and decoding procedures for SW networks, and show some properties of the procedures. First, we shall describe the encoding procedure.

Encoding procedure: Denote all pairs of mappings in $C(n, k_1, \mathcal{X}) \times C(n, k_2, \mathcal{Y})$ as (f_i, g_i) $(i = 1, 2, \cdots, N)$ where

$$N = |C(n, k_1, \mathcal{X})| \cdot |C(n, k_2, \mathcal{Y})| = |\mathcal{X}|^{k_1} \cdot |\mathcal{Y}|^{k_2}.$$

Then, consider the following fixed length code with block length $N_o = nN$, where n is an even integer.

1. Let a given sequence $\boldsymbol{x} \in \mathcal{X}^{N_o}$ be represented in the form of N-dimensional vector over $\mathcal{X}^n$, i.e. $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N)$ with $\boldsymbol{x}_i \in \mathcal{X}^n$ $(i = 1, 2, \cdots, N)$. Then, encode each subblock $\boldsymbol{x}_i$ into $f_i(\boldsymbol{x}_i) \in \mathcal{X}^{k_1}$ for $i = 1, 2, \cdots, N$. Similarly, for a given sequence $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N) \in \mathcal{Y}^{N_o}$ with $\boldsymbol{y}_i \in \mathcal{Y}^n$ $(i = 1, 2, \cdots, N)$, encode each subblock $\boldsymbol{y}_i$ into $g_i(\boldsymbol{y}_i) \in \mathcal{Y}^{k_2}$ for $i = 1, 2, \cdots, N$.

2. Encode the N-dimensional vector $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N)$ into $\boldsymbol{x}H_1^t \in \mathcal{X}^{nK_1}$ (K_1-dimensional vector over $\mathcal{X}^n$), and encode the vector $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N)$ into $\boldsymbol{y}H_2^t \in \mathcal{Y}^{nK_2}$, where H_1 and H_2 denote the parity check matrices of the code $C_H(N, N - K_1, D_1)$ over $\mathcal{X}^n$ and the code $C_H(N, N - K_2, D_2)$ over $\mathcal{Y}^n$, respectively.

Therefore, the encoded sequences consist of N pairs of $(f_i(\boldsymbol{x}), g_i(\boldsymbol{y})) \in \mathcal{X}^{k_1} \times \mathcal{Y}^{k_2}$ and a pair of $(\boldsymbol{x}H_1^t, \boldsymbol{y}H_2^t) \in \mathcal{X}^{nK_1} \times \mathcal{Y}^{nK_2}$.

Obviously, the proposed code is linear, and the overall computational complexity of the encoding procedure is at most $O(N_o^2)$. Further, let us define the rates of the

first and second steps of encodings by

$$\left.\begin{array}{ll} r_1 \stackrel{\triangle}{=} \frac{k_1}{n}\log|\mathcal{X}|, & r_2 \stackrel{\triangle}{=} \frac{k_2}{n}\log|\mathcal{Y}|, \\ \tilde{r}_1 \stackrel{\triangle}{=} \frac{K_1}{N}\log|\mathcal{X}|, & \tilde{r}_2 \stackrel{\triangle}{=} \frac{K_2}{N}\log|\mathcal{Y}|, \end{array}\right\} \tag{9}$$

then the rate pair (R_1, R_2) of the overall code are given by

$$(R_1, R_2) = (r_1 + \tilde{r}_1, r_2 + \tilde{r}_2). \tag{10}$$

Next, we shall describe the decoding procedure.

Decoding procedure: For obtained sequences $(f_i(\boldsymbol{x}), g_i(\boldsymbol{y}))$ $i = 1, 2, \cdots, N$ and $(\boldsymbol{x}H_1^t, \boldsymbol{y}H_2^t) \in \mathcal{X}^{nK_1} \times \mathcal{Y}^{nK_2}$, perform the following two-step decoding.

1. For $i = 1, 2, \cdots, N$, decode $(f_i(\boldsymbol{x}_i), g_i(\boldsymbol{y}_i))$ by a typical set decoder φ_i corresponding to (f_i, g_i), and obtain the first estimate $(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{y}}_i) \in \mathcal{X}^n \times \mathcal{Y}^n$. The overall first estimate $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) \in \mathcal{X}^{N_o} \times \mathcal{Y}^{N_o}$ of the encoded sequence can be described as $\hat{\boldsymbol{x}} = (\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, \cdots, \hat{\boldsymbol{x}}_N)$ and $\hat{\boldsymbol{y}} = (\hat{\boldsymbol{y}}_1, \hat{\boldsymbol{y}}_2, \cdots, \hat{\boldsymbol{y}}_N)$.

2. From two syndromes $\boldsymbol{s}_1 \stackrel{\triangle}{=} \boldsymbol{x}H_1^t - \hat{\boldsymbol{x}}H_1^t$ and $\boldsymbol{s}_2 \stackrel{\triangle}{=} \boldsymbol{y}H_2^t - \hat{\boldsymbol{y}}H_2^t$, find vectors with minimum Hamming weight $\boldsymbol{e}_1 \in \mathcal{X}^{N_o}$ and $\boldsymbol{e}_2 \in \mathcal{Y}^{N_o}$ such that $\boldsymbol{e}_1 H_1^t = \boldsymbol{s}_1$ and $\boldsymbol{e}_2 H_2^t = \boldsymbol{s}_2$. These vectors can be obtained efficiently by using the error correcting procedure of algebraic geometry code. Then, $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) = (\hat{\boldsymbol{x}} - \boldsymbol{e}_1, \hat{\boldsymbol{y}} - \boldsymbol{e}_2)$ is the final estimate of the encoded pair.

