論文 / 著書情報
Article / Book Information

Title	New Non-Asymptotic Bounds on Numbers of Codewords for the Fixed-Length Lossy Compression
Authors	Tetsunao Matsuta, Tomohiko Uyematsu
Citation	IEICE Trans. Fundamentals, Vol. E99-A, No. 12, pp. 2116-2129
Pub. date	2016, 12
URL	http://search.ieice.org/
Copyright	(c) 2016 Institute of Electronics, Information and Communication Engineers

IEICE TRANS. FUNDAMENTALS, VOL.E99–A, NO.12 DECEMBER 2016

PAPER Special Section on Information Theory and Its Applications

New Non-Asymptotic Bounds on Numbers of Codewords for the Fixed-Length Lossy Compression*

Tetsunao MATSUTA[†a)], ***Member and*** **Tomohiko UYEMATSU**[†b)], ***Fellow***

SUMMARY In this paper, we deal with the fixed-length lossy compression, where a fixed-length sequence emitted from the information source is encoded into a codeword, and the source sequence is reproduced from the codeword with a certain distortion. We give lower and upper bounds on the minimum number of codewords such that the probability of exceeding a given distortion level is less than a given probability. These bounds are characterized by using the α-mutual information of order infinity. Further, for i.i.d. binary sources, we provide numerical examples of tight upper bounds which are computable in polynomial time in the blocklength.

key words: *finite blocklength, non-asymptotic bound, rate-distortion function, source coding*

1. Introduction

The fixed-length lossy compression is a typical source coding, where a fixed-length sequence emitted from the information source is encoded into a codeword, and the source sequence is reproduced from the codeword with a certain distortion. One of the most important parameter of the fixed-length lossy compression is the *rate* of code such that the distortion between the source and the reproduction sequences is less than a given distortion level. Especially, the limit of the minimum rate as the blocklength tends to infinity is frequently called the *rate-distortion (RD) function*. There are a lot of studies dealing with the RD function (see e.g. [3]–[5]).

On the other hand, recently, non-asymptotic behavior of the minimum rate (i.e., the minimum rate for finite blocklengths) becomes an active target of the study [6]–[9]. Especially, Kostina and Verdú [7] reported a lot of non-asymptotic results of the minimum rate. They considered the minimum number of codewords such that the probability of exceeding a given distortion level is less than a given probability. This distortion criterion is known as the excess distortion criterion [7] or ϵ-fidelity criterion [3]. Since the minimum rate is easily calculated by the minimum number of codewords, they focus on it instead of the rate. Then, they gave an achievability bound (i.e., upper bound) and a converse bound (i.e., lower bound) to the minimum number of codewords.

In this paper, we also deal with the fixed-length lossy compression for the ϵ-fidelity criterion, and give a new achievability bound and a converse bound to the minimum number of codewords. Our bounds are characterized by using a quantity equivalent to the α-mutual information of order inifinity [10], where the α-mutual information is a generalized version of the mutual information and has many interesting properties (see [10], [11]). Our achievability bound is derived by using a slightly generalized covering lemma. We show that our converse bound is tighter than that of Kostina and Verdú [7]. By using these bounds, we attempt to characterize the limit of the minimum rate, and give several cases where the limit can be characterized by the α-mutual information of order infinity. We also show numerical examples of two achievability bounds which are induced by our achievability bound and computable in polynomial time in the blocklength. Then, we demonstrate that there exist cases where these two bounds are tighter than that of Kostina and Verdú [7].

The rest of this paper is organized as follows. In Sect. 2, we provide a precise definition of the fixed-length lossy compression and the minimum number of codewords. In Sect. 3, we give our achievability and converse bounds. In Sect. 4, we prove our main results. In Sect. 5, we give several cases where the limit of the minimum rate can be characterized by the α-mutual information. In Sect. 6, we provide numerical examples of two polynomial-time computable achievability bounds. In Sect. 7, we conclude the paper.

2. Preliminaries

In this section, we provide some notation, definitions, and a known result for the fixed-length lossy compression.

Unless otherwise stated, we will use the following notations. Let $\mathcal{X}$ and $\mathcal{Y}$ be finite or countably infinite sets which represent the source alphabet and the reproduction alphabet, respectively. We will denote the set of all probability distributions over $\mathcal{X}$ by $\mathcal{P}(\mathcal{X})$, and the set of all conditional distributions from $\mathcal{Y}$ to $\mathcal{X}$ by $\mathcal{W}(\mathcal{X}|\mathcal{Y})$. The probability distribution of a random variable (RV) X will be denoted by the subscripted notation P_X, and the conditional distribution for X given an RV Y will be denoted by $P_{X|Y}$.

Let X be an RV on $\mathcal{X}$ which represents a single source symbol. When we consider $\mathcal{X}$ and $\mathcal{Y}$ as n-fold Cartesian products of finite or countably infinite sets $\mathcal{A}$ and $\mathcal{B}$, respec-

Manuscript received January 18, 2016.
Manuscript revised June 17, 2016.

[†]The authors are with Dept. of Communications and Computer Engineering, Tokyo Institute of Technology, Tokyo, 152-8550 Japan.

*Portions of this paper were presented at the 37th Symposium on Information Theory and Its Applications [1], and at the 2015 IEEE International Symposium on Information Theory [2].

a) E-mail: tetsu@it.ce.titech.ac.jp
b) E-mail: uyematsu@ieee.org

DOI: 10.1587/transfun.E99.A.2116

tively, we can identify the source symbol X with the n-length source sequence. Thus, for the sake of the brevity, we only deal with the single source symbol unless otherwise stated.

For the lossy compression, the encoder is a map $f : \mathcal{X} \to \mathcal{Y}$. The decoder is implicitly employed in this encoding procedure. In fact, the encoder outputs a codeword, and the decoder outputs a reproduction symbol using a *one-to-one* mapping from the set of codewords to $\mathcal{Y}$.

In order to measure the distortion between the source symbol and the reproduction symbol, we introduce the distortion measure defined by a map $d : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$.

For this setting, we want to know the minimum number of the cardinality of the image of encoder f (i.e., the number of codewords) for the ϵ-fidelity criterion. To this end, we introduce the following definitions.

Definition 1. For a given source X, a distortion level $D \geq 0$, and a specified provability $0 \leq \epsilon \leq 1$, we say f is an (M, D, ϵ) code if and only if $\|f\| = M$ and $\Pr\{d(X, f(X)) > D\} \leq \epsilon$, where $\|f\|$ denotes the cardinality of the image of f.

Definition 2. For $D \geq 0$ and $\epsilon \geq 0$,

$$M^*(D, \epsilon) \triangleq \min\{M : \exists (M, D, \epsilon) \text{ code}\}.$$

When $\mathcal{X} = \mathcal{A}^n$ and $\mathcal{Y} = \mathcal{B}^n$, $\frac{1}{n}\log_2 \|f\|$ is called as the rate of the code, where $\mathcal{A}^n$ and $\mathcal{B}^n$ are n-fold Cartesian products of alphabets $\mathcal{A}$ and $\mathcal{B}$, respectively. Then, the minimum of rates of (M, D, ϵ) codes (say, $R(n, D, \epsilon)$) is sometimes called the *finite blocklength RD function* (with ϵ-fidelity criterion), i.e.,

$$R(n, D, \epsilon) = \frac{1}{n}\log_2 M^*(D, \epsilon).$$

For the minimum number $M^*(D, \epsilon)$, Kostina and Verdú [7] gave the following converse bound.

Theorem 1. For any given source X, distortion measure d, constant $D \geq 0$, and $0 \leq \epsilon \leq 1$, we have

$$\sup_{P_{\tilde{X}} \in \mathcal{P}(\mathcal{X})} \inf_{y \in \mathcal{Y}} \frac{\beta_{1-\epsilon}(P_X, P_{\tilde{X}})}{\mathrm{E}[1\{d(\tilde{X}, y) \leq D\}]} \leq M^*(D, \epsilon),$$

where $\tilde{X}$ is an RV subject to the distribution $P_{\tilde{X}}$ on $\mathcal{X}$, $\mathrm{E}[\cdot]$ denotes the expectation operator, $1\{\cdot\}$ denotes the indicator function, and

$$\beta_{1-\epsilon}(P_X, P_{\tilde{X}}) \triangleq \min_{\substack{\phi : \mathcal{X} \to [0,1], \\ \sum_{x \in \mathcal{X}} P_X(x)\phi(x) \geq 1-\epsilon}} \sum_{x \in \mathcal{X}} P_{\tilde{X}}(x)\phi(x). \tag{1}$$

3. Main Results

In this section, we give a new achievability bound and a converse bound, and show the relation between our converse bound and Theorem 1.

The next theorem shows our achievability bound.

Theorem 2. For any given source X, distortion measure d, constant $D \geq 0$, and $0 \leq \delta < \epsilon \leq 1$, we have

$$M^*(D, \epsilon) \leq \min_{\substack{P_{Y|X} \in \mathcal{W}(\mathcal{Y}|\mathcal{X}): \\ \Pr\{d(X,Y) > D\} \leq \delta}} \left\lceil \mu(P_{XY}) \ln \frac{1-\delta}{\epsilon - \delta} \right\rceil, \tag{2}$$

where $P_{XY}(x, y) = P_X(x)P_{Y|X}(y|x)$, $\lceil \cdot \rceil$ denotes the ceiling function, and

$$\mu(P_{XY}) \triangleq \sum_{y \in \mathcal{Y}} \sup_{x \in \mathcal{X} : P_X(x) > 0} P_{Y|X}(y|x).$$

The next theorem shows our converse bound.

Theorem 3. For any given source X, distortion measure d, constant $D \geq 0$, and $0 \leq \epsilon \leq 1$, we have

$$\min_{\substack{P_{Y|X} \in \mathcal{W}(\mathcal{Y}|\mathcal{X}): \\ \Pr\{d(X,Y) > D\} \leq \epsilon}} \lceil \mu(P_{XY}) \rceil \leq M^*(D, \epsilon). \tag{3}$$

We postpone the proof of these theorems to Sect. 4.

Remark 1. Since the ceiling function returns an integer value, the right-hand side (RHS) of (2) and the left-hand side of (3) can be minimized. Hence, we write min instead of inf.

