A Dichotomy of Functions in Distributed Coding: An Information Spectral Approach

Shigeaki Kuzuoka Member, IEEE and Shun Watanabe, Member, IEEE

Abstract

The problem of distributed data compression for function computation is considered, where (i) the function to be computed is not necessarily symbol-wise function and (ii) the information source has memory and may not be stationary nor ergodic. We introduce the class of slowly varying sources and give a sufficient condition on functions so that the achievable rate region for computing coincides with the Slepian-Wolf region (i.e., the rate region for reproducing the entire source) for any slowly varying sources. Moreover, for symbol-wise functions, the necessary and sufficient condition for the coincidence is established. All results are given not only for fixed-length coding but also for variable-length coding in a unified manner. Furthermore, for the full side-information case, the error probability in the moderate deviation regime is also investigated.

Index Terms

distributed computing, information-spectrum method, Slepian-Wolf coding

I. INTRODUCTION

We study the problem of distributed data compression for function computation described in Fig. 1 and Fig. 2, where the function to be computed is not necessarily symbol-wise function. In [1], Körner and Marton revealed that the achievable rate region for computing modulo-sum is strictly larger than the rate region that can be achieved by first applying Slepian-Wolf coding [2] and then computing the function. In [3], the error region for the function calculation is established. Since then, distributed coding schemes that are tailored for some classes of functions were studied (e.g., see [4, Chapter 21]). These results are the cases such that the structure of functions can be utilized for distributed coding. However, not all functions have such nice structures, and even for some classes of functions, it is known that the Slepian-Wolf region cannot be improved at all [4, 5], i.e., reproducing function value is as difficult as reproducing the entire source. Thus, it is important to...
understand what makes distributed computation difficult, which is the main theme of this paper. This direction of research has been studied for i.i.d. sources, which will be reviewed next.

In [4], Ahlswede and Csiszár investigated distributed coding for function computation when the full side-information is available at the decoder (see Fig. 2); they introduced the concept of sensitive functions, and showed that the achievable rate for computing sensitive functions coincides with the achievable rate of Slepian-Wolf coding (with full side-information) provided that the source is an i.i.d. source satisfying the positivity condition. Surprisingly, the class of sensitive functions includes a function such that the image size is just one bit. Later, El Gamal gave a simple proof of Ahlswede and Csiszár’s result [6].

In [5], Han and Kobayashi investigated distributed coding for function computation with two-encoders case (see Fig. 1); they considered the class of symbol-wise functions, and derived the necessary and sufficient condition of functions such that the achievable rate region coincides with that of Slepian-Wolf coding for any i.i.d. sources satisfying the positivity condition. In the rest of the paper, we shall call functions satisfying the Han and Kobayashi’s condition HK functions.

For the class of i.i.d. sources satisfying the positivity condition, the above mentioned two results [4], [5] showed some classes of functions that are difficult to compute via distributed coding. Then, a natural question is:

(*) Are functions in those classes difficult to compute even for wider classes of sources that have memory and may not be stationary nor ergodic?

In order to answer this question in a unified manner, we study distributed computation problem by information-spectral approach [7]. Our contributions are summarized as follows.

A. Contributions

First, we introduce a class of sources which we called slowly varying sources; other than the slowly varying condition, we do not impose any condition on sources, i.e., we consider general sources. Roughly speaking, the slowly varying condition says that the probability of a sequence does not change significantly when we flip a symbol of the sequence. When we restrict sources to be i.i.d., then the slowly varying condition coincides with the positivity condition studied in [4], [5]. However, the class of slowly varying sources is much wider than the class of i.i.d. sources satisfying the positivity condition. In fact, it includes Markov sources with positive transition matrices or mixtures of i.i.d. sources satisfying positivity condition.

They also introduced the concept of highly sensitive functions and showed the same result under a slightly weaker condition on the source.
Next, we introduce the concept of *joint sensitivity*; a function \( f_n \) is said to be jointly sensitive if \( f_n(x, y) \neq f_n(\hat{x}, \hat{y}) \) whenever \( x \neq \hat{x} \) and \( y \neq \hat{y} \). Then, we introduce the class of *totally sensitive* functions as the set of all functions that are sensitive in the sense of [4] and also jointly sensitive. When we restrict functions to be symbol-wise, the class of totally sensitive functions is a strict subset of the class of HK functions. However, totally sensitive functions are not necessarily symbol-wise. The inclusive relation among the classes of functions is summarized in Fig. 3.

[Fig. 3. The inclusive relation among the classes of functions.]

When the full side-information is available at the decoder, we show that the Slepian-Wolf rate cannot be improved if the function is sensitive and the source is slowly varying. This result generalizes the result in [4] for slowly varying sources. Thus, for the class of sensitive functions, the answer to Question (♣) is positive in the sense that the Slepian-Wolf rate cannot be improved.

For the two-encoders case, we show that the Slepian-Wolf region cannot be improved if the function is totally sensitive and the source is slowly varying. Furthermore, for symbol-wise functions, we show that the achievable region coincides with the Slepian-Wolf region for any slowly varying sources if and only if the function is totally sensitive. In fact, for a function that satisfies HK condition but is not totally sensitive, there exists a finite state source, which is slowly varying, such that the Slepian-Wolf region can be improved. This dichotomy theorem can be regarded as a slowly varying source counterpart of Han and Kobayashi’s dichotomy theorem [5]; we need the condition that is more strict than HK condition because we broaden the class of sources. Consequently, for the class of HK functions, the answer to Question (♣) is negative in the sense that the Slepian-Wolf region can be improved; but we can say that totally sensitive functions are difficult to compute via distributed coding for any slowly varying sources.

We also derive the following refinements of the above results. So far, the study of distributed computing has been restricted to the fixed-length coding in the literatures [4], [5]. In this paper, by using the techniques developed...
by the authors in [8], we show that the above mentioned results also hold even for the variable-length coding. Furthermore, for the full side-information case, we show that the Slepian-Wolf rate cannot be improved even in the moderate deviation regime [9], [10].

Although our main contributions of this paper are structural connections between the achievable rate regions (or rates) for function computing and the Slepian-Wolf regions (or rates), as a byproduct, we can derive explicit forms of the achievable regions (or rates) by using the corresponding results on the Slepian-Wolf regions (or rates). It is also known that distributed computing can be regarded as a special case of distributed lossy coding studied by Yamamoto [11]. Thus, our results may be interesting from the view point of distributed lossy coding for slowly varying sources.

From technical perspective, we elaborate El Gamal’s argument [6] so that it can be used for the wider class of sources; Lemma 1 is the core of the proofs, and it enables us to prove our main results for both fixed-length coding and variable-length coding in a unified manner. The bounds in Lemma 1 is also tight enough to be used for the moderate deviation analysis.

B. Organization of Paper

In Section II, we introduce the coding problem investigated in this paper, and also introduce classes of functions and classes of sources. Then, in Section III main coding theorems are stated. The proofs of main results are given in Section IV, where proofs of some lemmas are shown in Appendices.

C. Notation

Throughout this paper, random variables (e.g., \(X\)) and their realizations (e.g., \(x\)) are denoted by capital and lower case letters respectively. All random variables take values in some finite alphabets which are denoted by the respective calligraphic letters (e.g., \(\mathcal{X}\)). Similarly, \(X^n \triangleq (X_1, X_2, \ldots, X_n)\) and \(x^n \triangleq (x_1, x_2, \ldots, x_n)\) denote, respectively, a random vector and its realization in the \(n\)th Cartesian product \(\mathcal{X}^n\) of \(\mathcal{X}\). We will use bold lower letters to represent vectors if the length \(n\) is apparent from the context; e.g., we use \(x\) instead of \(x^n\).

For a finite set \(S\), \(|S|\) denotes the cardinality of \(S\) and \(S^*\) denotes the set of all finite strings drawn from \(S\). For a sequence \(s \in S^*\), \(|s|\) denotes the length of \(s\). The Hamming distance between two sequences \(s, \hat{s} \in S^n\) is defined as \(d(s, \hat{s}) \triangleq |\{i : s_i \neq \hat{s}_i\}|\). \(S^c\) denotes the complement of \(S\).

Information-theoretic quantities are denoted in the usual manner [12], [13]. For example, \(H(X|Y)\) denotes the conditional entropy of \(X\) given \(Y\). All logarithms are with respect to base 2.

Moreover, we will use quantities defined by using the information-spectrum method [7]. Here, we recall the probabilistic limit operation: For a sequence \(Z \triangleq \{Z_n\}_{n=1}^{\infty}\) of real-valued random variables, the \(\text{limit superior in probability}\) of \(Z\) is defined as

\[
\limsup_{n \to \infty} Z_n \triangleq \inf \left\{ \alpha : \lim_{n \to \infty} \Pr\{Z_n > \alpha\} = 0 \right\}.
\]
II. PROBLEM

A. General Setting

Let \((X, Y) = \{(X^n, Y^n)\}_{n=1}^\infty\) be a general correlated source with finite alphabets \(X\) and \(Y\). We consider a sequence \(f = \{f_n\}_{n=1}^\infty\) of functions \(f_n : X^n \times Y^n \to \mathbb{Z}_n\). A variable-length code \(\Phi\) for computing \(f_n\) is defined by a triplet \((\varphi_n^{(1)}, \varphi_n^{(2)}, \psi_n)\) of the first encoder \(\varphi_n^{(1)} : X^n \to \{0, 1\}^*\), the second encoder \(\varphi_n^{(2)} : Y^n \to \{0, 1\}^*\), and a decoder \(\psi_n : C_n^{(1)} \times C_n^{(2)} \to \mathbb{Z}_n\), where \(C_n^{(1)} \triangleq \{\varphi_n^{(1)}(x) : x \in X^n\} \subseteq \{0, 1\}^*\) and \(C_n^{(2)} \triangleq \{\varphi_n^{(2)}(y) : y \in Y^n\} \subseteq \{0, 1\}^*\). We assume that both of \(C_n^{(1)}\) and \(C_n^{(2)}\) satisfy the prefix condition.