In the first step of decoding, if the decoding of the i-th code (f_i, g_i) is successful, both $\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i$ and $\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i$ become zero vectors. Hence, by using the linearity of the syndrome, we can obtain the syndromes $\boldsymbol{s}_1 = (\boldsymbol{x} - \hat{\boldsymbol{x}})H_1^t$ and $\boldsymbol{s}_2 = (\boldsymbol{y} - \hat{\boldsymbol{y}})H_2^t$ corresponding to the errors produced by incorrect estimations of the first step decoding.

It is essential to use linear codes in the second step of encoding, since we cannot know which code (f_i, g_i, φ_i) produces errors at the time of encoding.

Since the decoding of the generalized Hermitian code can be done in the computational complexity of $O(N^3)$ (e.g. [13]), the overall computational complexity of the decoding procedure is $O(N_o^3)$.

Remark 1: If we are allowed to search a code for the first step of encoding, we can construct a code with much smaller block length. For an arbitrary $\epsilon > 0$ and sufficiently large block length n, assume that we can find the code (f_o, g_o, φ_o) satisfying $p_e(f_o, g_o, \varphi_o) \leq \epsilon$. Then, we modify the proposed encoding procedure as follows: In the first step of encoding, for a given $N(\leq \min(|\mathcal{X}|^n, |\mathcal{Y}|^n) - 1)$, only one encoder pair (f_o, g_o) is used repeatedly for all $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ $(i = 1, 2, \cdots, N)$. In the second step of encoding, Reed-Solomon codes over $\mathcal{X}^n$ and $\mathcal{Y}^n$ with length N are used instead of generalized Hermitian codes. This construction is analogous to the original concatenated code proposed by Forney for channel coding [8].

Remark 2: By using a random coding technique, Csiszár [9] showed that there exists a linear code for SW networks, if the rate pair (R_1, R_2) is an inner point of the achievable region. However, our result is different from his results in the following points. (i) Our code construction is deterministic, and once we construct the Galois fields $\mathcal{X}^n$, $\mathcal{X}^{k_1}$, $\mathcal{Y}^n$, and $\mathcal{Y}^{k_2}$, the construction of codes $C(n, k_1, \mathcal{X})$, $C(n, k_2, \mathcal{Y})$, and parity check matrices for $C_H(N, N - K_1, D_1)$ and $C_H(N, N - K_2, D_2)$ can be done within the computational complexity of $O(N_o^2)$, i.e. the polynomial order of the

block length. On the contrary, in order to obtain a good linear code by Csiszár's approach, it requires the computational complexity of exponential order of the block length. (ii) Csiszár proposed to employ minimum entropy decoding for the linear codes, but its computational complexity is an exponential order of the block length. On the contrary, the computational complexity of the decoding for our code is only $O(N_o^3)$. Hence, our code is much more efficient than that proposed by Csiszár. (iii) We consider the code construction for stationary ergodic sources, whereas Csiszár restricted his attention to memoryless sources.

The next theorem shows that the probability of error can become arbitrary small for sufficiently large block length.

Theorem 2: Suppose that the proposed code is applied to a jointly ergodic source (X, Y). Then, for every $\delta > 0$, we have $p_e \leq \delta$ for sufficiently large N_o, provided that $r_1 > H(X|Y)$, $r_2 > H(Y|X)$, $r_1 + r_2 > H(X,Y)$, $\tilde{r}_1 > 0$ and $\tilde{r}_2 > 0$ hold simultaneously.

If the rate pair (R_1, R_2) is an inner point of the achievable region of the SW network, according to Theorem 1, we have $R_1 > H(X|Y)$, $R_2 > H(Y|X)$ and $R_1 + R_2 > H(X,Y)$. Since $R_1 = r_1 + \tilde{r}_1$ and $R_2 = r_2 + \tilde{r}_2$, we can choose a pair (r_1, r_2) such that $r_1 > H(X|Y)$, $r_2 > H(Y|X)$, $r_1 + r_2 > H(X,Y)$, $\tilde{r}_1 > 0$ and $\tilde{r}_2 > 0$. Therefore, if the rate pair (R_1, R_2) is an inner point of the achievable region, we can construct the code achieving arbitrary small probability of error. This discussion is summarized in the following corollary.

Corollary 1: Let (X,Y) be a jointly ergodic source. Assume that the rate pair (R_1,R_2) satisfies all the conditions $R_1 > H(X|Y)$, $R_2 > H(Y|X)$ and $R_1 + R_2 > H(X,Y)$. Then, for any $\epsilon > 0$ we can choose rate pairs for the first and second encoding (r_1,r_2) and $(\tilde{r}_1,\tilde{r}_2)$ such that the resulting code satisfies $p_e \leq \epsilon$ for sufficiently large N_o.