Remark 2. $\mu(P_{XY})$ is equivalent to the α-mutual information [10] of order infinity $I_\infty(X; Y)$. In fact, we have

$$I_\infty(X; Y) = \log \mu(P_{XY}).$$

A conditional distribution $P_{Y|X} \in \mathcal{W}(\mathcal{Y}|\mathcal{X})$ is equivalent to a *stochastic* encoder. In fact, if we consider $\Pr\{y$ is a reproduction symbol of $x\}$ as $P_{Y|X}(y|x)$, a stochastic encoder gives a conditional distribution and *vice versa*, where the probability is induced by the stochastic encoder. Especially, a *deterministic* encoder f gives a distribution $P_{Y|X}$ as a special case. In fact, if we consider $1\{y$ is a reproduction symbol of $x\}$ as $P_{Y|X}(y|x)$, it gives a conditional distribution. Then, $\sup_{x \in \mathcal{X} : P_X(x) > 0} P_{Y|X}(y|x)$ is equivalent to $1\{y$ is a reproduction symbol of some source symbols$\}$. Hence, for the conditional distribution representing a deterministic encoder, $\mu(P_{XY})$ is just counting the number of codewords. Now, for stochastic encoder, since $P_{Y|X}(y|x)$ can also be considered as the expected number of choosing y as a reproduction symbol of x, $\sup_{x \in \mathcal{X} : P_X(x) > 0} P_{Y|X}(y|x)$ can be considered as the worst expected number of choosing y as a reproduction symbol. Hence, $\mu(P_{XY})$ can be considered as the worst expected number of codewords of the stochastic encoder. According to this perspective, our achievability and converse bounds may be considered as a characterization of the minimum number of codewords of *deterministic* encoder from a viewpoint of the minimum expected number of codewords of *stochastic* encoder.

The next theorem shows that our converse bound is tighter than the bound of Theorem 1.

Theorem 4. For any given source X such that $P_X(x) > 0$ for all $x \in \mathcal{X}$, distortion measure d, constant $D \geq 0$, and $0 \leq \epsilon \leq 1$, we have

$$\sup_{P_{\tilde{X}} \in \mathcal{P}(\mathcal{X})} \inf_{y \in \mathcal{Y}} \frac{\beta_{1-\epsilon}(P_X, P_{\tilde{X}})}{\mathrm{E}[1\{d(\tilde{X}, y) \leq D\}]} \leq \min_{\substack{P_{Y|X} \in \mathcal{W}(\mathcal{Y}|\mathcal{X}): \\ \Pr\{d(X,Y) > D\} \leq \epsilon}} \lceil \mu(P_{XY}) \rceil.$$

Proof. For $P_{Y|X} \in \mathcal{W}(\mathcal{Y}|\mathcal{X})$ satisfying

$$\Pr\{d(X,Y) > D\} \le \epsilon, \tag{4}$$

let $\phi : \mathcal{X} \to [0,1]$ be a map defined by

$$\phi(x) \triangleq \sum_{y\in\mathcal{Y}} P_{Y|X}(y|x)1\{d(x,y) \le D\}, \quad \forall x \in \mathcal{X}. \tag{5}$$

Then, due to (4), we have

$$\sum_{x\in\mathcal{X}} P_X(x)\phi(x) \ge 1-\epsilon. \tag{6}$$

On the other hand, for any $P_{\tilde{X}} \in \mathcal{P}(\mathcal{X})$, we have

$$\begin{aligned}
&\lceil \mu(P_{XY}) \rceil \left(\sup_{y\in\mathcal{Y}} \mathrm{E}[1\{d(\tilde{X},y) \le D\}] \right) \\
&\ge \sum_{y\in\mathcal{Y}} \left(\sup_{x:P_X(x)>0} P_{Y|X}(y|x) \right) \left(\sum_{x\in\mathcal{X}} P_{\tilde{X}}(x)1\{d(x,y) \le D\} \right) \\
&\ge \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} P_{\tilde{X}}(x)P_{Y|X}(y|x)1\{d(x,y) \le D\} \\
&\overset{\text{(a)}}{\ge} \beta_{1-\epsilon}(P_X, P_{\tilde{X}}),
\end{aligned}$$

where (a) comes from (1), (5), and (6). Thus, we have

$$\frac{\beta_{1-\epsilon}(P_X, P_{\tilde{X}})}{\sup_{y\in\mathcal{Y}} \mathrm{E}[1\{d(\tilde{X},y) \le D\}]} \le \lceil \mu(P_{XY}) \rceil .$$

Since this inequality holds for any $P_{\tilde{X}} \in \mathcal{P}(\mathcal{X})$, and any $P_{Y|X} \in \mathcal{W}(\mathcal{Y}|\mathcal{X})$ satisfying (4), we have the theorem. □

In order to employ our converse bound, we need to compute $\mu(P_{XY})$. If we compute this quantity straightforwardly, the time complexity will be $O(|\mathcal{X}||\mathcal{Y}|)$ in general. Further, we also need to solve the optimization problem over the set $\mathcal{W}(\mathcal{Y}|\mathcal{X})$ of conditional probability distributions. Hence, computing our bound will be rather difficult especially when cardinalities of alphabets $\mathcal{X}$ and $\mathcal{Y}$ are very large. For example, if $\mathcal{X} = \mathcal{A}^n$ and $\mathcal{Y} = \mathcal{B}^n$, the time complexity grows exponentially in the blocklength n even if the source is stationary and memoryless. On the other hand, according to Theorem 1, for an arbitrarily fixed $P_{\tilde{X}} \in \mathcal{P}(\mathcal{X})$, we have the next converse bound.

$$\inf_{y\in\mathcal{Y}} \frac{\beta_{1-\epsilon}(P_X, P_{\tilde{X}})}{\mathrm{E}[1\{d(\tilde{X},y) \le D\}]} \le M^*(D,\epsilon).$$

According to [7], if we choose $P_{\tilde{X}}$ properly, this converse bound is tight and can be computed in polynomial time in the blocklength. Hence, although the bound of Theorem 1 is looser than our converse bound, it will be somewhat more tractable than ours.

Now, according to Theorems 2 and 3, we have the next slightly loose bounds:

$$\begin{aligned}
\inf_{\substack{P_{Y|X}\in\mathcal{W}(\mathcal{Y}|\mathcal{X}):\\ \Pr\{d(X,Y)>D\}\le\epsilon}} \mu(P_{XY}) &\le M^*(D,\epsilon) \\
&\le \inf_{\substack{P_{Y|X}\in\mathcal{W}(\mathcal{Y}|\mathcal{X}):\\ \Pr\{d(X,Y)>D\}\le\delta}} \mu(P_{XY}) \ln\frac{1-\delta}{\epsilon-\delta} + 1.
\end{aligned}$$

Since $\mu(P_{XY})$ is a convex function on $P_{Y|X} \in \mathcal{W}(\mathcal{Y}|\mathcal{X})$, and $\{P_{Y|X} \in \mathcal{W}(\mathcal{Y}|\mathcal{X}) : \Pr\{d(X,Y) > D\} \le \epsilon\}$ is a closed convex set, computing the above bounds is equivalent to solve the convex optimization problem. Thus, these slightly loose bounds may be easily computed when cardinalities of alphabets $\mathcal{X}$ and $\mathcal{Y}$ are small.

4. Proof of Theorems

In this section, we prove Theorem 2 and Theorem 3.

4.1 Achievability Bound

First, we prove Theorem 2. To this end, we introduce the *generalized* covering lemma.

Lemma 1 (Generalized covering lemma)**.** Let X, Y, and $\tilde{Y}$ be RVs respectively on $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Y}$. For an integer $M > 0$, let $\tilde{Y}_1, \tilde{Y}_2, \cdots, \tilde{Y}_M$ be RVs which are independent of each other and of X, and each distributed according to $P_{\tilde{Y}}$. Then, for any subset $\mathcal{F} \subseteq \mathcal{X}\times\mathcal{Y}$, any function $\psi : \mathcal{X}\times\mathcal{Y} \to [0,1]$, and any constant $0 \le \alpha \le 1$ such that

$$P_X(x)P_{\tilde{Y}}(y) \ge \alpha\psi(x,y)P_{XY}(x,y), \forall (x,y) \in \mathcal{X}\times\mathcal{Y}, \tag{7}$$

we have

$$\begin{aligned}
&\Pr\left\{ \bigcap_{i=1}^{M} \{(X,\tilde{Y}_i) \notin \mathcal{F}\} \right\} \\
&\le 1 - (\mathrm{E}[\psi(X,Y)] - \Pr\{(X,Y) \notin \mathcal{F}\})(1-e^{-\alpha M}).
\end{aligned} \tag{8}$$

Proof. We have

$$\begin{aligned}
&\Pr\left\{ \bigcap_{i=1}^{M} \{(X,\tilde{Y}_i) \notin \mathcal{F}\} \right\} \\
&= \sum_{x\in\mathcal{X}} P_X(x)\Big(1 - \sum_{y:(x,y)\in\mathcal{F}} P_{\tilde{Y}}(y)\Big)^M \\
&\overset{\text{(a)}}{\le} \sum_{x\in\mathcal{X}} P_X(x)\Big(1 - \alpha\sum_{y:(x,y)\in\mathcal{F}} \psi(x,y)P_{Y|X}(y|x)\Big)^M \\
&\overset{\text{(b)}}{\le} 1 - \sum_{(x,y)\in\mathcal{F}} \psi(x,y)P_{XY}(x,y)(1-e^{-\alpha M}) \\
&\le 1 - (\mathrm{E}[\psi(X,Y)] - \Pr\{(X,Y) \notin \mathcal{F}\})(1-e^{-\alpha M}),
\end{aligned}$$

where (a) comes from (7), and (b) follows since $(1-xy)^M \le 1 - x(1-e^{-yM})$ for $0 \le x, y \le 1$ and $M > 0$ (cf. e.g. [12, Lemma 10.5.3]). □

By changing function ψ and constant α, we can obtain many types of covering lemma (e.g., [13, Lemma 5], [14, Lemma 3.3], [9, Lemma 4]) as corollaries of Lemma 1, which is the reason why we call it "generalized". Due to space limitation, we only show a corollary used to prove

Theorem 2.