For each \(i = 1, 2\), \(\varphi_n^{(i)}\) is said to be a fixed-length encoder if \(C_n^{(i)}\) consists of codewords of the same length. A code \(\Phi_n\) is called a fixed-length code if both of \(\varphi_n^{(i)}\) \((i = 1, 2)\) are fixed-length encoders. Clearly, the class of all variable-length codes includes that of all fixed-length codes as a strict subclass.

The average codeword length and the error probability of \(\Phi_n\) are respectively defined as

\[
\mathbb{E} \left[ |\varphi_n^{(1)}(X^n)| \right] \triangleq \sum_x P_X(x) |\varphi_n^{(1)}(x)|, \tag{2}
\]

\[
\mathbb{E} \left[ |\varphi_n^{(2)}(Y^n)| \right] \triangleq \sum_y P_Y(y) |\varphi_n^{(2)}(y)|, \tag{3}
\]

and

\[
P_e(\Phi_n | f_n) \triangleq \text{Pr} \left\{ f_n(X^n, Y^n) \neq \psi_n \left( \varphi_n^{(1)}(X^n), \varphi_n^{(2)}(Y^n) \right) \right\}. \tag{4}
\]

**Definition 1.** Given a source \((X, Y)\) and a sequence of functions \(f = \{f_n\}_{n=1}^\infty\), a pair \((R_1, R_2)\) of rates is said to be achievable, if there exists a sequence \(\{\Phi_n\}_{n=1}^\infty\) of codes satisfying

\[
\lim_{n \to \infty} P_e(\Phi_n | f_n) = 0
\]

and

\[
\limsup_{n \to \infty} \frac{1}{n} \mathbb{E} \left[ |\varphi_n^{(1)}(X^n)| \right] \leq R_1, \tag{5}
\]

\[
\limsup_{n \to \infty} \frac{1}{n} \mathbb{E} \left[ |\varphi_n^{(2)}(Y^n)| \right] \leq R_2. \tag{6}
\]

The set of all achievable rate pairs is denoted by \(\mathcal{R}_v(X, Y | f)\).

**Definition 2.** Given a source \((X, Y)\) and a sequence of functions \(f = \{f_n\}_{n=1}^\infty\), a pair \((R_1, R_2)\) of rates is said to be achievable by fixed-length coding, if there exists a sequence \(\{\Phi_n\}_{n=1}^\infty\) of fixed-length codes satisfying \((5)\), \((6)\), and \((7)\). The set of all rate pairs that are achievable by fixed-length coding is denoted by \(\mathcal{R}_f(X, Y | f)\).

A variable-length (resp. fixed-length) code \(\Phi_n\) for computing the identity function \(f_n^{id}(x, y) \triangleq (x, y)\) is called a variable-length (resp. fixed-length) Slepian-Wolf (SW) code.

**Definition 3 (SW region).** For a source \((X, Y)\), the achievable rate region \(\mathcal{R}_v(X, Y | f^{id})\) for \((X, Y)\) and the sequence \(f^{id} \triangleq \{f_n^{id}\}_{n=1}^\infty\) of identity functions is called the Slepian-Wolf (SW) region and denoted by \(\mathcal{R}_v^{SW}(X, Y)\).

By considering only fixed-length codes, \(\mathcal{R}_v^{SW}(X, Y)\) is defined similarly.
Remark 1. From the definitions, it is apparent that $\mathcal{R}_{SW}^{vl}(X, Y) \subseteq \mathcal{R}^{vl}(X, Y|f)$ and $\mathcal{R}_{SW}^{fl}(X, Y) \subseteq \mathcal{R}^{fl}(X, Y|f)$ for any $(X, Y)$ and $f$.

Remark 2. A general formula for the SW region for fixed-length coding was given by Miyake and Kanaya [14] as

$$\mathcal{R}_{SW}^{fl}(X, Y) = \left\{ (R_1, R_2) : R_1 \geq \mathcal{T}(X|Y), R_2 \geq \mathcal{T}(Y|X), R_1 + R_2 \geq \mathcal{T}(X, Y) \right\}$$

(8)

where

$$\mathcal{T}(X, Y) \triangleq \limsup_{n \to \infty} \frac{1}{n} \log \frac{1}{P_{X^nY^n}(X^nY^n)},$$

(9)

$$\mathcal{T}(X|Y) \triangleq \limsup_{n \to \infty} \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(X^n|Y^n)},$$

(10)

$$\mathcal{T}(Y|X) \triangleq \limsup_{n \to \infty} \frac{1}{n} \log \frac{1}{P_{Y^n|X^n}(Y^n|X^n)}.$$  

(11)

As long as the authors know, a general formula for $\mathcal{R}_{SW}^{vl}(X, Y)$ is not known. One of our contributions is to demonstrate that we can discuss the equivalence between $\mathcal{R}_{SW}^{vl}(X, Y)$ and $\mathcal{R}^{vl}(X, Y|f)$ without knowing the precise form of $\mathcal{R}_{SW}^{vl}(X, Y)$; for specific sources such that the precise form of $\mathcal{R}_{SW}^{vl}(X, Y)$ is known, we can get the precise form of $\mathcal{R}^{vl}(X, Y|f)$ as a byproduct.

As a special case of distributed computation, we are interested in the case where $y \in Y^n$ is completely known at the decoder as the side-information. We call this case as the “full-side-information case”. The optimal coding rates which are achievable in full-side-information case are defined as follows.

Definition 4. For any $(X, Y)$ and $f$, let

$$\mathcal{R}^{vl}(X|Y|f) \triangleq \inf \left\{ R_1 : (R_1, \log |Y|) \in \mathcal{R}^{vl}(X, Y|f) \right\},$$

(12)

$$\mathcal{R}^{fl}(X|Y|f) \triangleq \inf \left\{ R_1 : (R_1, \log |Y|) \in \mathcal{R}^{fl}(X, Y|f) \right\}.$$  

(13)

Similarly, for any $(X, Y)$, let

$$\mathcal{R}_{SW}^{vl}(X|Y) \triangleq \inf \left\{ R_1 : (R_1, \log |Y|) \in \mathcal{R}_{SW}^{vl}(X, Y) \right\},$$

(14)

$$\mathcal{R}_{SW}^{fl}(X|Y) \triangleq \inf \left\{ R_1 : (R_1, \log |Y|) \in \mathcal{R}_{SW}^{fl}(X, Y) \right\}.$$  

(15)

Remark 3. From (8), we have $\mathcal{R}_{SW}^{fl}(X|Y) = \mathcal{T}(X|Y)$. A general formula for $\mathcal{R}_{SW}^{vl}(X|Y)$ is recently given by the authors [8].

B. Function Classes

In this subsection, we introduce important classes of functions investigated in this paper. First, we state the concept of sensitivity introduced in [4] and related properties.

August 27, 2014 DRAFT
Definition 5 (Sensitivity). A function \( f_n : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{Z}_n \) is said to be sensitive conditioned on \( \mathcal{Y}^n \) if it satisfies the following property: If \( x, \hat{x}, y \) satisfy \( f_n(x, y) = f_n(\hat{x}, y) \) and \( x_i \neq \hat{x}_i \) for some \( i \) then there exists \( \hat{y} \in \mathcal{Y}^n \) such that \( \hat{y}_i \neq y_i \), \( \hat{y}_j = y_j \) for any \( j \neq i \) and \( f_n(x, \hat{y}) \neq f_n(\hat{x}, \hat{y}) \).

Similarly, a function \( f_n : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{Z}_n \) is said to be sensitive conditioned on \( \mathcal{X}^n \) if it satisfies the property, where the role of \( x \) (resp. \( \hat{x} \)) in Definition 5 is switched with that of \( y \) (resp. \( \hat{y} \)).

Remark 4. In [6], the concept of \( \alpha \)-sensitive functions, which includes sensitive functions as a special case, is introduced, and it is shown that the result of [4], which is proved for sensitive functions, can be proved also for \( \alpha \)-sensitive functions. Although our results for sensitive functions hold also for \( \alpha \)-sensitive functions, we consider only sensitive functions for simplicity.

Definition 6 (Joint sensitivity). A function \( f_n : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{Z}_n \) is said to be jointly sensitive if \( f_n(x, y) \neq f_n(\hat{x}, \hat{y}) \) holds for every \( x \neq \hat{x} \) and \( y \neq \hat{y} \).

Definition 7 (Total sensitivity). A function \( f_n : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{Z}_n \) is said to be totally sensitive if it is jointly sensitive and sensitive conditioned on both of \( \mathcal{X}^n \) and \( \mathcal{Y}^n \).