Proof of Theorem 2: For $\boldsymbol{x},\boldsymbol{x}' \in \mathcal{X}^n$ with $\boldsymbol{x} \neq \boldsymbol{x}'$, we define $\nu_1(\boldsymbol{x},\boldsymbol{x}')$ by

$$\nu_1(\boldsymbol{x},\boldsymbol{x}') \stackrel{\triangle}{=} |\{f \in C(n,k_1,\mathcal{X}) : f(\boldsymbol{x}) = f(\boldsymbol{x}')\}|.$$

Similarly, for $\boldsymbol{y},\boldsymbol{y}' \in \mathcal{Y}^n$ with $\boldsymbol{y} \neq \boldsymbol{y}'$, define

$$\nu_2(\boldsymbol{y},\boldsymbol{y}') \stackrel{\triangle}{=} |\{g \in C(n,k_2,\mathcal{Y}) : g(\boldsymbol{y}) = g(\boldsymbol{y}')\}|.$$

Since f is linear, the condition $f(\boldsymbol{x}) = f(\boldsymbol{x}')$ is equivalent to $f(\boldsymbol{x}'') = 0$ for a nonzero vector $\boldsymbol{x}''(= \boldsymbol{x} - \boldsymbol{x}')$. Hence, according to (6) and (9), it is easy to see that

$$\left.\begin{array}{rcl} \frac{\nu_1(\boldsymbol{x},\boldsymbol{x}')}{|C(n,k_1,\mathcal{X})|} & \leq & u_1|\mathcal{X}|^{-k_1} = u_1 \exp\{-nr_1\} \\ \frac{\nu_2(\boldsymbol{y},\boldsymbol{y}')}{|C(n,k_2,\mathcal{Y})|} & \leq & u_2|\mathcal{Y}|^{-k_2} = u_2 \exp\{-nr_2\} \end{array}\right\} \tag{11}$$

for every $\boldsymbol{x} \neq \boldsymbol{x}'$ and $\boldsymbol{y} \neq \boldsymbol{y}'$, where u_1 and u_2 are given by

$$\begin{aligned} u_1 &\stackrel{\triangle}{=} \lceil n/k_1 \rceil - 1 = \lceil \log|\mathcal{X}|/r_1 \rceil - 1, \\ u_2 &\stackrel{\triangle}{=} \lceil n/k_2 \rceil - 1 = \lceil \log|\mathcal{Y}|/r_2 \rceil - 1. \end{aligned}$$

Since we employ typical set decoders, for any $\epsilon > 0$, the probability of error for the i-th code (f_i,g_i,φ_i) can be bounded as

$$p_e(f_i,g_i,\varphi_i) \leq Q^n((A_\epsilon^n)^c) + Q^n(E_1(f_i)) + Q^n(E_2(g_i)) + Q^n(E_3(f_i,g_i)),$$

where A_ϵ^n is the set of typical sequence defined in (2), and

$$\begin{aligned}
E_1(f) &\stackrel{\triangle}{=} \{(\boldsymbol{x},\boldsymbol{y}) \in A_\epsilon^n : f(\boldsymbol{x}') = f(\boldsymbol{x}) \text{ and } (\boldsymbol{x}',\boldsymbol{y}) \in A_\epsilon^n \text{ for some } \boldsymbol{x}'(\neq \boldsymbol{x})\} \\
E_2(g) &\stackrel{\triangle}{=} \{(\boldsymbol{x},\boldsymbol{y}) \in A_\epsilon^n : g(\boldsymbol{y}') = g(\boldsymbol{y}) \text{ and } (\boldsymbol{x},\boldsymbol{y}') \in A_\epsilon^n \text{ for some } \boldsymbol{y}'(\neq \boldsymbol{y})\} \\
E_3(f,g) &\stackrel{\triangle}{=} \{(\boldsymbol{x},\boldsymbol{y}) \in A_\epsilon^n : f(\boldsymbol{x}') = f(\boldsymbol{x}), g(\boldsymbol{y}') = g(\boldsymbol{y}) \text{ and } (\boldsymbol{x}',\boldsymbol{y}') \in A_\epsilon^n \\
&\qquad \text{for some } \boldsymbol{x}'(\neq \boldsymbol{x}) \text{ and } \boldsymbol{y}'(\neq \boldsymbol{y})\}.
\end{aligned}$$

Then, by using (5) and (11), we have

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N} Q^n(E_1(f_i)) \\
&= \frac{1}{|C(n,k_1,\mathcal{X})|}\sum_{f\in C(n,k_1,\mathcal{X})} Q^n(E_1(f)) \\
&\le \frac{1}{|C(n,k_1,\mathcal{X})|}\sum_{(\boldsymbol{x},\boldsymbol{y})\in A_\epsilon^n} Q^n(\boldsymbol{x},\boldsymbol{y}) \sum_{\substack{\boldsymbol{x}'\neq\boldsymbol{x}:\\ (\boldsymbol{x}',\boldsymbol{y})\in A_\epsilon^n}} \nu_1(\boldsymbol{x},\boldsymbol{x}') \\
&\le \sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}^n\times\mathcal{Y}^n} Q^n(\boldsymbol{x},\boldsymbol{y}) u_1 \exp\{-nr_1\}|A_\epsilon^n(\boldsymbol{y})| \\
&\le u_1 \exp\{-n(r_1 - H(X|Y) - \epsilon)\}.
\end{aligned}$$