Corollary 1. For any integer $M > 0$, any subset $\mathcal{F} \subseteq \mathcal{X}\times\mathcal{Y}$, and any $P_{\tilde{Y}} \in \mathcal{P}(\mathcal{Y})$, we have

$$\Pr\left\{\bigcap_{i=1}^{M}\{(X,\tilde{Y}_i) \notin \mathcal{F}\}\right\} \le 1 - \Pr\{(X,Y)\in\mathcal{F}\}(1 - e^{-\exp(-D_\infty(P_{XY},P_{\tilde{Y}}))M}),$$

where $D_\infty(P_{XY}, P_{\tilde{Y}})$ is the Rényi divergence of order infinity [15] defined as

$$D_\infty(P_{XY}, P_{\tilde{Y}}) \triangleq \ln \sup_{(x,y):P_X(x)>0} \frac{P_{Y|X}(y|x)}{P_{\tilde{Y}}(y)}.$$

Proof. Let $\psi(x,y) = 1$ for all $(x,y) \in \mathcal{X}\times\mathcal{Y}$, and $\alpha = \exp(-D_\infty(P_{XY}, P_{\tilde{Y}}))$. Since $\psi(x,y)$ and α satisfy (7), by substituting these into (8), we have the corollary. □

We use the above corollary to prove the next theorem which implies our achievability bound.

Theorem 5. For any $D > 0$, $0 \le \delta < \epsilon \le 1$, $P_{\tilde{Y}} \in \mathcal{P}(\mathcal{Y})$, and $P_{Y|X} \in \mathcal{W}(\mathcal{Y}|\mathcal{X})$ such that

$$\Pr\{d(X,Y) > D\} \le \delta, \tag{9}$$

there exists an (M, D, ϵ) code which satisfies

$$M \le \left\lceil \exp(D_\infty(P_{XY}, P_{\tilde{Y}})) \ln\frac{1-\delta}{\epsilon-\delta}\right\rceil.$$

Proof. We generate $\tilde{y}_1, \tilde{y}_2, \cdots, \tilde{y}_M \in \mathcal{Y}$ independently subject to the probability distribution $P_{\tilde{Y}}$, and define the set $C \triangleq \{\tilde{y}_1, \tilde{y}_2, \cdots, \tilde{y}_M\}$. For a given set C and a given $x \in \mathcal{X}$, we define $f(x) = \tilde{y}_{i_0}$, where the index i_0 is determined by $i_0 = \arg\min_{1\le i\le M} d(x, \tilde{y}_i)$. Then, we define the probability $P_e(C) \triangleq \Pr\{d(X, f(X)) > D\}$. By taking the average over the random selection of C, the average probability $P_e(C)$ is bounded as follows:

$$\begin{aligned}
&\mathrm{E}[P_e(C)] \\
&= \Pr\left\{\bigcap_{i=1}^{M}\{d(X,\tilde{Y}_i) > D\}\right\} \\
&\overset{(a)}{\le} 1 - \Pr\{d(X,Y)\le D\}(1 - e^{-\exp(-D_\infty(P_{XY},P_{\tilde{Y}}))M}) \\
&\overset{(b)}{\le} 1 - (1-\delta)(1 - e^{-\exp(-D_\infty(P_{XY},P_{\tilde{Y}}))M}) \\
&= \delta + (1-\delta)e^{-\exp(-D_\infty(P_{XY},P_{\tilde{Y}}))M},
\end{aligned}$$

where (a) comes from Corollary 1, and (b) comes from (9). Thus, for the integer $M > 0$ such that

$$M = \left\lceil \exp(D_\infty(P_{XY}, P_{\tilde{Y}})) \ln\frac{1-\delta}{\epsilon-\delta}\right\rceil, \tag{10}$$

we have

$$\mathrm{E}[P_e(C)] \le \epsilon.$$

Hence, for any $P_{\tilde{Y}} \in \mathcal{P}(\mathcal{Y})$, any $P_{Y|X}$ satisfying (9), and any integer $M > 0$ satisfying (10), there exists an encoder f such that $\|f\| = M$, and $\Pr\{d(X, f(X)) > D\} \le \epsilon$. □

The next lemma shows that $D_\infty(P_{XY}, P_{\tilde{Y}})$ can be optimized with respect to $P_{\tilde{Y}} \in \mathcal{P}(\mathcal{Y})$.

Lemma 2.

$$\inf_{P_{\tilde{Y}}\in\mathcal{P}(\mathcal{Y})} D_\infty(P_{XY}, P_{\tilde{Y}}) = \ln\mu(P_{XY}). \tag{11}$$

Proof. We have

$$\begin{aligned}
\inf_{P_{\tilde{Y}}\in\mathcal{P}(\mathcal{Y})} D_\infty(P_{XY}, P_{\tilde{Y}}) &= \inf_{P_{\tilde{Y}}\in\mathcal{P}(\mathcal{Y})} \ln\sup_{y\in\mathcal{Y}} \frac{\sup_{x:P_X(x)>0} P_{Y|X}(y|x)}{P_{\tilde{Y}}(y)} \\
&\le \ln\sup_{y\in\mathcal{Y}} \frac{\sup_{x:P_X(x)>0} P_{Y|X}(y|x)}{\frac{\sup_{x:P_X(x)>0} P_{Y|X}(y|x)}{\mu(P_{XY})}} \\
&= \ln\mu(P_{XY}),
\end{aligned} \tag{12}$$

where the inequality comes from the fact that $\frac{\sup_{x:P_X(x)>0} P_{Y|X}(y|x)}{\mu(P_{XY})} \in \mathcal{P}(\mathcal{Y})$, i.e.,

$$\frac{\sup_{x:P_X(x)>0} P_{Y|X}(y|x)}{\mu(P_{XY})} \ge 0, \quad \forall y \in \mathcal{Y},$$

and

$$\sum_{y\in\mathcal{Y}} \frac{\sup_{x:P_X(x)>0} P_{Y|X}(y|x)}{\mu(P_{XY})} = 1.$$

On the other hand, we have for any $P_{\tilde{Y}} \in \mathcal{P}(\mathcal{Y})$,

$$\ln\sup_{y\in\mathcal{Y}} \frac{\sup_{x:P_X(x)>0} P_{Y|X}(y|x)}{P_{\tilde{Y}}(y)} \ge \ln\mu(P_{XY}), \tag{13}$$

where the inequality comes from the fact [12, Lemma 16.7.1] that for non-negative real valued functions $f(y)$ and $g(y)$,

$$\sup_{y\in\mathcal{Y}} \frac{f(y)}{g(y)} \ge \frac{\sum_{y\in\mathcal{Y}} f(y)}{\sum_{y\in\mathcal{Y}} g(y)},$$

and settings $f(y) = \sup_{x:P_X(x)>0} P_{Y|X}(y|x)$ and $g(y) = P_{\tilde{Y}}(y)$. Thus, by taking infimum on both sides (13) over $P_{\tilde{Y}} \in \mathcal{P}(\mathcal{Y})$, we have the opposite inequality

$$\inf_{P_{\tilde{Y}}\in\mathcal{P}(\mathcal{Y})} D_\infty(P_{XY}, P_{\tilde{Y}}) \ge \ln\mu(P_{XY}). \tag{14}$$

By combining (12) and (14), we have (11). □

Remark 3. The bound in [7, Theorem 21] was not optimized with respect to the corresponding probability of the above $P_{\tilde{Y}}$. This might be a reason why, in later numerical examples, our achievability bound is tighter than their bound.

Now, we prove Theorem 2.

Proof of Theorem 2. By using the optimal $P_{\tilde{Y}} \in \mathcal{P}(\mathcal{Y})$ that

achieves the equality (11) to Theorem 5, we can show for any $P_{Y|X} \in \mathcal{W}(\mathcal{Y}|\mathcal{X})$ satisfying $\Pr\{d(X,Y) > D\} \le \delta$, the existence of an (M, D, ϵ) code such that

$$M \le \left\lceil \mu(P_{XY}) \ln \frac{1-\delta}{\epsilon-\delta} \right\rceil.$$

Hence, by choosing $P_{Y|X} \in \mathcal{W}(\mathcal{Y}|\mathcal{X})$ minimizing the RHS of this inequality, we have the theorem. □

4.2 Converse Bound

Finally, we prove Theorem 3.

Proof of Theorem 3. For an (M, D, ϵ) code, we define the conditional distribution $P_{\hat{Y}|X} \in \mathcal{W}(\mathcal{Y}|\mathcal{X})$ as

$$P_{\hat{Y}|X}(y|x) = \begin{cases} 1 & \text{if } y = f(x), \\ 0 & \text{otherwise.} \end{cases}$$

Then, by letting $P_{X\hat{Y}}(x,y) = P_X(x)P_{\hat{Y}|X}(y|x)$, we have

$$\begin{aligned} \lceil \mu(P_{X\hat{Y}}) \rceil &= \sum_{y\in\mathcal{Y}} \sup_{x:P_X(x)>0} 1\{y = f(x)\} \\ &\le \sum_{y\in\mathcal{Y}} \sup_{x\in\mathcal{X}} 1\{y = f(x)\} \\ &= M. \end{aligned}$$

On the other hand, by the definition, we have

$$\Pr\{d(X,\hat{Y}) > D\} = \Pr\{d(X, f(X)) > D\} \le \epsilon.$$

Thus, any (M, D, ϵ) code must satisfy

$$\min_{\substack{P_{\hat{Y}|X}\in\mathcal{W}(\mathcal{Y}|\mathcal{X}):\\ \Pr\{d(X,\hat{Y})>D\}\le\epsilon}} \lceil \mu(P_{X\hat{Y}}) \rceil \le M.$$

This completes the proof. □

5. The Limit of the Minimum Rate

In this section, we will deal with n-length source sequences, and characterize the limit (superior) of the minimum rate defined by

$$R^*(D,\epsilon) \triangleq \limsup_{n\to\infty} R(n, D, \epsilon).$$

In the following, by abuse of notation, we treat the n-fold Cartesian product $\mathcal{X}^n$ (resp. $\mathcal{Y}^n$) as the alphabet $\mathcal{X}$ (resp. $\mathcal{Y}$). We will denote an n-length sequence of symbols $(a_1, a_2, \cdots, a_n)$ by a^n. By considering the n-length source sequence X^n (resp. Y^n) as the source symbol X (resp. Y), and the distortion measure d_n for each blocklength n as the distortion measure d, our main results can be directly applied to n-length source sequences. Then, we have the next upper and lower bounds to $R^*(D,\epsilon)$.