Example 1. Let \( P_{xy} \) be the joint type of \( (x, y) \) [13]; i.e., \( P_{xy} \) is a joint distribution on \( \mathcal{X} \times \mathcal{Y} \) such as

\[
P_{xy}(a, b) \triangleq \frac{|\{i : (x_i, y_i) = (a, b)\}|}{n}, \quad (a, b) \in \mathcal{X} \times \mathcal{Y}.
\]

(16)

The type function \( f_n(x, y) \triangleq P_{xy} \) is sensitive conditioned on both of \( \mathcal{X}^n \) and \( \mathcal{Y}^n \) but is not jointly sensitive. Hence, it is not totally sensitive.

Example 2. The function defined by

\[
f_n(x, y) \triangleq \begin{cases} 
(>, x) & \text{if } x > y \\
(=, x) & \text{if } x = y \\
(<, y) & \text{if } x < y
\end{cases}
\]

(17)

where \( > \) and \( < \) are with respect to arbitrary ordering on \( \mathcal{X}^n = \mathcal{Y}^n \) is jointly sensitive but is not sensitive conditioned on \( \mathcal{X}^n \) (nor \( \mathcal{Y}^n \)). On the other hand,

\[
f_n'(x, y) \triangleq (P_{xy}, f_n(x, y))
\]

(18)

is totally sensitive.

Next, we consider special classes of symbol-wise functions. Given a function \( f \) on \( \mathcal{X} \times \mathcal{Y} \), the function \( f_n \) on \( \mathcal{X}^n \times \mathcal{Y}^n \) defined as \( f_n(x, y) \triangleq (f(x_1, y_1), f(x_2, y_2), \ldots, f(x_n, y_n)) \) is called the symbol-wise function defined by \( f \). Now, we introduce a special class of symbol-wise functions defined by Han and Kobayashi [5].

Definition 8 (HK functions). A function \( f_n \) is called a Han-Kobayashi (HK) function if \( f_n \) is a symbol-wise function defined by some \( f \) such that
1) for every \( a_1 \neq a_2 \) in \( X \), the functions \( f(a_1, \cdot) \) and \( f(a_2, \cdot) \) are distinct,
2) for every \( b_1 \neq b_2 \) in \( Y \), the functions \( f(\cdot, b_1) \) and \( f(\cdot, b_2) \) are distinct, and
3) \( f(a_1, b_1) \neq f(a_2, b_2) \) for every \( a_1 \neq a_2 \) and \( b_1 \neq b_2 \).

By definitions, it is easy to see that (i) an HK function is sensitive conditioned on both of \( X^n \) and \( Y^n \), but (ii) there exists an HK function which is not jointly sensitive (and thus not totally sensitive). On the other hand, it is necessary for a totally sensitive function be an HK function. Indeed, the next proposition gives the sufficient and necessary condition for symbol-wise functions to be totally sensitive. The proof of Proposition 1 is given in Appendix A.

**Proposition 1.** Let \( f \) be given and \( f_n \) be the symbol-wise function defined by \( f \). Then \( f_n \) \((n \geq 2)\) is totally sensitive if and only if \( f \) is an HK function satisfying at least one of the following two properties:

1) for all \( x \in X \), if \( f(x, y) = f(x, \hat{y}) \) then \( y = \hat{y} \), or
2) for all \( y \in Y \), if \( f(x, y) = f(\hat{x}, y) \) then \( x = \hat{x} \).

**Example 3.** The function shown in Table I is an HK function, but it does not satisfy 1) nor 2) of Proposition 1. Thus, any \( f_n \) defined by \( f \) is not jointly sensitive nor totally sensitive. Indeed, let \( x^2 = (0, 1) \), \( y^n = (0, 1) \), \( \hat{x}^2 = (1, 1) \), and \( \hat{y}^2 = (0, 2) \), then we have \( f_2(x^2, y^2) = f_2(\hat{x}^2, \hat{y}^2) = (0, 3) \) even though \( x^2 \neq \hat{x}^2 \) and \( y^2 \neq \hat{y}^2 \). The function shown in Table II (resp. Table III) is an HK function and satisfies 1) (resp. 2)) of Proposition 1. Hence, the symbol-wise function \( f_n \) defined by \( f \) in Tables II or III is totally sensitive.

**Remark 5.** In this subsection, several properties of functions on \( X^n \times Y^n \) are introduced. In the following, we say a sequence \( f = \{f_n\}^\infty_{n=1} \) of functions satisfies some property, if \( f_n \) satisfies that property for all \( n = 1, 2, \ldots \); e.g., we say “\( f \) is totally sensitive” meaning “\( f_n \) is totally sensitive for all \( n = 1, 2, \ldots \)”.  

**C. Classes of General Sources**

In this subsection, we introduce the concept of slowly varying sources.

**Definition 9.** A general source \((X, Y)\) is said to be *slowly varying* with respect to \( Y \) if there exists a constant \( 0 < q < 1 \), which does not depend on \( n \), satisfying

\[
P_{X^n \times Y^n}(x, \hat{y}) \geq qP_{X^n \times Y^n}(x, y)
\]

\((19)\)
for any $x \in \mathcal{X}^n$ and any $y, \hat{y}$ such that $d(y, \hat{y}) = 1$.

The definition implies that, for a slowly varying source with respect to $Y$, the probability of joint sequences $(x, y)$ does not drastically change even if a symbol of $y$ is replaced with another symbol. The constant $q$ is called as the varying constant.

**Example 4 (General Source with Positive Side-Information Channel).** If $Q(y|x) > q$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and

$$P_{X^nY^n}(x, y) = P_{X^n}(x) \prod_{i=1}^n Q(y_i|x_i)$$

(20)

then $(X, Y)$ is slowly varying with respect to $Y$.

Similarly, a source is said to be slowly varying with respect to $X$ if it satisfies the property, where the role of $x$ in Definition 9 is switched with that of $y$. If a source is slowly varying with respect to both $X$ and $Y$ then we just call it a slowly varying source.

As shown in the following proposition, the slowly vanishing property is identical with the positivity condition when we consider only i.i.d. sources.

**Proposition 2.** Let $(X, Y)$ be an i.i.d. source with the joint distribution $P_{X_1Y_1} = P_{XY}$. Then, $(X, Y)$ is slowly varying if and only if $P_{XY}$ satisfies the positivity condition $P_{XY}(a, b) > 0 ((a, b) \in \mathcal{X} \times \mathcal{Y})$.

On the other hand, as shown in following examples, the class of slowly varying sources includes not only i.i.d. sources but also Markov sources and mixed sources.

**Example 5 (Markov Source).** Let $(X, Y)$ be the source induced by a positive transition matrix $W(x, y|x', \hat{y})$ and a positive initial distribution $P_{X_1Y_1}(x, y)$. Then, by setting

$$q_1 \triangleq \min_{(x_1, y_1), (x_2, y_2), (x_3, y_3)} W(x_3, y_3|x_2, y_2)W(x_2, y_2|x_1, y_1),$$

(21)

$$q_2 \triangleq \min_{(x_1, y_1), (x_2, y_2)} W(x_2, y_2|x_1, y_1)P_{X_1Y_1}(x_1, y_1),$$

(22)

we can find that $(X, Y)$ is a slowly varying source with varying constant $q \triangleq \min\{q_1, q_2\}$.

**Example 6 (Mixed Source).** Let $(X_i, Y_i)$ be a slowly varying source with the varying constant $q_i$ $(i = 1, 2, \ldots, k)$ and consider a mixture $(X, Y)$ of them such that

$$P_{X^nY^n}(x, y) = \sum_{i=1}^k \alpha_i P_{X^n_iY^n_i}(x, y), \quad (x, y) \in \mathcal{X}^n \times \mathcal{Y}^n$$

(23)

where $\alpha_i > 0$ for all $i = 1, \ldots, k$ and $\sum_i \alpha_i = 1$. Then, $(X, Y)$ is also a slowly varying source with the varying constant $q \triangleq \min q_i$. 

August 27, 2014 DRAFT
III. CODING THEOREMS

A. Two Encoders Case

Our first result shows that, given a code \( \{ \Phi_n \}_{n=1}^{\infty} \) for computing a totally sensitive function \( f \), we can construct a SW code \( \{ \hat{\Phi}_n \}_{n=1}^{\infty} \) such that the coding rates of \( \{ \Phi_n \}_{n=1}^{\infty} \) are asymptotically same as \( \{ \hat{\Phi}_n \}_{n=1}^{\infty} \) and the error probability of \( \{ \hat{\Phi}_n \}_{n=1}^{\infty} \) is vanishing as \( n \to \infty \), provided that \( (X, Y) \) is slowly varying.

**Theorem 1.** Suppose that \( (X, Y) \) is slowly varying and \( f \) is totally sensitive. Then, for any variable-length (resp. fixed-length) code \( \{ \Phi_n \}_{n=1}^{\infty} \) for computing \( f \) satisfying (5)–(7), there exists a variable-length (resp. fixed-length) SW code \( \{ \hat{\Phi}_n \}_{n=1}^{\infty} = \{ (\hat{\varphi}_n^{(1)}, \hat{\varphi}_n^{(2)}, \hat{\psi}_n) \}_{n=1}^{\infty} \) such that

\[
\lim_{n \to \infty} P_e(\hat{\Phi}_n/f_n) = 0
\]

and

\[
\limsup_{n \to \infty} \frac{1}{n} \mathbb{E} \left[ |\hat{\varphi}_n^{(1)}(X^n)| \right] \leq R_1, \\
\limsup_{n \to \infty} \frac{1}{n} \mathbb{E} \left[ |\hat{\varphi}_n^{(2)}(Y^n)| \right] \leq R_2.
\]

The proof will be given in the next section. As a consequence of Theorem 1, we have the following theorem, which shows that the achievable rate region for a slowly varying source \((X, Y)\) and a totally sensitive function \( f \) is identical with the SW region.