Since we can choose $\epsilon(>0)$ such that $r_1 > H(X|Y) + \epsilon$, we have

$$\frac{1}{N}\sum_{i=1}^{N} Q^n(E_1(f_i)) \le \epsilon,$$

for sufficiently large n. Similarly, we have

$$\frac{1}{N}\sum_{i=1}^{N} Q^n(E_2(g_i)) \le \epsilon,$$

for sufficiently large n. On the other hand, by using (4) and (11), we have

$$\frac{1}{N}\sum_{i=1}^{N} Q^n(E_3(f_i,g_i))$$

$$
\begin{aligned}
&= \frac{1}{|C(n,k_1,\mathcal{X})|\cdot|C(n,k_2,\mathcal{Y})|}\sum_{\substack{f\in C(n,k_1,\mathcal{X})\\ g\in C(n,k_2,\mathcal{Y})}} Q^n(E_3(f,g))\\
&\le \frac{1}{|C(n,k_1,\mathcal{X})|\cdot|C(n,k_2,\mathcal{Y})|}\sum_{(\boldsymbol{x},\boldsymbol{y})\in A_\epsilon^n} Q^n(\boldsymbol{x},\boldsymbol{y})\sum_{\substack{(\boldsymbol{x}',\boldsymbol{y}')\in A_\epsilon^n:\\ \boldsymbol{x}'\neq\boldsymbol{x},\boldsymbol{y}'\neq\boldsymbol{y}}}\nu_1(\boldsymbol{x},\boldsymbol{x}')\nu_2(\boldsymbol{y},\boldsymbol{y}')\\
&\le \sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}^n\times\mathcal{Y}^n} Q^n(\boldsymbol{x},\boldsymbol{y})u_1u_2\exp\{-n(r_1+r_2)\}|A_\epsilon^n|\\
&\le u_1u_2\exp\{-n(r_1+r_2-H(X,Y)-\epsilon)\}.
\end{aligned}
$$

Since we can choose $\epsilon(>0)$ such that $r_1+r_2>H(X,Y)+\epsilon$, we have

$$\frac{1}{N}\sum_{i=1}^{N} Q^n(E_3(f_i,g_i))\le\epsilon,$$

for sufficiently large n. Hence, the average probability of error for the code (f_i,g_i,φ_i) $i=1,2,\cdots,N$ can be bounded by

$$\frac{1}{N}\sum_{i=1}^{N} p_e(f_i,g_i,\varphi_i)\le Q^n((A_\epsilon^n)^c)+3\epsilon\le 4\epsilon,$$

for sufficiently large n. This implies that for $\boldsymbol{x}=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)\in\mathcal{X}^{N_o}$ and $\boldsymbol{y}=(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_N)\in\mathcal{Y}^{N_o}$

$$
\begin{aligned}
&\sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{X}^{N_o}\times\mathcal{Y}^{N_o}} Q^{N_o}(\boldsymbol{x},\boldsymbol{y})\sum_{i=1}^{N}1\{\varphi_i(f_i(\boldsymbol{x}_i),g_i(\boldsymbol{y}_i))\neq(\boldsymbol{x}_i,\boldsymbol{y}_i)\}\\
&= \sum_{i=1}^{N}\sum_{(\boldsymbol{x}_i,\boldsymbol{y}_i)\in\mathcal{X}^n\times\mathcal{Y}^n} Q^n(\boldsymbol{x}_i,\boldsymbol{y}_i)1\{\varphi_i(f_i(\boldsymbol{x}_i),g_i(\boldsymbol{y}_i))\neq(\boldsymbol{x}_i,\boldsymbol{y}_i)\}\\
&\le 4N\epsilon,
\end{aligned}
$$

where $1\{\cdot\}$ denotes the indicator function. Hence, among N decoders in the first step of decoding, the average number of decoders which produce errors can be bounded by $4N\epsilon$. Since both generalized Hermitian codes can correct at least $\lfloor(\min(D_1,D_2)-$

$1)/2\rfloor$ errors, the average probability of error p_e after the second step of decoding can be bounded by

$$\begin{aligned} p_e &\leq \Pr\Big\{(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^{N_o} \times \mathcal{Y}^{N_o} : \\ &\qquad \sum_{i=1}^{N} 1\{\varphi_i(f_i(\boldsymbol{x}_i), g_i(\boldsymbol{y}_i)) \neq (\boldsymbol{x}_i, \boldsymbol{y}_i)\} \geq \lfloor (\min(D_1, D_2)+1)/2 \rfloor \Big\} \\ &\leq \frac{\sum_{i=1}^{N} \sum_{(\boldsymbol{x}_i,\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n} Q^n(\boldsymbol{x}_i, \boldsymbol{y}_i) 1\{\varphi_i(f_i(\boldsymbol{x}), g_i(\boldsymbol{y})) \neq (\boldsymbol{x}, \boldsymbol{y})\}}{\lfloor (\min(D_1, D_2)+1)/2 \rfloor} \\ &\leq \frac{8N\epsilon}{\min(D_1, D_2)}, \end{aligned} \tag{12}$$

where the first inequality comes from Markov's inequality. For the code $C_H(N, K_1, D_1)$, the parameter ℓ is an arbitrarily fixed integer satisfying $\ell > 2(r_1 + r_2)/\log|\mathcal{X}| - 1$. Then, according to (7) and (8), we have

$$\lim_{N\to\infty} \frac{D_1}{N} \geq \lim_{n\to\infty} \left(\frac{\tilde{r}_1}{\log|\mathcal{X}|} - \frac{g(\ell, |\mathcal{X}|^{n/2}) - 1}{N} \right) = \frac{\tilde{r}_1}{\log|\mathcal{X}|}, \tag{13}$$

Similarly we also have $\lim_{N\to\infty} D_2/N = \tilde{r}_2/\log|\mathcal{Y}|$. Combining these results with (12), we complete the proof. □

IV. Discrete Memoryless Sources

In this chapter, we restrict our attention to discrete memoryless sources (DMS's). First, we show that our proposed code can be modified such that both encoding and decoding of the code do not depend on the generic distribution of DMS's. Next, we clarify that the probability of error can vanish exponentially as the block length tends to infinity.

First, we need some preliminaries about the type of sequence [4, 9], and the minimum entropy decoder [9].

Definition 7 (Joint Type): Let the joint type $P_{\boldsymbol{xy}}$ of sequences $\boldsymbol{x} \in \mathcal{X}^n$, $\boldsymbol{y} \in \mathcal{Y}^n$ be the distribution on $\mathcal{X} \times \mathcal{Y}$ defined by the empirical distribution of the elements of $\mathcal{X} \times \mathcal{Y}$ in $(\boldsymbol{x}, \boldsymbol{y})$.