Theorem 6. For any given source $\{X^n\}$, sequence of distortion measures $\{d_n\}$, constant $D \ge 0$, and $0 \le \delta < \epsilon \le 1$, we have

$$\rho(D,\epsilon) \le R^*(D,\epsilon) \le \rho(D,\delta),$$

where

$$\begin{aligned} \rho(D,\epsilon) &\triangleq \limsup_{n\to\infty} \frac{1}{n} \log_2 \min_{\substack{P_{Y^n|X^n}\in\mathcal{W}(\mathcal{Y}^n|\mathcal{X}^n):\\ \Pr\{d_n(X^n,Y^n)>D\}\le\epsilon}} \mu(P_{X^nY^n}) \\ &= \limsup_{n\to\infty} \min_{\substack{P_{Y^n|X^n}\in\mathcal{W}(\mathcal{Y}^n|\mathcal{X}^n):\\ \Pr\{d_n(X^n,Y^n)>D\}\le\epsilon}} \frac{1}{n} I_\infty(X^n; Y^n). \end{aligned}$$

Since this theorem can be easily proved by Theorems 2 and 3, we omit the proof.

By using this theorem, we can characterize $R^*(D,\epsilon)$ for some special cases as shown in the following corollaries.

Corollary 2. When $\rho(D,\epsilon)$ is upper semicontinuous (cf. e.g. [16] or [17]) at $\epsilon \in (0,1)$, i.e., $\limsup_{\delta\to\epsilon} \rho(D,\delta) \le \rho(D,\epsilon)$, we have

$$R^*(D,\epsilon) = \rho(D,\epsilon) = \lim_{\delta\uparrow\epsilon} \rho(D,\delta).$$

Corollary 3. When $R^*(D,\epsilon)$ is upper semicontinuous at $\epsilon \in (0,1)$, we have,

$$R^*(D,\epsilon) = \lim_{\delta\uparrow\epsilon} \rho(D,\delta). \tag{15}$$

When $R^*(D,\epsilon)$ is lower semicontinuous (cf. e.g. [16] or [17]) at $\epsilon \in (0,1)$, i.e., $\liminf_{\delta\to\epsilon} R^*(D,\delta) \ge R^*(D,\epsilon)$, we have,

$$R^*(D,\epsilon) = \rho(D,\epsilon). \tag{16}$$

Corollary 4. When the source satisfies the strong converse property [3], i.e.,

$$R^*(D,\epsilon) = \lim_{\delta\downarrow 0} R^*(D,\delta), \quad \forall \epsilon \in (0,1),$$

we have for any $\epsilon \in (0,1)$,

$$R^*(D,\epsilon) = \rho(D,\epsilon) = \lim_{\delta\uparrow\epsilon} \rho(D,\delta).$$

Proof. According to the assumption of this corollary, $R^*(D,\epsilon)$ is lower and upper semicontinuous at $\epsilon \in (0,1)$. Hence, the corollary follows from Corollary 3. □

Remark 4. This corollary holds for i.i.d. sources, because they satisfy the strong converse property (cf. [3] or [7])

Since Corollaries 2 and 3 can be proved easily from a simple property of semicontinuity, we give proofs of these corollaries in Appendix A.

In the above theorem, we attempt to characterize the limit of the minimum rate of codes such that the error probability must be less than ϵ for every blocklength n. Since this condition is somewhat rigid, we fail to characterize the limit in general. So, instead of characterizing $R^*(D,\epsilon)$

directly, we consider "pessimistic" and "optimistic" limits, i.e., $\limsup_{\delta\to\epsilon} R^*(D,\delta)$ and $\liminf_{\delta\to\epsilon} R^*(D,\delta)$, respectively.

Theorem 7. For any given source $\{X^n\}$, sequence of distortion measures $\{d_n\}$, constant $D \geq 0$, and $0 \leq \epsilon \leq 1$, we have

$$\limsup_{\delta\to\epsilon} R^*(D,\delta) = \lim_{\delta\uparrow\epsilon} \rho(D,\delta), \tag{17}$$

$$\liminf_{\delta\to\epsilon} R^*(D,\delta) = \lim_{\delta\downarrow\epsilon} \rho(D,\delta). \tag{18}$$

Since this theorem can be also proved easily from a simple property of semicontinuity, we give the proof in Appendix A.

Remark 5. The RD function $R^*(D)$ can be defined by (cf. [3])

$$R^*(D) \triangleq \lim_{\delta\downarrow 0} R^*(D,\delta).$$

According to Lemma 8 (in Appendix A) and Theorem 7, the RD function can be characterized as

$$R^*(D) = \lim_{\delta\downarrow 0} \rho(D,\delta).$$

6. Polynomial-Time Computable Achievability Bounds for i.i.d. Binary Sources

In order to demonstrate the tightness of our results, we show numerical examples of our achievability bound for i.i.d. sources.

In this section, we once again deal with n-length source sequences. Especially, we consider the binary case $\mathcal{X}^n = \mathcal{Y}^n = \{0,1\}^n$, and the i.i.d. binary source sequence X^n on $\mathcal{X}^n$ such that $P_{X^n}(x^n) = \prod_{i=1}^n P_X(x_i)$, where $P_X(0) = 1-p$ and $P_X(1) = p$ for a certain constant $0 \leq p \leq 1$. Throughout this section, we assume that $0 \leq D < p \leq 1/2$.

We use the Hamming distance par blocklength as a distortion measure, i.e., $d(x^n,y^n) = \frac{1}{n} d_{\mathrm{H}}^n(x^n,y^n)$, where d_{H}^n is the Hamming distance defined as $d_{\mathrm{H}}^n(x^n,y^n) \triangleq \sum_{i=1}^n 1\{x_i \neq y_i\}$. For $a^n \in \{0,1\}^n$, let $w(a^n)$ be the Hamming weight of a^n, i.e., $w(a^n) = d_{\mathrm{H}}^n(a^n,0^n)$, where 0^n denotes the n-length all-zero sequence. Then, for $a^n \in \{0,1\}^n$ and $r \geq 0$, we define the ball $\mathcal{B}_r(a^n)$ of radius r centered at a^n as $\mathcal{B}_r(a^n) \triangleq \{b^n \in \{0,1\}^n : d_{\mathrm{H}}^n(a^n,b^n) \leq r\}$, and define the sphere $\mathcal{S}_r$ of the ball $\mathcal{B}_r(0^n)$ as $\mathcal{S}_r \triangleq \{a^n \in \{0,1\}^n : w(a^n) = r\}$.

Since our achievability bound involves optimization problem over conditional probability distributions, it is hard to compute directly in general. Hence, instead of finding the optimal distribution, we fix a conditional distribution, and use a somewhat loose bound. Specifically, instead of using the bound of Theorem 2 (i.e., the following (19)), we use the following loose bound (20) for a fixed conditional distribution $\tilde{P}_{Y^n|X^n} \in \{P_{Y^n|X^n} \in \mathcal{W}(\mathcal{Y}^n|\mathcal{X}^n) : \Pr\{d(X^n,Y^n) > D\} \leq \delta\}$:

$$M^*(D,\epsilon) \leq \min_{\substack{P_{Y^n|X^n}\in\mathcal{W}(\mathcal{Y}^n|\mathcal{X}^n):\\ \Pr\{d(X^n,Y^n)>D\}\leq\delta}} \left\lceil \mu(P_{X^nY^n}) \ln \frac{1-\delta}{\epsilon-\delta} \right\rceil \tag{19}$$

$$\leq \left\lceil \mu(P_{X^n}\cdot\tilde{P}_{Y^n|X^n}) \ln \frac{1-\delta}{\epsilon-\delta} \right\rceil, \tag{20}$$

where $P_{X^n}\cdot\tilde{P}_{Y^n|X^n} \in \mathcal{P}(\mathcal{X}^n\times\mathcal{Y}^n)$ is a joint probability distribution such that $P_{X^n}\cdot\tilde{P}_{Y^n|X^n}(x^n,y^n) = P_{X^n}(x^n)\tilde{P}_{Y^n|X^n}(y^n|x^n)$. Although this bound (20) does not involve the optimization problem, it may be looser than that of Theorem 2 and it still needs to compute $\mu(P_{X^n}\cdot\tilde{P}_{Y^n|X^n})$. Hence, we need to choose a "good" distribution $\tilde{P}_{Y^n|X^n}$ such that the above bound (20) becomes tight and $\mu(P_{X^n}\cdot\tilde{P}_{Y^n|X^n})$ can be easily computed.

In the following two subsections, we will give two such good distributions. But before we give them, we consider two simple distributions

$$\tilde{P}^{(1)}_{Y^n|X^n}(y^n|x^n) = 1\{y^n = x^n\},$$

$$\tilde{P}^{(2)}_{Y^n|X^n}(y^n|x^n) = \frac{1\{d(x^n,y^n)\leq D\}}{\sum_{y^n\in\mathcal{Y}^n} 1\{d(x^n,y^n)\leq D\}}$$

for a better understanding of the bound (20). Since these distributions satisfy $\Pr\{d(X^n,Y^n) > D\} = 0$, these can be used to the bound (20). Then, one can easily check that

$$\mu(P_{X^n}\cdot\tilde{P}^{(1)}_{Y^n|X^n}) = 2^n, \tag{21}$$

$$\mu(P_{X^n}\cdot\tilde{P}^{(2)}_{Y^n|X^n}) = \frac{2^n}{\left\langle {n \atop \lfloor nD\rfloor} \right\rangle}, \tag{22}$$

where we use the notation

$$\left\langle {n \atop k} \right\rangle \triangleq \sum_{i=0}^{k} \binom{n}{i}$$

and the fact that $\sum_{y^n\in\mathcal{Y}^n} 1\{d(x^n,y^n)\leq D\} = \left\langle {n \atop \lfloor nD\rfloor} \right\rangle$, where $\lfloor\cdot\rfloor$ denotes the floor function, and $\binom{n}{i}$ denotes the binomial coefficient. These can be easily computed, but the bound using (21) is trivial, and the bound using (22) does not depend on the probability distribution of the i.i.d source. Hence, these distributions give loose bounds in general.