**Theorem 2.** Suppose that \( (X, Y) \) is slowly varying and \( f \) is totally sensitive. Then we have

\[
\mathcal{R}^l(X, Y|f) = \mathcal{R}^l_{SW}(X, Y)
\]

and

\[
\mathcal{R}^vl(X, Y|f) = \mathcal{R}^vl_{SW}(X, Y).
\]

It should be noted that total sensitivity may not be necessary for \( \mathcal{R}^{**}(X, Y|f) = \mathcal{R}^{**}_{SW}(X, Y) \) (\( ** = fl/vl \)). Indeed a close inspection of the proof of Theorem 2 reveals that it is sufficient that if \( f \) satisfies

\[
\lim_{n \to \infty} \frac{1}{n} \log \max_{z_n \in \mathcal{Z}_n} \text{Equiv}(z_n|f_n) = 0,
\]

where \( \text{Equiv}(z_n|f_n) \) denotes the maximum number \( J \) such that we can choose \( J \) pairs \((x_1, y_1), (x_2, y_2), \ldots, (x_J, y_J) \in \mathcal{X}^n \times \mathcal{Y}^n \) such that \( x_i \neq x_j \) and \( y_i \neq y_j \) for all \( i \neq j \) and \( z_n = f_n(x_1, y_1) = f_n(x_2, y_2) = \cdots = f_n(x_J, y_J) \). Note that total sensitivity requires \( \text{Equiv}(z_n|f_n) \leq 1 \).

Theorem 2 states that the total sensitivity is a sufficient condition for the set of all achievable rates to coincide with the SW region. If we restrict our attention to the class of symbol-wise functions, we can also prove the converse statement, i.e., the total sensitivity is the necessary and sufficient condition for the set of all achievable rates to coincide with the SW region. More precisely, we have the following theorem.
**Theorem 3.** Let \( f \) be a sequence of symbol-wise functions. Then \( R^f(X,Y|f) = R^f_{SW}(X,Y) \) for all slowly varying sources \((X,Y)\) if and only if \( f \) is totally sensitive.

Now, let us compare our result with that of Han and Kobayashi [5].

**Proposition 3 (Theorem 1 of [5]).** Let \( f \) be a sequence of symbol-wise functions. Then \( R^f(X,Y|f) = R^f_{SW}(X,Y) \) for all i.i.d. sources \((X,Y)\) satisfying the positivity condition \( P_{X,Y}(x,y) > 0 \) if and only if \( f \) is a HK function.

Comparison of Theorem 3 with Proposition 3 implies that the condition given by Han and Kobayashi [5] is no longer sufficient for \( R^f(X,Y|f) = R^f_{SW}(X,Y) \), when we consider not only i.i.d. sources but also sources with memory.

Further, we can generalize the result for the variable-length coding case.

**Theorem 4.** Let \( f \) be a sequence of symbol-wise functions. Then \( R^{vl}(X,Y|f) = R^{vl}_{SW}(X,Y) \) for all slowly varying sources \((X,Y)\) if and only if \( f \) is totally sensitive.

### B. Full-Side-Information Case

Theorem 2 assumes the slowly varying property of the source and the total sensitivity of functions. In the full-side-information case, weaker conditions are sufficient to show the corresponding result. Indeed we have the following theorem.

**Theorem 5.** Suppose that \((X,Y)\) is slowly varying with respect to \( Y \) and \( f \) is sensitive conditioned on \( Y^n \). Then we have

\[
R^f(X|Y|f) = R^f_{SW}(X|Y).
\]  

As a corollary of the theorem, we can derive the first half of Theorem 3 of [4].

**Corollary 1 ([4]).** Suppose that \((X,Y)\) is an i.i.d. source satisfying the positivity condition \( P_{X,Y}(x,y) > 0 \) and \( f \) is sensitive conditioned on \( Y^n \). Then we have

\[
R^f(X|Y|f) = R^f_{SW}(X|Y).
\]  

**Remark 6.** We can also derive Lemmas 1 and 2 of [5] by applying Theorem 5 (or Corollary 1) to symbol-wise functions.

**Remark 7.** In the second half of Theorem 3 of [4], it is shown that if \( f \) is highly sensitive then Corollary 1 holds even under the weaker condition. Similarly, we can prove that if \( f \) is highly sensitive then Theorem 5 holds

\[\text{Note that neither Theorem 3 nor Proposition 3 subsumes the other.}\]
even under the condition weaker than the slowly varying property, and thus, we can derive also the second half of Theorem 3 of [4] as a corollary. See Section III-C for more details.

Further, we can generalize the result for the variable-length coding case.

**Theorem 6.** Suppose that \((X, Y)\) is slowly varying with respect to \(Y\) and \(f\) is sensitive conditioned on \(Y^n\). Then we have
\[
R_{vl}^{el}(X \mid Y \mid f) = R_{Sw}^{el}(X \mid Y).
\]

**C. Weaker Condition on Sources**

So far, we consider only slowly varying sources for simplicity. In this subsection, we show that all our results in Sections III-A and III-B are true even under weaker condition than the slowly varying, provided that the function \(f\) is highly sensitive in the sense of [4].

**Definition 10.** A function \(f_n: \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{Z}^n\) is said to be highly sensitive conditioned on \(Y^n\) if for any \(a_1 \neq a_2\) in \(\mathcal{X}\) and \(b_1 \neq b_2\) in \(\mathcal{Y}\) the following property holds: If \(x, \hat{x}, y\) satisfy \(f_n(x, y) = f_n(\hat{x}, y)\), \(x_i = a_1\), \(\hat{x}_i = a_2\), and \(y_i = b_1\) for some \(i\) then for \(\hat{y} \in \mathcal{Y}^n\) obtained from \(y\) by replacing the \(i\)th component by \(b_2\) we always have \(f_n(x, \hat{y}) \neq f_n(\hat{x}, \hat{y})\).

Similarly, the concept of “the highly sensitivity conditioned on \(X^n\)” is defined. Further, by replacing the sensitivity with the highly sensitivity in Definition 7, the highly total sensitivity is defined.

Now, we define a class of sources which is wider than the class of slowly varying sources.

**Definition 11.** A general source \((X, Y)\) is said to be weakly slowly varying with respect to \(Y\) if there exists a constant \(0 < q < 1\), which does not depend on \(n\), satisfying the following property: For any \(x \neq \hat{x}\) and \(y\) satisfying \(P_{X^nY^n}(x, y) \cdot P_{X^nY^n}(\hat{x}, y) > 0\), whenever \(x_i \neq \hat{x}_i\), there exists \(\hat{y} \in \mathcal{Y}^n\) such that \(\hat{y}_i \neq y_i\), \(\hat{y}_j = y_j\) for any \(j \neq i\) and
\[
\begin{align*}
P_{X^nY^n}(x, \hat{y}) & \geq qP_{X^nY^n}(x, y), \\
P_{X^nY^n}(\hat{x}, \hat{y}) & \geq qP_{X^nY^n}(\hat{x}, y).
\end{align*}
\]

Then, we can modify theorems in Sections III-A and III-B as in the following theorem.

**Theorem 7.** Theorems 1, 2, 3, and 4 hold even when we replace “slowly varying” (resp. “totally sensitive”) with “weakly slowly varying” (resp. “highly totally sensitive”). Further, Theorems 5 and 6 hold even when we replace “slowly varying” (resp. “sensitive”) with “weakly slowly varying” (resp. “highly sensitive”).

Especially, as mentioned in Remark 7, the second half of Theorem 3 of [4] can be derived as a corollary of the above theorem, since the following proposition holds. The proof of Proposition 4 is given in Appendix B.
Proposition 4. Let $(X, Y)$ be an i.i.d. source with the joint distribution $P_{X, Y_1} = P_{X, Y}$. Then, $(X, Y)$ is weakly slowly varying if and only if $P_{X, Y}$ satisfies the condition that for every $a_1 \neq a_2$ in $\mathcal{X}$ the number of elements $b \in \mathcal{Y}$ with

$$P_{X, Y}(a_1, b) \cdot P_{X, Y}(a_2, b) > 0$$

is different from one.

D. Moderate Deviation

In this subsection, we assume that $(X, Y)$ is an i.i.d. source with the joint distribution $P_{X, Y_1} = P_{X, Y}$, and we consider the full side-information case. The results in Section III-B states that $R^{**}(X|Y|f) = R^{**}_{SW}(X|Y) = H(X|Y)$ ($** = fl/vl$). In the following, we conduct more refined analysis in the moderate deviation regime.