Definition 8 (Minimum Entropy Decoder): The minimum entropy decoder corresponding to a pair of encoders $f : \mathcal{X}^n \to \mathcal{M}_1$, $g : \mathcal{Y}^n \to \mathcal{M}_2$ maps each pair of codewords $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$ into a pair $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ minimizing $H(P_{\boldsymbol{xy}})$ subject to $f(\boldsymbol{x}) = m_1$ and $g(\boldsymbol{y}) = m_2$, where $H(P_{\boldsymbol{xy}})$ denotes the entropy of the joint type $P_{\boldsymbol{xy}}$. Obviously this decoding rule does not depend on the generic distribution of DMS's.

Since the proposed encoding procedure does not depend on the generic distributions of DMS's, the encoding procedure remains unchanged. On the contrary, the proposed decoding procedure is modified as follows: In the first step of decoding, for $i = 1, 2, \cdots, N$ decode $(f_i(\boldsymbol{x}_i), g_i(\boldsymbol{y}_i))$ by using *minimum entropy decoder* instead of typical set decoder. Since minimum entropy decoder does not depend on the generic distribution of DMS's, this modified code is universal.

To describe our result, for $R \geq 0$ and distributions Q on $\mathcal{X} \times \mathcal{Y}$ define the

exponent functions as

$$\left.\begin{aligned} e_r^1(R,Q) &\stackrel{\triangle}{=} \min[D(P_{XY} \parallel Q) + |R - H(X|Y)|^+] \\ e_r^2(R,Q) &\stackrel{\triangle}{=} \min[D(P_{XY} \parallel Q) + |R - H(Y|X)|^+] \\ e_r^3(R,Q) &\stackrel{\triangle}{=} \min[D(P_{XY} \parallel Q) + |R - H(XY)|^+] \end{aligned}\right\}. \tag{14}$$

Here the minimizations are over all dummy random variables X, Y, and P_{XY} denotes their joint distribution; further $|t|^+ \stackrel{\triangle}{=} \max(0,t)$, and $D(P \parallel Q)$ denotes Kullback-Leibler informational divergence [4].

The next theorem shows that the probability of error for the proposed code vanishes exponentially with block length N_o.

Theorem 3: For any $\epsilon > 0$ and sufficiently large block length N_o, the probability of error for the proposed code can be bounded, universally for every Q, as

$$p_e \leq 2 \exp\left[-N_o\left\{\frac{1}{2}\min\left(\frac{\tilde{r}_1}{\log|\mathcal{X}|}, \frac{\tilde{r}_2}{\log|\mathcal{Y}|}\right) E_r(r_1, r_2, Q) - \epsilon\right\}\right], \tag{15}$$

where

$$E_r(r_1, r_2, Q) \stackrel{\triangle}{=} \min_{1 \leq i \leq 3} e_r^i(r_i, Q), \tag{16}$$

with $r_3 = r_1 + r_2$.

As shown in [9], $E_r(r_1, r_2, Q) > 0$ whenever (r_1, r_2) is an inner point of the achievable region of the SW network with generic distribution Q. Hence, if (R_1, R_2) is an inner point of the achievable rate region, that is, if all the conditions $R_1 > H(X|Y)$, $R_2 > H(Y|X)$ and $R_1 + R_2 > H(X,Y)$ are satisfied, we can choose r_1 and r_2 such that $E_r(r_1, r_2, Q) > 0$, $\tilde{r}_1 > 0$ and $\tilde{r}_2 > 0$. Therefore, we can achieve a

positive error exponent by the proposed code, whenever (R_1, R_2) is an inner point of the achievable rate region of the SW network.

The next corollary shows the error exponent obtainable by the proposed code for the generic distribution Q.

Corollary 2: For the generic distribution Q, the lower bound on the error exponent obtainable by the proposed codes is given by

$$E_p(R_1, R_2, Q) = \max_{\substack{0<r_1<R_1 \\ 0<r_2<R_2}} \left[\frac{1}{2}\min\left(\frac{R_1 - r_1}{\log|\mathcal{X}|}, \frac{R_2 - r_2}{\log|\mathcal{Y}|}\right) E_r(r_1, r_2, Q)\right], \tag{17}$$

where $E_r(r_1, r_2, Q)$ is given by (16).

Before we prove Theorem 3, we shall explain some notations and the method of type.

To simplify the notation, (joint) types of sequences will be considered as (joint) distributions such as P_X (P_{XY}) of dummy random variables X and Y. The set of different types of sequences in $\mathcal{X}^n$ will be denoted by $\mathcal{P}_n(\mathcal{X})$.

For a type $P_X \in \mathcal{P}_n(\mathcal{X})$ and a joint type $P_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$, define T_X and $T_{Y|X}(\boldsymbol{x})$ as

$$\begin{aligned} T_X &\stackrel{\triangle}{=} \{\boldsymbol{x} \in \mathcal{X}^n : P_{\boldsymbol{x}} = P_X\}, \\ T_{Y|X}(\boldsymbol{x}) &\stackrel{\triangle}{=} \{\boldsymbol{y} \in \mathcal{Y}^n : P_{\boldsymbol{x}\boldsymbol{y}} = P_{XY}\} \quad \text{for } \boldsymbol{x} \in T_X. \end{aligned}$$