6.1 Uniform Weighting Distribution

For the probability distribution P_{X^n} of an i.i.d. binary source, and the equiprobable distribution $U_{\mathcal{X}^n}$ on $\mathcal{X}^n = \{0,1\}^n$, i.e., $U_{\mathcal{X}^n}(x^n) = \frac{1}{2^n}$ for all $x^n \in \mathcal{X}^n$, Kostina and Verdú [7] showed that $\beta_{1-\delta}(P_{X^n}, U_{\mathcal{X}^n})$ can be optimized by the function $\phi^* : \mathcal{X}^n \to [0,1]$ such that

$$\phi^*(x^n) = \begin{cases} 1 & \text{if } w(x^n) \leq r^*_\delta, \\ \alpha_\delta & \text{if } w(x^n) = r^*_\delta + 1, \\ 0 & \text{if } w(x^n) \geq r^*_\delta + 2, \end{cases}$$

where

$$r^*_\delta = \max\left\{ r : \sum_{k=0}^{r} \binom{n}{k} p^k (1-p)^{n-k} \leq 1-\delta \right\},$$

and $\alpha_\delta \in [0,1)$ is a solution to

$$\sum_{k=0}^{r_\delta^*}\binom{n}{k}p^k(1-p)^{n-k}+\alpha_\delta\binom{n}{r_\delta^*+1}p^{r_\delta^*+1}(1-p)^{n-r_\delta^*-1}=1-\delta.$$

On the other hand, let us recall that, in the proof of Theorem 4, we set

$$\phi^*(x^n)=\sum_{y^n\in\mathcal{Y}^n}P_{Y^n|X^n}(y^n|x^n)1\{d(x^n,y^n)\le D\}.$$

This implies that $P_{Y^n|X^n}$ satisfying this equality may be a good distribution of our achievability bound. For this reason, we consider the conditional probability distribution $\tilde{P}^{[\mathrm{u}]}_{Y^n|X^n}\in\mathcal{W}(\mathcal{Y}^n|\mathcal{X}^n)$ defined as

$$\begin{aligned}&\tilde{P}^{[\mathrm{u}]}_{Y^n|X^n}(y^n|x^n)\\&=\begin{cases}\dfrac{1\{d(x^n,y^n)\le D\}}{\left\langle{n\atop\lfloor nD\rfloor}\right\rangle} & \text{if } w(x^n)\le r_\delta^*,\\ \alpha_\delta\dfrac{1\{d(x^n,y^n)\le D\}}{\left\langle{n\atop\lfloor nD\rfloor}\right\rangle}\\ \quad+(1-\alpha_\delta)\dfrac{1\{d(x^n_{\max},y^n)\le D\}}{\left\langle{n\atop\lfloor nD\rfloor}\right\rangle} & \text{if } w(x^n)=r_\delta^*+1,\\ \dfrac{1\{d(x^n_{\max},y^n)\le D\}}{\left\langle{n\atop\lfloor nD\rfloor}\right\rangle} & \text{if } w(x^n)\ge r_\delta^*+2,\end{cases}\end{aligned}\tag{23}$$

where

$$x^n_{\max}\triangleq\operatorname*{argmax}_{\tilde{x}^n\in\mathcal{X}^n:w(\tilde{x}^n)\le\lfloor nD\rfloor}d(x^n,\tilde{x}^n).\tag{24}$$

If there are more than one sequences maximizing $d(x^n,\tilde{x}^n)$, choose any one of them as $x^n_{\max}$. Note that $x^n_{\max}$ is a function of x^n.

According to Lemma 3 given below, this distribution is well-defined when $r_\delta^*\ge\lfloor nD\rfloor$. When $r_\delta^*+1\le\lfloor nD\rfloor$ for $0\le\delta<\epsilon\le1$, we employ the encoder that encodes all sequences $x^n\in\mathcal{X}^n$ into 0^n. Then, we have $\|f\|=1$ and

$$\Pr\{d(X^n,f(X^n))\le D\}\ge\sum_{k=0}^{r_\delta^*+1}\binom{n}{k}p^k(1-p)^{n-k}\ge1-\epsilon.$$

This implies that $M^*(D,\epsilon)=1$. Hence, our main interest is the case where $r_\delta^*+1>\lfloor nD\rfloor$, i.e., $r_\delta^*\ge\lfloor nD\rfloor$.

From a viewpoint of the usual random coding method (which we also use to give the achievability bound), a conditional distribution $P_{Y^n|X^n}$ plays a role of weighting to candidates of reproduction sequences for a given source sequence. Intuitively, the above conditional distribution $\tilde{P}^{[\mathrm{u}]}_{Y^n|X^n}$ uniformly maps each high probability source sequence to reproduction sequences within the distortion D, and each low probability source sequence to reproduction sequences near the origin 0^n. In other words, this distribution uniformly weights all reproduction sequences within the distortion D for possible source sequences.

By employing this distribution to the bound (20), we have the next upper bound for i.i.d. sources.

Theorem 8. Let $P_{X^n}(x^n)=\prod_{i=1}^nP_X(x_i)$ be the probability distribution for the i.i.d. binary source, where $P_X(1)=p$. Then, for $0\le D<p\le\frac{1}{2}$, and $0\le\delta<\epsilon\le1$, if $r_\delta^*\ge\lfloor nD\rfloor$, we have

$$M^*(D,\epsilon)\le\left\lceil\frac{\left\langle{n\atop r_\delta^*+\lfloor nD\rfloor}\right\rangle+\alpha_\delta\left\langle{n\atop r_\delta^*+1+\lfloor nD\rfloor}\right\rangle}{\left\langle{n\atop\lfloor nD\rfloor}\right\rangle}\ln\frac{1-\delta}{\epsilon-\delta}\right\rceil.$$

Remark 6. Since the time complexity of binomial coefficient $\binom{n}{k}$ is $O(nk)$, the time complexity of $\left\langle{n\atop k}\right\rangle$ is $O(nk^2)$. Hence, this bound can be computed in polynomial time at most $O(n^3)$ because time complexity of each component of the bound is at most $O(n^3)$.

Remark 7. By using the same argument of the proof of [7, Theorem 23], the RHS of the above inequality will be asymptotically characterized by $\exp(nh(p+D)-nh(D)+o(n))$. Thus, this bound is not asymptotically optimal, but as will be later shown, there exist some cases where it is tight for small blocklengths.

To prove Theorem 8, we use the following two lemmas. We give proofs of these lemmas in Appendix B.

Lemma 3. Suppose that $r_\delta^*\ge\lfloor nD\rfloor$. Then, for any given $(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n$ with $w(x^n)\ge r_\delta^*+1$, if $d(x^n_{\max},y^n)\le D$, then we have $d(x^n,y^n)>D$.

Lemma 4. Suppose that $r_\delta^*\ge\lfloor nD\rfloor$. Then, we have

$$\begin{aligned}&\sup_{x^n\in\mathcal{X}^n}\tilde{P}^{[\mathrm{u}]}_{Y^n|X^n}(y^n|x^n)\\&=\begin{cases}\dfrac{1}{\left\langle{n\atop\lfloor nD\rfloor}\right\rangle} & \text{if } w(y^n)\le r_\delta^*+\lfloor nD\rfloor, &(25)\\ \dfrac{\alpha_\delta}{\left\langle{n\atop\lfloor nD\rfloor}\right\rangle} & \text{if } w(y^n)=r_\delta^*+1+\lfloor nD\rfloor, &(26)\\ 0 & \text{if } w(y^n)\ge r_\delta^*+2+\lfloor nD\rfloor. &(27)\end{cases}\end{aligned}$$

Now, we prove Theorem 8.

Proof of Theorem 8. For the distribution $\tilde{P}^{[\mathrm{u}]}_{Y^n|X^n}$ defined by (23), we have

$$\begin{aligned}&\Pr\{d(X^n,Y^n)\le D\}\\&=\sum_{x^n\in\mathcal{X}^n}P_{X^n}(x^n)\sum_{y^n\in\mathcal{Y}^n:d(x^n,y^n)\le D}\tilde{P}^{[\mathrm{u}]}_{Y^n|X^n}(y^n|x^n)\\&\overset{(a)}{=}\sum_{x^n\in\mathcal{X}^n:w(x^n)\le r_\delta^*}P_{X^n}(x^n)\sum_{y^n\in\mathcal{Y}^n:d(x^n,y^n)\le D}\frac{1}{\left\langle{n\atop\lfloor nD\rfloor}\right\rangle}\\&\quad+\sum_{x^n\in\mathcal{X}^n:w(x^n)=r_\delta^*+1}P_{X^n}(x^n)\sum_{y^n\in\mathcal{Y}^n:d(x^n,y^n)\le D}\frac{\alpha_\delta}{\left\langle{n\atop\lfloor nD\rfloor}\right\rangle}\\&=\sum_{x^n\in\mathcal{X}^n:w(x^n)\le r_\delta^*}P_{X^n}(x^n)+\alpha_\delta\sum_{x^n\in\mathcal{X}^n:w(x^n)=r_\delta^*+1}P_{X^n}(x^n)\\&=1-\delta,\end{aligned}$$

where (a) comes from Lemma 3.