For real numbers $t \in (0, 1/2)$ and $\gamma > 0$, and a sequence of functions $f = \{f_n\}_{n=1}^\infty$, let

$$e^{vl}(t, \gamma|f) \triangleq \liminf_{n \to \infty} -\frac{1}{n^{1-2t}} \log \min_{\Phi_n} P_e(\Phi_n|f_n)$$

(36)

where the minimum is taken over all sequences of codes $\{\Phi_n\}_{n=1}^\infty = \{(\varphi_n^{(1)}, \psi_n)\}_{n=1}^\infty$ for computing $f = \{f_n\}_{n=1}^\infty$ satisfying

$$\limsup_{n \to \infty} \frac{1}{n^{1-t}} \left( \mathbb{E} \left[ \varphi_n^{(1)}(X^n) \right] - nH(X|Y) \right) \leq \gamma.$$  

(37)

Similarly, by taking the minimum over all fixed-length codes, $e^{fl}(t, \gamma|f)$ is defined. Further, by considering the identity function $f_n^{id}$ and SW codes, $e_{SW}^{vl}(t, \gamma)$ and $e_{SW}^{fl}(t, \gamma)$ are defined. The single-letter characterization of $e_{SW}^{vl}(t, \gamma)$ and $e_{SW}^{fl}(t, \gamma)$ are obtained by He et al. [9]. The following theorem states that computing sensitive function is as difficult as reproducing $X$ itself even for the moderate deviation regime.

Theorem 8. Suppose that $P_{X, Y}$ satisfies positivity condition and $f$ is sensitive. Then, we have

$$e^{vl}(t, \gamma|f) = e_{SW}^{vl}(t, \gamma),$$

(38)

$$e^{fl}(t, \gamma|f) = e_{SW}^{fl}(t, \gamma)$$

(39)

for every $t \in (0, 1/2)$ and $\gamma > 0$.

IV. PROOF OF THEOREMS

A. Preliminaries for Proofs

First we introduce some notations used in this section. $1$ denotes the indicator function, e.g., $1[s \in \mathcal{S}] = 1$ if $s \in \mathcal{S}$ and 0 otherwise; $h(p)$ is binary entropy function; $[a]^+$ is 0 if $a < 1$ and $|a|$ if $a \geq 1$. For given $0 < \beta < 1/2$, August 27, 2014 DRAFT
let
\[
\begin{align*}
v_n(\beta) & \triangleq \sum_{i=1}^{\lceil n\beta \rceil - 1} (|X| - 1)^i \binom{n}{i} \\
& \leq n|X|^n 2^{nh(\beta)},
\end{align*}
\]
and
\[
\begin{align*}
u_n(\beta) & \triangleq \sum_{i=1}^{\lceil n\beta \rceil - 1} (|Y| - 1)^i \binom{n}{i} \\
& \leq n|Y|^n 2^{nh(\beta)}.
\end{align*}
\]
(40)

For a given code \(\Phi_n = (\phi_n^{(1)}, \phi_n^{(2)}, \psi_n)\), we abbreviate the length of codewords by
\[
\ell_n^{(1)}(x) \triangleq |\phi_n^{(1)}(x)|,
\]
(44)
\[
\ell_n^{(2)}(x) \triangleq |\phi_n^{(2)}(x)|.
\]
(45)

Without loss of generality, we assume that there are \(L_1 < \infty\) and \(L_2 < \infty\) such that
\[
\begin{align*}
\ell_n^{(1)}(x) & \leq nL_1, \quad \text{for all } x \in X^n, \\
\ell_n^{(2)}(x) & \leq nL_2, \quad \text{for all } y \in Y^n.
\end{align*}
\]
(46)

Let
\[
D_n \triangleq \left\{ (x, y) : \psi_n(\phi_n^{(1)}(x), \phi_n^{(2)}(y)) = f_n(x, y) \right\}
\]
be the set of all correctly decodable sequences. When we analyze the performance of a variable length code via information spectrum approach, the following typical-like sets play an important role:
\[
\begin{align*}
T_{n,1} & \triangleq \left\{ (x, y) : \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(x|y)} \leq \frac{\ell_n^{(1)}(x)}{n} + \delta \right\}, \\
T_{n,2} & \triangleq \left\{ (x, y) : \frac{1}{n} \log \frac{1}{P_{Y^n|X^n}(y|x)} \leq \frac{\ell_n^{(2)}(y)}{n} + \delta \right\}, \\
T_{n,0} & \triangleq \left\{ (x, y) : \frac{1}{n} \log \frac{1}{P_{X^nY^n}(x, y)} \leq \frac{\ell_n^{(1)}(x)}{n} + \frac{\ell_n^{(2)}(y)}{n} + \delta \right\},
\end{align*}
\]
(49)
(50)
(51)
where \(\delta > 0\) is any real number specified later. The following lemma is the core of the proofs of coding theorems, which connect the combinatorial property, i.e., the sensitivity of a function, to a probabilistic analysis. The proof of Lemma will be given in Appendix C.

**Lemma 1.** For any code \(\Phi_n\) and real numbers \(\beta, \delta > 0\), if \(f_n\) is sensitive conditioned on \(Y^n\) and \((X, Y)\) is slowly varying with respect to \(Y\), then we have
\[
P_{X^nY^n}(D_n \cap \overline{T_{n,1}}) \leq \frac{2|Y|}{\beta q} P_e(\Phi_n | f_n) + (v_n(\beta) + 1)2^{-n\delta}.
\]
(52)

\(^5\)Note that for any encoder \(\phi_n^{(1)}\), we can modify \(\phi_n^{(1)}\) without increasing the error probability and obtain an encoder \(\tilde{\phi}_n^{(1)}\) satisfying \(|\tilde{\phi}_n^{(1)}(x)| \leq \max \left[ 1 + |\phi_n^{(1)}(x)|, 1 + n|\log |X|\rceil \right].\)
Similarly, if \( f_n \) is sensitive conditioned on \( X^n \) and \((X, Y)\) is slowly varying with respect to \( X \), then we have
\[
P_{X^n | Y^n} (D_n \cap T_{n,2}^c) \leq \frac{2(|X| + |Y|)}{\beta q} P_n (f_n) \leq \beta (u_n(\beta) + 1)2^{-\delta}.
\] (53)

Furthermore, if \( f_n \) is totally sensitive and \((X, Y)\) is slowly varying, then we have
\[
P_{X^n | Y^n} (D_n \cap T_{n,0}^c) \leq \frac{2(|X| + |Y|)}{\beta q} P_n (f_n) \leq \beta (u_n(\beta) + 1)2^{-\delta}.
\] (54)

The following lemma is an immediate consequence of Lemma [1]

**Lemma 2.** For any \( \delta > 0 \) and any code satisfying [5], if \( f \) is totally sensitive, we have
\[
\lim_{n \to \infty} P_{X^n | Y^n} (T_{n,1}^c \cup T_{n,2}^c \cup T_{n,0}^c) = 0.
\] (55)

**Proof:** We have
\[
P_{X^n | Y^n} (T_{n,1}^c \cup T_{n,2}^c \cup T_{n,0}^c)
\]
(56)
\[
= P_{X^n | Y^n} (D_n \cap (T_{n,1}^c \cup T_{n,2}^c \cup T_{n,0}^c)) + P_{X^n | Y^n} (D_n \cap (T_{n,1}^c \cup T_{n,2}^c \cup T_{n,0}^c))
\]
(57)
\[
\leq P_n (f_n) + P_{X^n | Y^n} (D_n \cap T_{n,1}^c) + P_{X^n | Y^n} (D_n \cap T_{n,2}^c) + P_{X^n | Y^n} (D_n \cap T_{n,0}^c).
\] (58)

Then, we apply Lemma [1] by taking sufficiently small \( \beta > 0 \) so that \( v_n(\beta)2^{-\delta} \) and \( u_n(\beta)2^{-\delta} \) converges to 0 as \( n \to \infty \).

\[\square\]

**B. Proof of Theorem 1**

For a given (variable-length) code \( \Phi_n \) for function computation, we construct a SW code by using a random binning of adaptive length.\footnote{We only show the statement for variable-length coding since the statement for fixed-length coding can be proved as a special case of the former.}

Let
\[
\tilde{\ell}_n^{(1)}(x) \triangleq \lceil \ell_n^{(1)}(x) + 2n\delta \rceil,
\] (61)
\[
\tilde{\ell}_n^{(2)}(x) \triangleq \lceil \ell_n^{(2)}(x) + 2n\delta \rceil,
\] (62)

and for each integer \( l \), let
\[
S_{n,1}(l) \triangleq \{ (x, y) : \frac{1}{n} \log \frac{1}{P_{X^n | Y^n}(x|y)} \leq \frac{l}{n} - \delta \},
\] (63)
\[
S_{n,2}(l) \triangleq \{ (x, y) : \frac{1}{n} \log \frac{1}{P_{Y^n | X^n}(y|x)} \leq \frac{l}{n} - \delta \},
\] (64)
\[
S_{n,0}(l) \triangleq \{ (x, y) : \frac{1}{n} \log \frac{1}{P_{X^n | Y^n}(x, y)} \leq \frac{l}{n} - \delta \}.
\] (65)

Further, for integers \( l_1 \) and \( l_2 \), let
\[
S_n(l_1, l_2) \triangleq S_{n,1}(l_1) \cap S_{n,2}(l_2) \cap S_{n,0}(l_1 + l_2).
\] (66)

\[\square\]

August 27, 2014 DRAFT
Note that, for any $y \in \mathcal{Y}^n$, we have
\begin{equation}
|\{\hat{x} : (\hat{x}, y) \in S_n(l_1, l_2)\}| \leq |\{\hat{x} : (\hat{x}, y) \in S_{n,1}(l_1)\}| \leq 2^{l_1 - n\delta}.
\end{equation}

Similarly, for any $x \in \mathcal{X}^n$, we have
\begin{equation}
|\{\hat{y} : (x, \hat{y}) \in S_n(l_1, l_2)\}| \leq 2^{l_2 - n\delta}.
\end{equation}

and
\begin{equation}
|\{(\hat{x}, \hat{y}) : (\hat{x}, \hat{y}) \in S_n(l_1, l_2)\}| \leq 2^{l_1 + l_2 - n\delta}.
\end{equation}

Now, we construct a SW code as follows:

- Given $x \in \mathcal{X}^n$, the encoder 1
  1) sends the integer $l_1 = \ell_n^{(1)}(x)$ by using at most $2(\log \ell_n^{(1)}(x) + 1)$ bits \cite{15}, and then
  2) by using a random bin-code with $\ell_n^{(1)}(x)$ bits, sends the bin-index $m_1$ of $x$.