The following upper bounds on $\mathcal{P}_n$, T_X and $T_{Y|X}(\boldsymbol{x})$ are well known (see e.g.[4]):

$$\left.\begin{array}{rcl} |\mathcal{P}_n(\mathcal{X})| & \leq & (n+1)^{|\mathcal{X}|}, \\ |T_X| & \leq & \exp\{nH(X)\}, \\ |T_{Y|X}(\boldsymbol{x})| & \leq & \exp\{nH(Y|X)\}, \quad \boldsymbol{x} \in T_X. \end{array}\right\} \tag{18}$$

For each joint type $P_{X\tilde{X}Y\tilde{Y}} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Y})$ and a pair of encoders (f, g) we first define the subset $N_{f,g}(X\tilde{X}Y\tilde{Y})$ of T_{XY} as follows:

$$\begin{aligned} N_{f,g}(X\tilde{X}Y\tilde{Y}) \stackrel{\triangle}{=} \{&(\boldsymbol{x}, \boldsymbol{y}) \in T_{XY} : \text{there exist some } (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \neq (\boldsymbol{x}, \boldsymbol{y}) \text{ such that} \\ &P_{\boldsymbol{x}\tilde{\boldsymbol{x}}\boldsymbol{y}\tilde{\boldsymbol{y}}} = P_{X\tilde{X}Y\tilde{Y}}, f(\boldsymbol{x}) = f(\tilde{\boldsymbol{x}}) \text{ and } g(\boldsymbol{y}) = g(\tilde{\boldsymbol{y}})\} \end{aligned} \tag{19}$$

The next lemma is a strong version of Csiszár's Lemma [9, p.587], and is essential in this chapter.

Lemma 2: For any $\epsilon > 0$, any positive integers k_1, k_2 and sufficiently large integer n, there exist at least $N(1 - \exp(-n\epsilon))$ pairs of linear encoders $(f, g) \in C(n, k_1, \mathcal{X}) \times C(n, k_2, \mathcal{Y})$ satisfying

$$N_{f,g}(X\tilde{X}Y\tilde{Y}) \leq \left\{\begin{array}{l} |T_{XY}| \exp\{-n(r_1 + r_2 - H(\tilde{X}\tilde{Y}|XY) - 2\epsilon)\} \\ \qquad \text{if } X \neq \tilde{X}, Y \neq \tilde{Y} \\ |T_{XY}| \exp\{-n(r_1 - H(\tilde{X}|XY) - 2\epsilon)\} \\ \qquad \text{if } Y = \tilde{Y} \\ |T_{XY}| \exp\{-n(r_2 - H(\tilde{Y}|XY) - 2\epsilon)\} \\ \qquad \text{if } X = \tilde{X} \end{array}\right. , \tag{20}$$

for every joint type $P_{X\tilde{X}Y\tilde{Y}} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Y})$.

Proof: This lemma can be proven in a similar manner as Csiszár's Lemma in [9, p.588]. According to (11), for every nonzero vectors $\boldsymbol{x} \in \mathcal{X}^n$ and $\boldsymbol{y} \in \mathcal{Y}^n$, we have

$$\left.\begin{array}{rcl} \frac{|\{f \in C(n,k_1,\mathcal{X}): f(\boldsymbol{x})=\boldsymbol{O}\}|}{|C(n,k_1,\mathcal{X})|} & \leq & u_1 \exp\{-nr_1\}, \\ \frac{|\{g \in C(n,k_2,\mathcal{Y}): g(\boldsymbol{y})=\boldsymbol{O}\}|}{|C(n,k_2,\mathcal{Y})|} & \leq & u_2 \exp\{-nr_2\}. \end{array}\right\} \tag{21}$$

Consider any joint type $P_{X\tilde{X}Y\tilde{Y}} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Y})$ with $X \neq \tilde{X}$ and $Y \neq \tilde{Y}$. Since both f and g are linear, the condition in (19) is equivalent to $f(\boldsymbol{x} - \tilde{\boldsymbol{x}}) = 0$, $g(\boldsymbol{y} - \tilde{\boldsymbol{y}}) = 0$. Hence, according to (18) and (21), we immediately have

$$\begin{aligned} & \frac{1}{N} \sum_{f \in C(n,k_1,\mathcal{X})} \sum_{g \in C(n,k_2,\mathcal{Y})} N_{f,g}(X\tilde{X}Y\tilde{Y}) \\ \leq\ & \frac{1}{N} \sum_{f \in C(n,k_1,\mathcal{X})} \sum_{g \in C(n,k_2,\mathcal{Y})} \\ & \times \sum_{(\boldsymbol{x},\boldsymbol{y}) \in T_{XY}} \sum_{\substack{(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}) \in T_{\tilde{X}\tilde{Y}|XY}(\boldsymbol{x},\boldsymbol{y}): \\ (\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}) \neq (\boldsymbol{x},\boldsymbol{y})}} 1\{f(\boldsymbol{x} - \tilde{\boldsymbol{x}}) = \boldsymbol{O} \text{ and } g(\boldsymbol{x} - \tilde{\boldsymbol{x}}) = \boldsymbol{O}\} \\ \leq\ & \sum_{(\boldsymbol{x},\boldsymbol{y}) \in T_{XY}} \sum_{\substack{(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}) \in T_{\tilde{X}\tilde{Y}|XY}(\boldsymbol{x},\boldsymbol{y}): \\ (\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}) \neq (\boldsymbol{x},\boldsymbol{y})}} \frac{|\{f \in C(n,k_1,\mathcal{X}) : f(\boldsymbol{x} - \tilde{\boldsymbol{x}}) = \boldsymbol{O}\}|}{|C(n,k_1,\mathcal{X})|} \\ & \times \frac{|\{g \in C(n,k_2,\mathcal{Y}) : g(\boldsymbol{y} - \tilde{y}) = \boldsymbol{O}\}|}{|C(n,k_2,\mathcal{Y})|} \\ \leq\ & u_1 u_2 |T_{XY}| \exp\{-n(r_1 + r_2 - H(\tilde{X}\tilde{Y}|XY))\}. \end{aligned} \tag{22}$$