On the other hand, we have

$$\begin{aligned}&\sum_{y^n\in\mathcal{Y}^n}\sup_{x^n\in\mathcal{X}^n}\tilde{P}^{[\mathrm{u}]}_{Y^n|X^n}(y^n|x^n)\\&=\sum_{y^n\in\mathcal{Y}^n:w(y^n)\le r^*_\delta+\lfloor nD\rfloor}\frac{1}{\left\langle{n\atop\lfloor nD\rfloor}\right\rangle}+\sum_{y^n\in\mathcal{Y}^n:w(y^n)=r^*_\delta+1+\lfloor nD\rfloor}\frac{\alpha_\delta}{\left\langle{n\atop\lfloor nD\rfloor}\right\rangle}\\&=\frac{\left\langle{n\atop r^*_\delta+\lfloor nD\rfloor}\right\rangle+\alpha_\delta\left\langle{n\atop r^*_\delta+1+\lfloor nD\rfloor}\right\rangle}{\left\langle{n\atop\lfloor nD\rfloor}\right\rangle},\end{aligned}$$

where the first equality comes from Lemma 4. Thus, by using the bound (20), we have the theorem. □

6.2 Spherical Weighting Distribution

Let r_δ be an integer such that

$$\sum_{k=0}^{r_\delta}\binom{n}{k}p^k(1-p)^{n-k}\ge 1-\delta, \tag{28}$$

and w_δ be an integer such that

$$w_\delta=\max\{\lfloor nD\rfloor,\min\{r_\delta,\lfloor(1/D-1)\lfloor nD\rfloor\rfloor-1\}\}.$$

Then, we employ the next distribution as a good distribution.

$$\begin{aligned}&\tilde{P}^{[\mathrm{s}]}_{Y^n|X^n}(y^n|x^n)\\&=\begin{cases}\dfrac{1\{y^n\in\mathcal{A}^{k(w(x^n),D)}_{\lfloor nD\rfloor}(x^n)\}}{|\mathcal{A}^{k(w(x^n),D)}_{\lfloor nD\rfloor}(x^n)|} & \text{if } \lfloor nD\rfloor<w(x^n)\le w_\delta,\\ 1\{y^n=x^n\} & \text{if } w_\delta<w(x^n)\le r_\delta,\\ 1\{y^n=0^n\} & \text{otherwise,}\end{cases}\end{aligned} \tag{29}$$

where $\mathcal{A}^k_w(x^n)\triangleq\mathcal{B}_w(x^n)\cap\mathcal{S}_k$,

$$k(r,D)\triangleq\begin{cases}f(r,D) & \text{if } f(r,D)=r-\lfloor nD\rfloor \pmod 2,\\ f(r,D)+1 & \text{otherwise,}\end{cases}$$

and $f(r,D)\triangleq\left\lfloor\frac{r-\lfloor nD\rfloor}{1-2D}\right\rfloor$.

Intuitively, this conditional distribution uniformly maps each high probability source sequence to reproduction sequences within the distortion D on the partial sphere $\mathcal{A}^k_w(x^n)$, and each low probability source sequence to the origin 0^n. Thus, for possible source sequences, this distribution uniformly weights all reproduction sequences in the partial sphere $\mathcal{A}^k_w(x^n)$. Dumer et al. [18] used this kind of distribution to derive an upper bound to the minimum number of balls of radius r covering a ball of radius $s(>r)$, and they showed the bound is asymptotically tight.

For $\mathcal{A}^k_w(x^n)$ defined above, we have the following two lemmas.

Lemma 5. For integers $r,k,w\in\{0,1,\cdots,n\}$, and any $x^n\in\mathcal{S}_r$, we have

$$|\mathcal{A}^k_w(x^n)|=A^k_w(r,n), \tag{30}$$

where

$$A^k_w(r,n)\triangleq\begin{cases}0 & \text{if } \max\{0,r-k\}>\left\lfloor\frac{w+r-k}{2}\right\rfloor,\\ \displaystyle\sum_{i=\max\{0,r-k\}}^{\min\{r,n-k,\lfloor\frac{w+r-k}{2}\rfloor\}}\binom{r}{i}\binom{n-r}{i+k-r} & \text{otherwise.}\end{cases}$$

Lemma 6. $\mathcal{A}^{k(w(x^n),D)}_{\lfloor nD\rfloor}(x^n)$ is not empty for any $x^n\in\{0,1\}^n$ satisfying $\lfloor nD\rfloor<w(x^n)\le\lfloor(1/D-1)\lfloor nD\rfloor\rfloor-1$.

We give proofs of these lemmas in Appendix C. We note that, according to Lemma 6, distribution $\tilde{P}^{[\mathrm{s}]}_{Y^n|X^n}$ is well-defined because $|\mathcal{A}^{k(w(x^n),D)}_{\lfloor nD\rfloor}(x^n)|\ge 1$ if x^n satisfies $\lfloor nD\rfloor<w(x^n)\le w_\delta$.

By employing the above distribution $\tilde{P}^{[\mathrm{s}]}_{Y^n|X^n}$ to the bound (20), we have the next upper bound for i.i.d. sources.

Theorem 9. Let $P_{X^n}(x^n)=\prod_{i=1}^n P_X(x_i)$ be the probability distribution for the i.i.d. binary source, where $P_X(1)=p$. Then, for any $0\le D<p\le\frac{1}{2}$, $0\le\delta<\epsilon\le 1$, and any $n>0$, we have

$$M^*(D,\epsilon)\le\left\lfloor\left(1+\sum_{k=w_\delta+1}^{r_\delta}\binom{n}{k}+\sum_{r=\lfloor nD\rfloor+1}^{w_\delta}\frac{\binom{n}{k(r,D)}}{A^{k(r,D)}_{\lfloor nD\rfloor}(r,n)}\right)\ln\frac{1-\delta}{\epsilon-\delta}\right\rfloor.$$

Remark 8. Since the time complexity of binomial coefficient $\binom{n}{k}$ is $O(nk)$, the time complexity of $A^{k(r,D)}_{\lfloor nD\rfloor}(r,n)$ is at most $O(n^3)$. Hence, $\sum_{r=\lfloor nD\rfloor+1}^{w_\delta}\frac{\binom{n}{k(r,D)}}{A^{k(r,D)}_{\lfloor nD\rfloor}(r,n)}$ can be computed in polynomial time at most $O(n^4)$, and also this achievability bound can be computed in the same order.

Remark 9. Since the RHS of the above inequality is almost same as [18, RHS of (14)], it will be asymptotically characterized by $\exp(nh(p)-nh(D)+o(n))$. Thus, this bound may be asymptotically optimal. In fact, for later numerical examples, this bound is very tight.

To prove Theorem 9, we use the following lemma. The proof of the lemma is shown in Appendix D.

Lemma 7. We have

$$\begin{aligned}&\sup_{x^n\in\mathcal{X}^n}\tilde{P}^{[\mathrm{s}]}_{Y^n|X^n}(y^n|x^n)\\&=\begin{cases}1 & \text{if } y^n\in\{0^n\}\cup\mathcal{S}^{r_\delta}_{w_\delta+1},\\ \dfrac{1}{A^{k(r,D)}_{\lfloor nD\rfloor}(r,n)} & \text{if } y^n\in\mathcal{S}_{k(r,D)}\cap\left[\{0^n\}\cup\mathcal{S}^{r_\delta}_{w_\delta+1}\right]^c\\ & \text{for some } r\in\{\lfloor nD\rfloor+1,\cdots,w_\delta\},\\ 0 & \text{otherwise,}\end{cases}\end{aligned}$$

where $[\cdot]^c$ denotes the complement of the set, and $\mathcal{S}^{r_\delta}_{w_\delta+1}=\bigcup_{w_\delta+1\le k\le r_\delta}\mathcal{S}_k$.

Now we prove Theorem 9.

Proof of Theorem 9. For the distribution $\tilde{P}^{[\mathrm{s}]}_{Y^n|X^n}$ defined by (29), we have

$$
\begin{aligned}
&\Pr\{d(X^n, Y^n) \le D\} \\
&\ge \sum_{x^n : w(x^n) \le \lfloor nD \rfloor} P_{X^n}(x^n) + \sum_{x^n : \lfloor nD \rfloor < w(x^n) \le w_\delta} P_{X^n}(x^n) \\
&\quad \times \sum_{y^n \in \mathcal{A}_{\lfloor nD \rfloor}^{k(w(x^n),D)}(x^n) : d(x^n, y^n) \le D} \frac{1}{|\mathcal{A}_{\lfloor nD \rfloor}^{k(w(x^n),D)}(x^n)|} \\
&\quad + \sum_{x^n : w_\delta < w(x^n) \le r_\delta} P_{X^n}(x^n) \\
&= \sum_{x^n \in \mathcal{X}^n : w(x^n) \le r_\delta} P_{X^n}(x^n) \\
&\ge 1 - \delta,
\end{aligned}
$$

where the last inequality comes from the property (28) of r_δ.

On the other hand, according to Lemma 7, we have

$$
\begin{aligned}
&\sum_{y^n \in \mathcal{Y}^n} \sup_{x^n \in \mathcal{X}^n} \tilde{P}^{[s]}_{Y^n|X^n}(y^n|x^n) \\
&\le 1 + \sum_{k=w_\delta+1}^{r_\delta} |\mathcal{S}_k| + \sum_{r=\lfloor nD \rfloor+1}^{w_\delta} \sum_{y^n \in \mathcal{S}_{k(r,D)}} \frac{1}{A_{\lfloor nD \rfloor}^{k(r,D)}(r,n)} \\
&= 1 + \sum_{k=w_\delta+1}^{r_\delta} \binom{n}{k} + \sum_{r=\lfloor nD \rfloor+1}^{w_\delta} \frac{\binom{n}{k(r,D)}}{A_{\lfloor nD \rfloor}^{k(r,D)}(r,n)},
\end{aligned}
$$

where the last equality comes from the fact that $|\mathcal{S}_k| = \binom{n}{k}$. Thus, by using the bound (20), we have the theorem. □

6.3 Numerical Examples

In this section, we show numerical examples for our two polynomial-time computable achievability bounds, and compare our bounds to Kostina and Verdú's achievability and converse bounds [7, Theorems 20 and 21].

Figure 1 shows bounds to the rate $R(n, D, \epsilon)$ for a less biased source, i.e., p is close to a half. As shown in this figure, the achievability bound of Theorem 8 is tighter than other achievability bounds for small blocklengths. Unfortunately, this bound is looser than other bounds for large blocklengths. On the other hand, the achievability bound of Theorem 9 is tighter than other achievability bounds for large blocklengths, where we use $r^*_\delta + 1$ as r_δ.

Figure 2 shows bounds to the rate $R(n, D, \epsilon)$ for a biased source, i.e., p is close to zero. Unlike the previous example, the achievability bound of Theorem 8 is looser than other achievability bounds for almost all blocklengths. On the other hand, the achievability bound of Theorem 9 is very tight for almost all blocklengths.

It is confirmed by several experiments that the achievability bound of Theorem 9 has more advantage over Kostina and Verdú's bound as ϵ becomes smaller. On the other hand, it is also confirmed that for large ϵ (roughly over 0.01), our bound almost coincides to their bound.

In these examples, we set $\delta = 0.9\epsilon$. According to our several experiments, this value tightens our achievability bounds. In fact, when δ is increased or decreased from 0.9ϵ, these bounds become loose in most cases.