- Given $y \in \mathcal{Y}^n$, the encoder 2
  1) sends the integer $l_2 = \ell_n^{(2)}(y)$ by using at most $2(\log \ell_n^{(1)}(x) + 1)$ bits \cite{15}, and then
  2) by using a random bin-code with $\ell_n^{(2)}(y)$ bits, sends the bin-index $m_2$ of $y$.

- The decoder
  1) extracts $l_1, l_2, m_1$, and $m_2$ from the received codewords, and then
  2) looks for the unique pair $(x, y)$ such that $(x, y) \in S_n(l_1, l_2)$, $\ell_n^{(1)}(x) = l_1$, $\ell_n^{(2)}(y) = l_2$, and the bin-index of $x$ (resp. $y$) is $m_1$ (resp. $m_2$).

By using the standard argument, we can upper bound the average error probability $\mathbb{E}[P_e(\hat{\Phi}_n|f_{ad}^n)]$ of the constructed SW code with respect to random bin-coding by
\begin{equation}
\mathbb{E}[P_e(\hat{\Phi}_n|f_{ad}^n)] \leq \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(X^n|Y^n)} > \frac{\tilde{\ell}_n^{(1)}(X^n)}{n} - \delta \right. \text{ or } \frac{1}{n} \log \frac{1}{P_{Y^n|X^n}(Y^n|X^n)} > \frac{\tilde{\ell}_n^{(2)}(Y^n)}{n} - \delta \right. \text{ or }
\frac{1}{n} \log \frac{1}{P_{X^n|Y^n}(X^n, Y^n)} > \frac{\tilde{\ell}_n^{(1)}(X^n) + \tilde{\ell}_n^{(2)}(Y^n)}{n} - \delta \right\}
\end{equation}
\begin{equation}
+ \sum_{x, y} P_{X^nY^n}(x, y) \left\{ \frac{1}{2^{\tilde{\ell}_n^{(1)}(x)}} \right\}
\end{equation}
\begin{equation}
+ \sum_{x, y} P_{X^nY^n}(x, y) \left\{ \frac{1}{2^{\tilde{\ell}_n^{(2)}(y)}} \right\}
\end{equation}
\begin{equation}
+ \sum_{x, y} P_{X^nY^n}(x, y) \left\{ \frac{1}{2^{\tilde{\ell}_n^{(1)}(x) + \tilde{\ell}_n^{(2)}(y)}} \right\}
\end{equation}
Hence, by Lemma 2, we have

\[ P \text{D. Proofs of Theorem 5 and Theorem 6} \]

\[ C. \text{Proof of Theorem 3 and Theorem 4} \]

August 27, 2014 DRAFT

\[ \text{a given variable-length sources such that} \]

\[ f \text{g} \]

\[ \text{coding, the SW region for fixed length coding is identical with that of variable length coding if the source is} \]

\[ x \]

\[ \text{In other words, let us consider a class of} \]

\[ \text{an i.i.d. source} \]

\[ \text{Since we can take} \]

\[ \phi \]

\[ n \]

\[ \text{On the other hand, the codeword length of the encoder} \]

\[ \hat{\phi}_n \]

\[ \text{of the constructed SW code satisfies that} \]

\[ \left| \hat{\phi}_n^{(1)}(x) \right| \leq \ell_n^{(1)}(x) + 2(\lfloor \log \ell_n^{(1)}(x) \rfloor + 1) \]

\[ \leq \ell_n^{(1)}(x) + 2n\delta + 2\log(\ell_n^{(1)}(x) + 2n\delta) + 3 \]

\[ \leq \ell_n^{(1)}(x) + 2n\delta + 2\log(nL_1 + 2n\delta) + 3. \]

Since we can take \( \delta > 0 \) arbitrarily small, \( \hat{\phi}_n^{(1)} \) satisfies (25). Similarly, we can prove (26).

\[ \Box \]

\[ C. \text{Proof of Theorem 2 and Theorem 4} \]

\[ \text{Since “if” part is obvious from Theorem 2, we only prove “only if” part. Let us consider a class of finite-state} \]

\[ \text{sources that} \]

\[ P_{X^{2n}Y^{2n}}(x,y) = \prod_{i=1}^{n} P_{X_{2i-1}X_{2i}Y_{2i-1}Y_{2i}}. \]

In other words, let us consider a class of two-symbol-wise i.i.d. sources. Note that such a source \((X, Y)\) includes an i.i.d. source \((U, V)\) with alphabets \(U = X^2\) and \(V = Y^2\).

Assume that \( f \) is symbol-wise but not jointly sensitive. Then, as shown in the proof of Proposition 1 there exists \( x^2 = (a_0, a_1), \hat{x}^2 = (a_0, a_2), y^2 = (b_1, b_0), \) and \( \hat{y}^2 = (b_2, b_0) \) such that \( x^2 \neq \hat{x}^2, y^2 \neq \hat{y}^2, \) and \( f_2(x^2, y^2) = f_2(\hat{x}^2, \hat{y}^2). \) Note that \( f \) induces a function \( g \) on \( U \times V \) which is not a HK function.

Now, we can prove the theorem by applying the result of Han and Kobayashi [5, Theorem 1] to \((U, V)\) and \( g_n(u, v) = (g(u_1, v_1), \ldots , g(u_n, v_n)); \) it should be noted that, while Han and Kobayashi deal with fixed length coding, the SW region for fixed length coding is identical with that of variable length coding if the source is i.i.d.. Further, it is not hard to see that \((X, Y)\) is slowly varying if \((U, V)\) satisfies the positivity condition, i.e., \( P_{U|V}(u, v) > 0 \) for all \((u, v)\).

\[ \Box \]

\[ D. \text{Proofs of Theorem 5 and Theorem 6} \]

The proof of these theorems are almost the same as that of Theorem 1. Thus, we only show the outline. For a given variable-length code \( \Phi_n = (\hat{\phi}_n^{(1)}, \psi_n) \) for computing \( f_n, \) by a similar argument as Section IV-B we can

\[ 7\text{Again, the result for fixed-length (Theorem 5) can be proved as a special case of the variable-length code.} \]
show that there exists a SW code (with full side-information) \( \hat{\Phi}_n = (\hat{\varphi}_n^{(1)}, \hat{\psi}_n) \) satisfying

\[
P_e(\hat{\Phi}_n | f_n^{id}) \leq P_{X^n Y^n}(T_{n,1}^c) + 2^{-n\delta}
\]

\[
= P_{X^n Y^n}(D_n^c \cap T_{n,1}^c) + P_{X^n Y^n}(D_n \cap T_{n,1}^c) + 2^{-n\delta}
\]

\[
\leq P_e(\Phi_n | f_n) + P_{X^n Y^n}(D_n \cap T_{n,1}^c) + 2^{-n\delta}
\]

and

\[
|\hat{\varphi}_n^{(1)}(x)| \leq \ell^{(1)}(x) + 2n\delta + 2\log(nL_1 + 2n\delta) + 3.
\]

Now, we apply Lemma 1 to (85), and obtain

\[
P_e(\hat{\Phi}_n | f_n^{id}) \leq \left(1 + \frac{2|Y|}{\beta q}\right) P_e(\Phi_n | f_n) + (v_n(\beta) + 2)2^{-n\delta}.
\]

Thus, by taking \( \beta > 0 \) sufficiently small compared to \( \delta > 0 \), we can derive the statement of the theorem.

**E. Proof of Theorem 7**

The only modifications we need is the proof of Lemma 1. In the proof of Lemma 1, we use the properties of sensitivity and slowly varying in (101) and (102). Suppose that \( P_{X^n Y^n}(x'_k, y) \cdot P_{X^n Y^n}(x''_k, y) > 0 \) (otherwise, since \( P_{X^n Y^n}(x_{v+2k}, y) = 0 \), the desired inequality \( P_{X^n Y^n}(x'_{k, j}, y_j) \geq qP_{X^n Y^n}(x_{v+2k}, y) \) holds trivially) and \( x'_k \) and \( x''_k \) differ in \( i_1 \)th, \( \ldots \), \( i_{[\beta n]} \)th positions. Since \( (X, Y) \) is weakly slowly varying, for each \( j = 1, \ldots, [\beta n] \), there exists \( y_j \) that differs from \( y \) only in \( i_1 \)th position and

\[
P_{X^n Y^n}(x'_k, y_j) \geq qP_{X^n Y^n}(x'_k, y),
\]

\[
P_{X^n Y^n}(x''_k, y_j) \geq qP_{X^n Y^n}(x''_k, y).
\]

Furthermore, since \( f_n \) is highly sensitive conditioned on \( Y^n \), we have \( f_n(x'_k, y_j) \neq f_n(x''_k, y_j) \), which implies either of the events in (101) is true. Then, by defining \( x^*_k, j \) in the same manner as the proof of Lemma 1 (102) also holds. The rest of the proof goes through exactly in the same manner.