For joint types $P_{X\tilde{X}Y\tilde{Y}}$ with $Y = \tilde{Y}$, $N_{f,g}(X\tilde{X}Y\tilde{Y})$ is equal to the number of pairs $(\boldsymbol{x}, \boldsymbol{y}) \in T_{XY}$ such that for some $\tilde{\boldsymbol{x}} \in T_{\tilde{X}|XY}(\boldsymbol{x}, \boldsymbol{y})$ the relation $f(\boldsymbol{x}) = f(\tilde{\boldsymbol{x}})$ holds. Thus, by the same argument as above, we have

$$\begin{aligned} & \frac{1}{N} \sum_{f \in C(n,k_1,\mathcal{X})} \sum_{g \in C(n,k_2,\mathcal{Y})} N_{f,g}(X\tilde{X}Y\tilde{Y}) \\ & \quad \leq\ u_1 |T_{XY}| \exp\{-n(r_1 - H(\tilde{X}|XY))\}. \end{aligned} \tag{23}$$

Analogous bound holds also for joint types $P_{X\tilde{X}Y\tilde{Y}}$ with $X = \tilde{X}$.

Using Markov's inequality, the above bounds imply that

$$\begin{aligned} &\tfrac{1}{N}|\{(f,g) \in C(n,k_1,\mathcal{X}) \times C(n,k_2,\mathcal{Y})\text{: } (f,g) \text{ fails to satisfy (20)} \\ &\quad \text{for some } P_{X\tilde{X}Y\tilde{Y}} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Y})\}| \\ &\leq \quad u_1 u_2 (n+1)^{|\mathcal{X}|^2|\mathcal{Y}|^2} \exp(-2n\epsilon) \leq \exp(-n\epsilon) \end{aligned}$$

for sufficiently large n. Therefore, at most $N\exp\{-n\epsilon\}$ pairs of (f,g) in $C(n,k_1,\mathcal{X})\times C(n,k_2,\mathcal{Y})$ fail to meet the requirement in the lemma for some $P_{X\tilde{X}Y\tilde{Y}} \in \mathcal{P}(\mathcal{X}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Y})$. This completes the proof. □

The next lemma is a direct application of Lemma 2.

Lemma 3: For any $\epsilon > 0$, any positive integers k_1, k_2 and sufficiently large integer n, there exist at least $N(1-\exp(-n\epsilon))$ pairs of linear encoders $(f,g) \in C(n,k_1,\mathcal{X})\times C(n,k_2,\mathcal{Y})$ such that for the pair (f,g) and the corresponding minimum entropy decoder φ, the probability of error p_e is bounded, universally for every Q as

$$p_e \leq \exp\{-n(E_r(r_1,r_2,Q)-\epsilon)\}, \tag{24}$$

where $E_r(r_1,r_2,Q)$ is defined by (16).

The proof of this lemma can be done essentially in the same manner as Theorem 1 in [9, pp.589-590] by using Lemma 2 instead of Csiszár's Lemma. So, we omit the proof.

Proof of Theorem 3: The proof of Theorem 3 can be done by using a technique developed by Delsarte and Piret for concatenated codes in [6, IV-c]. From Lemma

3, at least $N(1-\exp(-n\epsilon))$ pairs of encoders have the probability of error bounded by $\hat{p}_e \stackrel{\triangle}{=} \exp\{-n(E_r(r_1,r_2,Q)-\epsilon)\}$, and we call these encoders as *good encoders* and other encoders as *bad encoders*. Suppose that after the second step of decoding, we obtain the estimate $(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}) \in \mathcal{X}^{N_o} \times \mathcal{Y}^{N_o}$ of an encoded sequence $(\boldsymbol{x},\boldsymbol{y})$. Since the code $C_H(N,K_1,D_1)$ can correct at least $d = \lfloor (D_1-1)/2 \rfloor$ errors, the probability of erroneous estimation $\tilde{\boldsymbol{x}} \neq \boldsymbol{x}$ after the second step of decoding can be bounded by the probability that errors occur at least $d+1$ codes among (f_i,g_i,φ_i) $i=1,2,\cdots,N$ at the first step of decoding. Assuming that all bad encoders always give errors, we have

$$\begin{aligned}
\Pr\{\boldsymbol{x} \in \mathcal{X}^{N_o} : \tilde{\boldsymbol{x}} \neq \boldsymbol{x}\} &\leq \sum_{j=d+1-z}^{N-z} \binom{N-z}{j} (\hat{p}_e)^j (1-\hat{p}_e)^{N-z-j} \\
&\leq 2^N (\hat{p}_e)^{d-z} \\
&\leq 2^N \cdot \exp\{-n(d-z)(E_r(r_1,r_2,Q)-\epsilon)\}, \qquad (25)
\end{aligned}$$

where $z \stackrel{\triangle}{=} N\exp(-n\epsilon)$ denotes the maximum number of bad encoders. On the other hand, in a similar manner as (13), we have

$$\begin{aligned}
\frac{d-z}{N} &> \frac{1}{2}\left(\frac{\tilde{r}_1}{\log|\mathcal{X}|} - \frac{g(\ell_1,|\mathcal{X}|^{n/2})+1}{N}\right) - \exp(-n\epsilon) \\
&= \frac{1}{2}\left(\frac{\tilde{r}_1}{\log|\mathcal{X}|} - o(1)\right), \qquad (26)
\end{aligned}$$