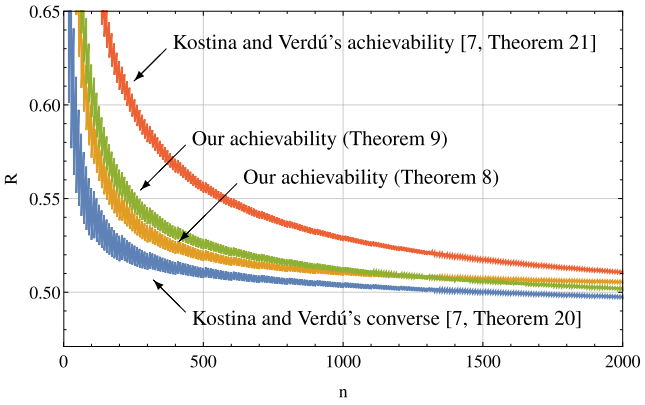

Fig. 1 Bounds to $R(n, D, \epsilon)$ in the case where $p = 0.4$, $D = 0.11$, $\epsilon = 10^{-8}$, and $\delta = 9 \cdot 10^{-9}$.

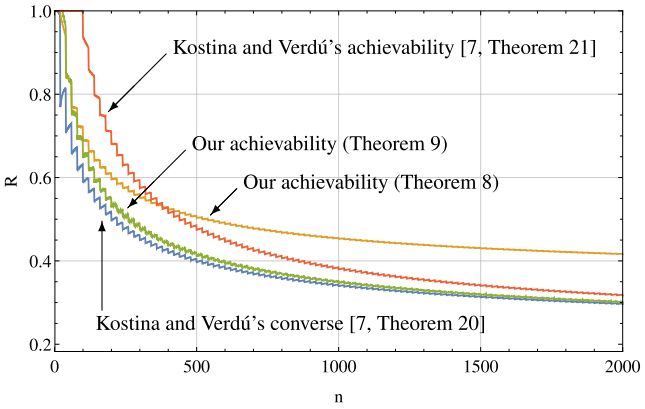

Fig. 2 Bounds to $R(n, D, \epsilon)$ in the case where $p = 0.1$, $D = 0.05$, $\epsilon = 10^{-8}$, and $\delta = 9 \cdot 10^{-9}$.

7. Conclusion

This paper has dealt with the fixed-length lossy compression, and gave achievability bound (Theorem 2) and converse bound (Theorem 3) for the minimum number $M^*(D, \epsilon)$ of codewords by using $\mu(P_{XY})$ and equivalently the α-mutual information of order infinity. By using these bounds, in Corollaries 2–4, we have shown that there exist several cases where $R^*(D, \epsilon)$ is characterized by the α-mutual information of order infinity. On the other hand, in Theorem 7, we have shown that $\limsup_{\delta \to \epsilon} R^*(D, \delta)$ and $\liminf_{\delta \to \epsilon} R^*(D, \delta)$ are completely characterized by the α-mutual information of order infinity. We have also gave numerical examples of two polynomial-time computable achievability bounds for i.i.d. binary sources in Sect. 6. Each of these bounds is induced by our achievability bound (i.e., Theorem 2) by choosing a good conditional distribution. Then, we have shown that there exist cases where these achievability bounds are tighter than that of Kostina and Verdú [7].

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

This work was supported in part by JSPS KAKENHI Grant Number 15K15935.

References

[1] T. Matsuta and T. Uyematsu, "New non-asymptotic achievability and converse bounds for fixed-length lossy compression," Proc. 37th Symp. Inform. Theory and its Apps. (SITA2014), pp.189–194, Dec. 2014.
[2] T. Matsuta and T. Uyematsu, "Non-asymptotic bounds for fixed-length lossy compression," Proc. 2015 IEEE Int. Symp. Inform. Theory (ISIT), pp.1811–1815, 2015.
[3] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, 2nd ed., Cambridge University Press, 2011.
[4] T.S. Han, Information-Spectrum Methods in Information Theory, Stochastic Modelling and Applied Probability, vol.50, Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.
[5] R. Nomura and H. Yagi, "Information spectrum approach to fixed-length lossy source coding problem with some excess distortion probability," Proc. 2015 IEEE Int. Symp. Inform. Theory (ISIT), pp.306–310, 2015.
[6] A. Ingber and Y. Kochman, "The dispersion of lossy source coding," 2011 Data Compression Conference, pp.53–62, 2011.
[7] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," IEEE Trans. Inform. Theory, vol.58, no.6, pp.3309–3338, June 2012.
[8] T. Uyematsu and T. Matsuta, "Revisiting the rate-distortion theory using smooth max Rényi divergence," Proc. 2014 IEEE Inform. Theory Workshop (ITW 2014), pp.202–206, 2014.
[9] N.A. Warsi, "One-shot source coding with coded side information available at the decoder," Proc. 2013 IEEE Int. Symp. Inform. Theory, pp.3070–3074, 2013.
[10] S. Verdú, "α-mutual information," 2015 Information Theory and Applications Workshop (ITA), pp.1–6, 2015.
[11] S.-W. Ho and S. Verdú, "Convexity/concavity of Rényi entropy and α-mutual information," Proc. 2015 IEEE Int. Symp. Inform. Theory (ISIT), pp.745–749, 2015.
[12] T.M. Cover and J.A. Thomas, Elements of Information Theory, 2nd ed., Wiley, New York, 2006.
[13] S. Verdú, "Non-asymptotic achievability bounds in multiuser information theory," Proc. 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp.1–8, 2012.
[14] A. El Gamal and Y.H. Kim, Network Information Theory, Cambridge University Press, 2011.
[15] A. Rényi, "On measures of entropy and information," Proc. 4th Berkeley Symp. Math Stat. Prob., pp.547–561, 1960.
[16] R.G. Bartle, A Modern Theory of Integration, Graduate Studies in Mathematics, vol.32, American Mathematical Society, Providence, Rhode Island, 2001.
[17] M. Giaquinta and G. Modica, Mathematical Analysis: Linear and Metric Structures and Continuity, Birkhäuser, 2007.
[18] I. Dumer, M.S. Pinsker, and V.V. Prelov, "On the thinnest coverings of spheres and ellipsoids with balls in Hamming and Euclidean spaces," General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, vol.4123, pp.898–925, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

Appendix A:

In this appendix, we give proofs of Corollary 2, Corollary 3, and Theorem 7.

Before we give the proofs, we show the next fundamental lemma.

Lemma 8. For a real-valued non-increasing function f, we have

$$\limsup_{x\to x_0} f(x) = \lim_{x\uparrow x_0} f(x), \tag{A·1}$$

$$\liminf_{x\to x_0} f(x) = \lim_{x\downarrow x_0} f(x), \tag{A·2}$$

where

$$\limsup_{x\to x_0} f(x) = \lim_{\epsilon\downarrow 0} \sup\{f(x) : 0 < |x - x_0| < \epsilon\},$$
$$\liminf_{x\to x_0} f(x) = \lim_{\epsilon\downarrow 0} \inf\{f(x) : 0 < |x - x_0| < \epsilon\}.$$

Proof. From the definition of the limit superior, we have

$$\begin{aligned}\limsup_{x\to x_0} f(x) &= \lim_{\epsilon\downarrow 0} \sup\{f(x) : x \in (x_0 - \epsilon, x_0 + \epsilon)\backslash\{x_0\}\} \\ &= \lim_{\epsilon\downarrow 0} \sup\{f(x) : x \in (x_0 - \epsilon, x_0 - \epsilon/2)\}.\end{aligned}$$

where the second equality comes from the fact that f is a non-increasing function, and

$$\begin{aligned}&\sup\{f(x) : x \in (x_0 - \epsilon, x_0 + \epsilon)\backslash\{x_0\}\} \\ &\quad = \sup\{f(x) : x \in (x_0 - \epsilon, x_0 - \epsilon/2)\}.\end{aligned}$$

On the other hand, we have

$$\begin{aligned}f(x_0 - \epsilon/2) &\le \sup\{f(x) : x \in (x_0 - \epsilon, x_0 - \epsilon/2)\} \\ &\le f(x_0 - \epsilon).\end{aligned}$$

By taking the right hand limit on both sides, we have

$$\lim_{\epsilon\downarrow 0} \sup\{f(x) : x \in (x_0 - \epsilon, x_0 - \epsilon/2)\} = \lim_{\epsilon\downarrow 0} f(x_0 - \epsilon).$$

This completes the proof of (A·1). The proof of (A·2) can be done similarly. □

Now, we give proofs of Corollaries 2 and 3.

Proof of Corollary 2. Since $\rho(D, \epsilon)$ is upper semicontinuous at ϵ, we have

$$\limsup_{\delta\to\epsilon} \rho(D, \delta) \le \rho(D, \epsilon) \le R^*(D, \epsilon) \le \lim_{\delta\uparrow\epsilon} \rho(D, \delta), \tag{A·3}$$

where the left-most inequality comes from the definition of the upper semicontinuity, and right two inequalities come from Theorem 6. On the other hand, since $\rho(D, \epsilon)$ is a non-increasing function for $\epsilon \in (0, 1)$, we have

$$\lim_{\delta\uparrow\epsilon} \rho(D, \delta) = \limsup_{\delta\to\epsilon} \rho(D, \delta), \tag{A·4}$$

where the equality comes from Lemma 8. By combining (A· 3) and (A· 4), we have the corollary. □

Proof of Corollary 3. First, we show (15). According to Theorem 6, we have

$$R^*(D,\epsilon) \le \lim_{\delta\uparrow\epsilon}\rho(D,\delta) \le \lim_{\delta\uparrow\epsilon} R^*(D,\delta). \tag{A· 5}$$

On the other hand, since $R^*(D,\epsilon)$ is upper semicontinuous at ϵ and a non-increasing function, we have

$$\lim_{\delta\uparrow\epsilon} R^*(D,\delta) = \limsup_{\delta\to\epsilon} R^*(D,\delta) \le R^*(D,\epsilon), \tag{A· 6}$$

where this equality comes from Lemma 8, and the inequality comes from the definition of the upper semicontinuity. Now, by combining (A· 5) and (A· 6), we have (15).