**F. Proof of Theorem 8**

We only prove (38) since (37) can be proved in a similar manner. It suffice to prove only one direction, i.e.,

\[
e^{vl}(t, \gamma | f) \leq e^{vl}_{SW}(t, \gamma).
\]

For a given code \( \{\Phi_n\}_{n=1}^\infty \) satisfying (37) and

\[
\lim_{n \to \infty} \frac{1}{n^{1-2t}} \log P_e(\Phi_n | f_n) \geq e^{vl}(t, \gamma | f),
\]

we can construct a SW code \( \hat{\Phi}_n = (\hat{\varphi}_n^{(1)}, \hat{\psi}_n) \) satisfying (85) and (87). We set \( \beta = \beta_n = \frac{1}{n} \) and \( \delta = \delta_n = \frac{1}{n^{1-2t}} \). Then, by noting \( h(\beta) \leq 2\beta + 2\beta \log(1/2\beta) \) for \( 0 < \beta < 1/2 \), we have \( v_n(\beta_n) \leq 16|X|n^3 \). Thus, we have

\[
\limsup_{n \to \infty} \frac{1}{n^{1-4t}} \left( \mathbb{E} \left[ |\hat{\varphi}_n^{(1)}(X^n)| \right] - nH(X|Y) \right) \leq \gamma
\]

\[
\liminf_{n \to \infty} \frac{1}{n^{1-2t}} \log P_e(\hat{\Phi}_n | f_n^{id}) \geq e^{vl}(t, \gamma | f).
\]
V. Conclusion

In this paper, we investigated a dichotomy of functions in distributed coding: for a sequence $f$ of functions, does the achievable rate region for computing $f$ coincide with the SW region? We introduced the class of slowly varying sources and gave a sufficient condition for the coincidence: if $f$ is totally sensitive then the achievable rate region for computing $f$ coincides with the SW region for any slowly varying sources. Further, we proved that, for symbol-wise functions, the total sensitivity is the necessary and sufficient condition for the coincidence of two regions. On the other hand, it remains as a future work to establish the necessary and sufficient condition on functions which may not be symbol-wise.

In our investigation, we used the information-spectrum approach so that we can establish the results in a unified way. This approach allows us to derive a refined result in the moderate deviation regime as given in Section IV-D. Although we consider only i.i.d. sources in Section IV-D for simplicity, it is not hard to generalize Theorem 8 for wider classes of sources. Indeed, the assumption of i.i.d. is not so critical in the proof of Theorem 8 given in Section IV-D. On the other hand, for general sources that have memory and may not be stationary nor ergodic, to characterize $e_{SW}^{(t)}(t, \gamma)$ and $e_{SW}^{(t)}(t, \gamma)$ itself remains as an important work.

In this paper, we considered only lossless computation, where the error probability is required to tend zero as the block size goes to infinity. It is an important future work to generalize our results for $\varepsilon$-error case, where the error probability is required only to be smaller than the given threshold $\varepsilon > 0$. When we consider $\varepsilon$-error case, the strong converse property is an important subject to be investigated; e.g., it is an interesting problem to establish the necessary and sufficient condition on functions so that the strong converse holds for function computation whenever the strong converse holds for SW coding. Furthermore, it is also an important future work to generalize our results for lossy case and to establish the condition so that the rate-distortion region for distributed computing coincides with that for distributed source coding.

APPENDIX A

Proof of Proposition 1

a) If part: At first, we assume that $f$ is a HK function and satisfies 1) of the proposition. Then we have

$$f(a_1, b_1) = f(a_2, b_2) \text{ means } b_1 = b_2. \quad (93)$$

Indeed, if $a_1 = a_2$ then (93) follows from 1) of the proposition. Moreover, if $a_1 \neq a_2$ then (93) follows from the condition 3) in the definition of HK functions.

Now, note that if $f_n(x, y) = f_n(\hat{x}, \hat{y})$ then $f(x_i, y_i) = f(\hat{x}_i, \hat{y}_i)$ for all $i = 1, 2, \ldots, n$, since $f_n$ is symbol-wise. Hence, by (93), we can see that if $f_n(x, y) = f_n(\hat{x}, \hat{y})$ then $y_i = \hat{y}_i$ for all $i = 1, 2, \ldots, n$, that is, $y = \hat{y}$.

On the other hand, similar argument holds for a case where $f$ satisfies 2) of the proposition, and we can show that if $f_n(x, y) = f_n(\hat{x}, \hat{y})$ then $x = \hat{x}$ in this case.

Summarizing the above, if $f$ is a HK function and satisfies 1) or 2) of the proposition then $f_n(x, y) = f_n(\hat{x}, \hat{y})$ implies $x = \hat{x}$ or $y = \hat{y}$. This completes the proof of “if part”.

August 27, 2014 DRAFT
holds, by the assumption, there exists obtained by replacing the
Next, we select the
Clearly, we have
\[ a, b, \hat{x} = (a, a), y^2 = (b, b), \text{ and } y^2 = (b, b) \text{ satisfy } x^2 \neq \hat{x}^2, y^2 \neq \hat{y}^2, \text{ and } f_2(x^2, y^2) = f_2(\hat{x}^2, \hat{y}^2). \] □

**APPENDIX B**

**PROOF OF PROPOSITION [4]**

\[ c \) If part: \) Let \( q \triangleq \min\{P_{XY}(a, b) : (a, b) \in \mathcal{X} \times \mathcal{Y}, P_{XY}(a, b) > 0\} \). Fix \( x \neq \hat{x} \) and \( y \) satisfying \( P_{X \times Y}(x, y) = P_{X \times Y}(\hat{x}, y) > 0 \) arbitrarily, and suppose that \( x_i \neq \hat{x}_i \). Since \( P_{X, Y}(x_i, y_i) \cdot P_{X, Y}(\hat{x}_i, y_i) > 0 \) holds, by the assumption, there exists \( b \neq y_i \) satisfying \( P_{X, Y}(x_i, b) \cdot P_{X, Y}(\hat{x}_i, b) > 0 \). We can see that \( \hat{y} \in \mathcal{Y}^n \) obtained by replacing the \( i \)th component of \( y \) with \( b \) satisfies (33) and (34).

\[ □ \]

\[ d \) Only if part: \) This part is obvious, since if the source is weakly slowly varying then the property required in Definition [11] holds for \( n = 1 \).

**APPENDIX C**

**PROOF OF LEMMA [1]**

Throughout the proof, we omit subscript \( n \) if it is obvious from the context. Furthermore, we also omit \( \beta \) from \( v_n(\beta) \) and \( u_n(\beta) \), and thus they are just denoted by \( v \) and \( u \). For \( a \in \mathcal{C}^{(1)} \) and \( b \in \mathcal{C}^{(2)} \), let

\[ \mathcal{D}_{a, b} \triangleq \left\{ (x, y) : \varphi(1)(x) = a, \varphi(2)(y) = b, (x, y) \in \mathcal{D} \right\}, \]

\[ \mathcal{D}_{a, y} \triangleq \left\{ (x, y) : \varphi(1)(x) = a, (x, y) \in \mathcal{D} \right\}, \]

\[ \mathcal{D}_{x, b} \triangleq \left\{ (x, y) : \varphi(2)(y) = b, (x, y) \in \mathcal{D} \right\}. \]

**Proof of [52]:** We leverage El Gamal’s argument [6]. For each \( (a, y) \), we sort the elements in \( \mathcal{D}_{a, y} \) in the decreasing order of probabilities, i.e.,

\[ P_{X \times Y}(x_1, y) \geq P_{X \times Y}(x_2, y) \geq \cdots \geq P_{X \times Y}(x_{|\mathcal{D}_{a, y}|}, y). \]

First, we take \( x_1' \triangleq x_1 \), and pair it with an \( x_1'' \in \mathcal{D}_{a, y} \) that satisfies \( d(x_1', x_1'') \geq \beta n \) and has the largest probability. Clearly, we have

\[ P_{X \times Y}(x_1', y) \geq P_{X \times Y}(x_{v+2}, y). \]