where $o(1) \to 0$ as $n \to \infty$. Substituting (26) into (25), we immediately obtain

$$\Pr\{\boldsymbol{x} \in \mathcal{X}^{N_o} : \tilde{\boldsymbol{x}} \neq \boldsymbol{x}\} \leq \exp\left[-N_o\left\{\frac{\tilde{r}_1 E_r(r_1,r_2,Q)}{2\log|\mathcal{X}|} - \epsilon\right\}\right],$$

for sufficiently large block length N_o. In a similar manner, we also have

$$\Pr\{\boldsymbol{y} \in \mathcal{Y}^{N_o} : \tilde{\boldsymbol{y}} \neq \boldsymbol{y}\} \leq \exp\left[-N_o\left\{\frac{\tilde{r}_2 E_r(r_1,r_2,Q)}{2\log|\mathcal{Y}|} - \epsilon\right\}\right].$$

By combining the above two equations, we finally obtain

$$\begin{aligned} p_e &= \Pr\{(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^{N_o} \times \mathcal{Y}^{N_o} : (\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}) \neq (\boldsymbol{x},\boldsymbol{y})\} \\ &\leq \Pr\{\boldsymbol{x} \in \mathcal{X}^{N_o} : \tilde{\boldsymbol{x}} \neq \boldsymbol{x}\} + \Pr\{\boldsymbol{y} \in \mathcal{Y}^{N_o} : \tilde{\boldsymbol{y}} \neq \boldsymbol{y}\} \\ &\leq 2\exp\left[-N_o\left\{\frac{1}{2}\min\left(\frac{\tilde{r}_1}{\log|\mathcal{X}|}, \frac{\tilde{r}_2}{\log|\mathcal{Y}|}\right) E_r(r_1, r_2, Q) - \epsilon\right\}\right], \end{aligned}$$

which completes the proof. □

V. Conclusion

We proposed an explicit construction of fixed length codes for Slepian-Wolf source networks. The proposed code is linear and has two-step encoding and decoding procedures similar to the concatenated code used for channel coding. Further, if the sources are memoryless, the proposed code is universal and the probability of error vanishes exponentially as the block length tends to infinity. Regarding future research, we have the problem to obtain tight upper and lower bounds on the error exponent obtainable by the proposed code for DMS's.

Acknowledgment

The author wishes to thank Prof. Te Sun Han of University of Electro-Communications for his valuable discussions and comments.

References

[1] D. Slepian and J. K. Wolf: "Noiseless coding of correlated information sources," *IEEE Trans. on Inform. Theory,* vol.IT-19, pp.471-480, July 1973.

[2] T. M. Cover: "A proof of the data compression theorems of Slepian and Wolf for ergodic sources," *IEEE Trans. on Inform. Theory,* vol.IT-21, pp.226-228, March 1975.

[3] R. Ahlswede and J. Körner: "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. on Inform. Theory,* vol.IT-21, pp.629-637, Nov. 1975.

[4] I. Csiszár and J. Körner: *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic, 1981.

[5] R. Ahlswede and I. Csiszár: "Common randomness in information theory and cryptography — Part I: Secret sharing, " *IEEE Trans. on Inform. Theory,* vol.39, pp.1121-1132, July 1993.

[6] P. Delsarte and P. Piret: "Algebraic constructions of Shannon codes for regular channels," *IEEE Trans. on Inform. Theory,* vol.IT-28, pp.593-599, July 1982.

[7] J. Justesen : "A class of constructive asymptotically good algebraic codes," *IEEE Trans. on Inform. Theory,* vol.IT-18, pp.652-656, Sep. 1972.

[8] G. D. Forney, Jr. : *Concatenated Codes.* Cambridge, MA: MIT Press, 1966.

[9] I. Csiszár: "Linear codes for sources and source networks: error exponents, universal coding," *IEEE Trans. on Inform. Theory,* vol.IT-28, pp.585-592, July 1982.

[10] J. H. van Lint: *Introduction to Coding Theory,* 2nd. Ed., Springer-Verlag, 1991.

[11] V. D. Goppa : "Codes on algebraic curves," *Sov. Math.-Dokl.,* vol.24, pp.170-172, 1981.

[12] B.-Z. Shen : "A Justesen construction of binary concatenated codes that asymptotically meet the Zyablov bound for low rate," *IEEE Trans. on Inform. Theory,* vol.IT-39, pp.239-242, Jan. 1993.

[13] B.-Z. Shen and K. K. Tzeng : "A code decomposition approach for decoding cyclic and algebraic-geometric codes," *IEEE Trans. on Inform. Theory,* vol.IT-41, pp.1969-1987, Nov. 1995.

Biography of the author

Tomohiko Uyematsu (M'95) received the B. E., M. E. and D. E. degrees from Tokyo Institute of Technology in 1982, 1984 and 1988, respectively. From 1984 to 1992, He was with the Department of Electrical and Electronic Engineering of Tokyo Institute of Technology, first as research associate, next as lecturer, and lastly as associate professor. From 1992 to 1997, he was with School of Information Science of Japan Advanced Institute of Science and Technology as associate professor. Since 1997, he returned to the Department of Electrical and Electronic Engineering of Tokyo Institute of Technology as associate professor. In 1992 and 1996, he was a visiting researcher at the Centre National de la Recherche Scientifique, France, and Delft University of Technology, Delft, The Netherlands, respectively. His current research interests are in the areas of information theory, especially Shannon theory and multiterminal information theory.