Next, we select the \( x_2' \in \mathcal{D}_{a, y} \backslash \{x_1', x_1''\} \) with the largest probability, and pair it with an unselected \( x_2'' \) satisfying \( d(x_2', x_2'') \geq \beta n \) and that has the largest probability. Clearly, we have

\[ P_{X \times Y}(x_2', y) \geq P_{X \times Y}(x_3, y), \]

\[ P_{X \times Y}(x_2'', y) \geq P_{X \times Y}(x_{v+4}, y). \]
We repeat this process until no more pairing is possible. Then, since \( f_n \) is sensitive conditioned on \( Y^n \), for pair \((x'_k, x''_k)\), we can find \( y_1, \ldots, y_{[\beta n]} \) such that \( d(y_i, y_j) = 1 \) and \( f_n(x'_k, y_j) \neq f_n(x''_k, y_j) \), which implies that either
\[
(x'_k, y_j) \in D^c \quad \text{or} \quad (x''_k, y_j) \in D^c
\]
(101) is true. For each \( j \), let \( x_{k,j}^* \in \{x'_k, x''_k\} \) be such that \((x_{k,j}^*, y_j) \in D^c\). Since \((X, Y)\) is slowly varying with respect to \( Y \), we have
\[
P_{X^nY^n}(x_{k,j}^*, y_j) \geq qP_{X^nY^n}(x_{k,j}^*, y)
\]
(102)
\[
\geq qP_{X^nY^n}(x_{v+2k}, y),
\]
(103)
where the second inequality follows from the procedure of pairing (cf. (99) and (100)). Thus, we have
\[
\left[ \frac{1}{2} \left| (D_{a,y}|-v) \right| \right]^+ \sum_{k=1}^{[\beta n]} P_{X^nY^n}(x_{v+2k}, y) \leq \sum_{j=1}^{[\beta n]} \sum_{k=1}^{[\beta n]} P_{X^nY^n}(x_{k,j}^*, y_j),
\]
(104)
Here, note that
\[
\bigcup_{a \in C^{(1)}} \bigcup_{y \in Y^n} \left\{x_{k,j}^*, y_j : k = 1, \ldots, \left[ \frac{1}{2} \left| (D_{a,y}|-v) \right| \right]^+, j = 1, \ldots, [\beta n] \right\} \subset D^c,
\]
(105)
and each element in \( D^c \) overlaps at most \( n|Y| \) times in the lefthand side. Thus, we have
\[
\sum_{a \in C^{(1)}} \sum_{y \in Y^n} \left[ \frac{1}{2} \left| (D_{a,y}|-v) \right| \right]^+ P_{X^nY^n}(x_{v+2k}, y) \leq \frac{n|Y|}{\beta q} \sum_{(x,y) \in D^c} P_{X^nY^n}(x, y)
\]
(106)
\[
= \frac{|Y|}{\beta q} P_e(\Phi_n | f_n).
\]
(107)
Now, we have
\[
P_{X^nY^n}(D \cap T^c_1) = \sum_{a \in C^{(1)}} \sum_{y \in D_{a,y}} P_{X^nY^n}(x, y) 1[(x, y) \in T^c_1]
\]
(108)
\[
\leq \sum_{a \in C^{(1)}} \sum_{y} (v+1)P_{Y^n}(y) 2^{-\ell(a) - n\delta}
\]
(109)
\[
+ \sum_{a \in C^{(1)}} \sum_{y} \left[ \frac{1}{2} \left| (D_{a,y}|-v) \right| \right]^+ (P_{X^nY^n}(x_{v+2k}, y) + P_{X^nY^n}(x_{v+2k+1}, y))
\]
(110)
\[
\leq (v+1)2^{-n\delta} + 2 \sum_{a \in C^{(1)}} \sum_{y} \left[ \frac{1}{2} \left| (D_{a,y}|-v) \right| \right]^+ P_{X^nY^n}(x_{v+2k}, y)
\]
(111)
\[
\leq (v+1)2^{-n\delta} + \frac{2|Y|}{\beta q} P_e(\Phi_n | f_n),
\]
(112)
where \( \ell(a) \) is the length of codeword \( a \); the first inequality is derived by splitting \( D_{a,y} \) into the first \( v+1 \) elements and the rest, and then by applying the property of \( T^c_1 \) to the former; and the second inequality follows from the Kraft inequality. Thus, we have the desired bound.

\[\square\]

\(^8\)This process continues at least \( \left[ \frac{1}{2} \left| (D_{a,y}|-v) \right| \right]^+ \) times, which may be 0.

\(^9\)It should be noted that \( x_{k,j}^* \) and \( y_j \) implicitly depend on \( y \).

August 27, 2014 DRAFT
Proof of (54): To bound $P_{X^nY^n} (D \cap T_0^c)$, we need the following observation. Since $f_n$ is jointly sensitive, if we pick arbitrary $(x^{*n}_a, y^{*n}_b) \in D_{a,b}$, one of the following must be true:

$$D_{a,b} = D^{*n}_{a,b}.$$

(113)

Otherwise, there exists $(x,y),(x',y') \in D_{a,b}$ such that $x \neq x'$ and $y \neq y'$, but it contradict with the fact that $f_n$ is jointly sensitive. Consequently, we have

$$P_{X^nY^n} (D \cap T_0^c) = \sum_{\alpha \in C(1)} \sum_{\beta \in C(2)} \sum_{(x,y) \in D_{a,b}} P_{X^nY^n} (x,y) 1 \{ (x,y) \in T_0^c \}$$

(114)

$$\leq \sum_{\alpha \in C(1)} \sum_{\beta \in C(2)} \left( \sum_{x \in D_{a,b}^{*n}} P_{X^nY^n} (x,y) 1 \{ (x,y) \in T_0^c \} \right)$$

(115)

$$+ \sum_{y \in D_{a,b}^{*n}} P_{X^nY^n} (x^{*n}_a, y) 1 \{ (x^{*n}_a, y) \in T_0^c \}$$

(116)

$$\leq \sum_{\alpha \in C(1)} \sum_{\beta \in C(2)} \left( [(\nu + 1) + (\nu + 1)] 2^{-\ell(a) - \ell(b) - n\delta} \cdot \sum_{k=1}^{[\frac{1}{4} |D_{a,b}^{*n}| - \nu]} (P_{X^nY^n} (x^{*n}_{v+2k}, y^{*n}_b) + P_{X^nY^n} (x^{*n}_{v+2k+1}, y^{*n}_b)) \right)$$

(117)

where $y_{u+2k}$ is defined in a similar manner as $x_{v+2k}$ by sorting the elements in $D_{a,b}$ for each $a$ and $b$ (cf. (97)), and where the inequality in (117) is derived in a similar manner as the inequality in (109). By the Kraft inequality, we have

$$\sum_{\alpha \in C(1)} \sum_{\beta \in C(2)} [(\nu + 1) + (\nu + 1)] 2^{-\ell(a) - \ell(b) - n\delta} \leq [(\nu + 1) + (\nu + 1)] 2^{-n\delta}.$$  

(120)

By using (107), we have

$$\sum_{\alpha \in C(1)} \sum_{\beta \in C(2)} \left[ \frac{1}{4} |D_{a,b}^{*n}| - \nu \right] (P_{X^nY^n} (x^{*n}_{v+2k}, y^{*n}_b) + P_{X^nY^n} (x^{*n}_{v+2k+1}, y^{*n}_b))$$

(121)

$$\leq 2 \sum_{\alpha \in C(1)} \sum_{\beta \in C(2)} \sum_{k=1} \left( \frac{1}{4} |D_{a,b}^{*n}| - \nu \right) P_{X^nY^n} (x^{*n}_{v+2k}, y)$$

(122)

$$\leq \frac{2|Y|}{\beta q} \mathbb{P}_e (\Phi_n | f_n).$$

(123)

Similarly, we have

$$\sum_{\alpha \in C(1)} \sum_{\beta \in C(2)} \left[ \frac{1}{4} |D_{a,b}^{*n}| - \nu \right] (P_{X^nY^n} (x^{*n}_a, y^{*n}_{u+2k}) + P_{X^nY^n} (x^{*n}_a, y^{*n}_{u+2k+1}))$$

(124)

$$\leq \frac{2|Y|}{\beta q} \mathbb{P}_e (\Phi_n | f_n).$$

(125)
Thus, we have the desired bound.

REFERENCES

[1] J. Körner and K. Marton, “How to encode the modulo-two sum of binary sources,” IEEE Trans. Inf. Theory, vol. IT-25, pp. 219–221, Mar. 1979.
[2] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” IEEE Trans. Inf. Theory, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.
[3] A. El Gamal and Y.-H. Kim, Network Information Theory. Cambridge University Press, 2011.
[4] R. Ahlswede and I. Csiszár, “To get a bit of information may be as hard as to get full information,” IEEE Trans. Inf. Theory, vol. 27, no. 4, pp. 398–408, 1981.
[5] T. S. Han and K. Kobayashi, “A dichotomy of functions $F(X,Y)$ of correlated sources $(X,Y)$ from the viewpoint of the achievable rate region,” IEEE Trans. Inf. Theory, vol. 33, no. 1, pp. 69–76, Jan. 1987.
[6] A. A. E. Gamal, “A simple proof of the Ahlswede-Csiszár one-bit theorem,” IEEE Trans. Inf. Theory, vol. 29, no. 6, pp. 931–933, 1983.
[7] T. S. Han, Information-spectrum methods in information theory. New York: Springer-Verlag, 2002.
[8] S. Kuzuoka and S. Watanabe, “An information-spectrum approach to weak variable-length source coding with side-information,” Jan. 2013, arXiv:1401.3809.
[9] D. K. He, L. A. Lastras-Montaño, E. H. Yang, A. Jagmohan, and J. Chen, “On the redundancy of Slepian-Wolf coding,” IEEE Trans. Inf. Theory, vol. 55, no. 12, pp. 5607–5627, Dec. 2009.
[10] Y. Altuğ and A. B. Wagner, “Moderate deviations in channel coding,” IEEE Trans. Inf. Theory, vol. 60, no. 8, pp. 4417–4426, 2014.
[11] H. Yamamoto, “Wyner-Ziv theory for a general function of the correlated sources,” IEEE Trans. Inf. Theory, vol. 28, no. 5, pp. 803–807, Sep. 1982.
[12] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd ed. John Wiley & Sons, Inc., 2006.
[13] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems. New York: Academic, 1981.
[14] S. Miyake and F. Kanaya, “Coding theorems on correlated general sources,” IEICE Trans. Fundamentals, vol. E78-A, no. 9, pp. 1063–1070, Sep. 1995.
[15] P. Elias, “Universal codeword sets and representations of the integers,” IEEE Trans. Inf. Theory, vol. 21, no. 2, pp. 194–203, 1975.