Next, we show (16). According to Theorem 6, we have

$$\lim_{\delta\downarrow\epsilon} R^*(D,\delta) \le \rho(D,\epsilon) \le R^*(D,\epsilon). \tag{A· 7}$$

On the other hand, since $R^*(D,\epsilon)$ is lower semicontinuous at ϵ and a non-increasing function, we have

$$R^*(D,\epsilon) \le \liminf_{\delta\to\epsilon} R^*(D,\delta) = \lim_{\delta\downarrow\epsilon} R^*(D,\delta), \tag{A· 8}$$

where the equality comes from Lemma 8, and the inequality comes from the definition of the lower semicontinuity. Now, by combining (A· 7) and (A· 8), we have (16). □

The proof of Theorem 7 is as follows.

Proof of Theorem 7. Since $R^*(D,\delta)$ is a non-increasing function for δ, we have

$$\begin{aligned}\limsup_{\delta\to\epsilon} R^*(D,\delta) &= \lim_{\delta\uparrow\epsilon} R^*(D,\delta)\\ &= \lim_{\gamma\downarrow 0} R^*(D,\epsilon-\gamma).\end{aligned}$$

On the other hand, according to Theorem 6, we have for any $\gamma > 0$,

$$\rho(D,\epsilon-\gamma) \le R^*(D,\epsilon-\gamma) \le \rho(D,\epsilon-2\gamma).$$

Now, by taking the right hand limit on both sides, we have

$$\lim_{\gamma\downarrow 0} R^*(D,\epsilon-\gamma) = \lim_{\gamma\downarrow 0}\rho(D,\epsilon-\gamma).$$

This implies (17).

The equality (18) can be proved in a similar way. □

Appendix B:

In this appendix, we give proofs of Lemmas 3 and 4.

In the proofs of the lemmas, for a sequence $x^n \in \{0,1\}^n$ and a permutation $\rho : \{1,2,\cdots,n\} \to \{1,2,\cdots,n\}$, we denote the permuted sequence of x^n as $\rho\circ x^n$, i.e., $\rho\circ x^n = (x_{\rho(1)}, x_{\rho(2)}, \cdots, x_{\rho(n)})$.

Now, we give proofs of the lemmas.

Proof of Lemma 3. Let $w(x^n) = \beta\,(\ge r^*_\delta + 1)$. For x^n, let ρ be the permutation such that $\rho\circ x^n = (\underbrace{1,\cdots,1}_{\beta},0,\cdots,0)$. In other words, ρ permutes x^n so that the first β elements are all 1. On the other hand, if $\beta \le 2\lfloor nD\rfloor + 1$, we consider the sequence $\tilde{x}^n \in \mathcal{X}^n$ such that

$$\tilde{x}^n = (\underbrace{0,\cdots,0}_{\beta},\underbrace{1,\cdots,1}_{2\lfloor nD\rfloor-\beta+1},0,\cdots,0).$$

Since $r^*_\delta \ge \lfloor nD\rfloor$ and $\beta \ge r^*_\delta + 1$, we have $w(\tilde{x}^n) \le \lfloor nD\rfloor$. Equivalently, for a permuted sequence $\rho^{-1}\circ\tilde{x}^n = (\tilde{x}_{\rho^{-1}(1)}, \tilde{x}_{\rho^{-1}(2)}, \cdots, \tilde{x}_{\rho^{-1}(n)})$, we have $w(\rho^{-1}\circ\tilde{x}^n) \le \lfloor nD\rfloor$, where ρ^{-1} is the inverse function of ρ. Hence, if $\beta \le 2\lfloor nD\rfloor + 1$, we have

$$\begin{aligned}d(x^n, x^n_{\max}) &\ge d(x^n, \rho^{-1}\circ\tilde{x}^n)\\ &= d(\rho\circ x^n, \rho\circ(\rho^{-1}\circ\tilde{x}^n))\\ &= d(\rho\circ x^n, \tilde{x}^n) = \frac{2\lfloor nD\rfloor+1}{n},\end{aligned} \tag{A· 9}$$

and if $\beta > 2\lfloor nD\rfloor + 1$, we have

$$d(x^n, x^n_{\max}) \ge d(x^n, 0^n) = \frac{\beta}{n} > \frac{2\lfloor nD\rfloor+1}{n}. \tag{A· 10}$$

Thus, for any $y^n \in \mathcal{Y}^n$ such that $d(x^n_{\max}, y^n) \le D$, we have

$$\begin{aligned}d(x^n, y^n) &\overset{\text{(a)}}{\ge} d(x^n, x^n_{\max}) - d(x^n_{\max}, y^n)\\ &\overset{\text{(b)}}{\ge} \frac{\lfloor nD\rfloor+1}{n} > D,\end{aligned}$$

where (a) comes from the triangle inequality of the distance, and (b) comes from (A· 9) and (A· 10). This completes the proof. □

Proof of Lemma 4. First, we show (25). Let $w(y^n) = \beta_1\,(\le r^*_\delta + \lfloor nD\rfloor)$, and ρ be the permutation such that first β_1 elements of $\rho\circ y^n$ are all 1. Then, for $\tilde{x}^n \in \mathcal{X}^n$ such that first $\beta_1 - \lfloor nD\rfloor$ elements are all 1 and the rest are all 0, we have $w(\rho^{-1}\circ\tilde{x}^n) \le r^*_\delta$ and $d(\rho^{-1}\circ\tilde{x}^n, y^n) = d(\tilde{x}^n, \rho\circ y^n) \le D$. Thus, by the definition (23) of $\tilde{P}^{[\mathrm{u}]}_{Y^n|X^n}$, we have (25).

Next, we show (26). Since $w(y^n) = r^*_\delta + 1 + \lfloor nD\rfloor$, for any $x^n \in \mathcal{X}^n$ such that $w(x^n) \le r^*_\delta$, we have

$$\begin{aligned}d(x^n, y^n) &\ge d(y^n, 0^n) - d(x^n, 0^n)\\ &\ge \frac{\lfloor nD\rfloor+1}{n} > D.\end{aligned}$$

Further, for any $x^n \in \mathcal{X}^n$, we have

$$\begin{aligned}d(x^n_{\max}, y^n) &\ge d(y^n, 0^n) - d(x^n_{\max}, 0^n)\\ &\overset{\text{(a)}}{\ge} \frac{r^*_\delta+1}{n}\\ &\overset{\text{(b)}}{\ge} \frac{\lfloor nD\rfloor+1}{n} > D,\end{aligned}$$

where (a) comes from the definition (24) of $x^n_{\max}$, and (b) comes from the assumption of this lemma. On the other

hand, let $w(y^n) = \beta_2 (= r^*_\delta + 1 + \lfloor nD \rfloor)$, and ρ be the permutation such that first β_2 elements of $\rho \circ y^n$ are all 1. Then, for $\tilde{x}^n \in \mathcal{X}^n$ such that first $\beta_2 - \lfloor nD \rfloor$ elements are all 1 and the rest are all 0, we have $w(\rho^{-1} \circ \tilde{x}^n) = r^*_\delta + 1$ and $d(\rho^{-1} \circ \tilde{x}^n, y^n) = d(\tilde{x}^n, \rho \circ y^n) \le D$. Thus, by the definition (23) of $\tilde{P}^{[\mathrm{u}]}_{Y^n|X^n}$, we have (26).

Finally, we show (27). For any $y^n \in \mathcal{Y}^n$ such that $w(y^n) \ge r^*_\delta + 2 + \lfloor nD \rfloor$, we have for any $x^n \in \mathcal{X}^n$ such that $w(x^n) \le r^*_\delta + 1$,

$$\begin{aligned} d(x^n, y^n) &\ge d(y^n, 0^n) - d(x^n, 0^n) \\ &\ge \frac{r^*_\delta + 2 + \lfloor nD \rfloor - r^*_\delta - 1}{n} \\ &= \frac{\lfloor nD \rfloor + 1}{n} > D, \end{aligned} \tag{A·11}$$

and for any $x^n \in \mathcal{X}^n$,

$$\begin{aligned} d(x^n_{\max}, y^n) &\ge d(y^n, 0^n) - d(x^n_{\max}, 0^n) \\ &\overset{(a)}{\ge} \frac{r^*_\delta + 2}{n} \\ &\overset{(b)}{>} D, \end{aligned} \tag{A·12}$$

where (a) comes from the definition (24) of $x^n_{\max}$, and (b) comes from the assumption of the lemma. According to (23), inequalities (A·11) and (A·12) imply (27). □

Appendix C:

In this appendix, we give proofs of Lemmas 5 and 6.

Proof of Lemma 5. For any given $x^n \in \mathcal{S}_r$, any $y^n \in \mathcal{A}^k_w(x^n)$ $(= \mathcal{B}_w(x^n) \cap \mathcal{S}_k)$ can be obtained from x^n if and only if we replace i ones in x^n with zeros and $i + k - r$ zeros in x^n with ones, where $0 \le i \le r$, $0 \le i+k-r \le n-r$, and $2i+k-r \le w$, i.e., $\max\{0, r-k\} \le i \le \min\left\{r, n-k, \left\lfloor \frac{w+r-k}{2} \right\rfloor\right\}$. This means that

$$\mathcal{A}^k_w(x^n) = \bigcup_{i=\max\{0,r-k\}}^{\min\left\{r,n-k,\left\lfloor \frac{w+r-k}{2} \right\rfloor\right\}} \tilde{\mathcal{A}}_i, \tag{A·13}$$

where

$$\tilde{\mathcal{A}}_i = \{y^n \in \{0,1\}^n : y^n \text{ is a sequence obtained by replacing } i \text{ ones in } x^n \text{ with zeros and } i + k - r \text{ zeros in } x^n \text{ with ones}\}.$$

According to (A·13), $\mathcal{A}^k_w(x^n)$ is empty if and only if $\max\{0, r-k\} > \left\lfloor \frac{w+r-k}{2} \right\rfloor$. Hence, by noting that $\tilde{\mathcal{A}}_i \cap \tilde{\mathcal{A}}_j = \emptyset$ for any $i \ne j$ and $|\tilde{\mathcal{A}}_i| = \binom{r}{i}\binom{n-r}{i+k-r}$, we have (30). □

Proof of Lemma 6. According to Lemma 5, $\mathcal{A}^k_w(x^n)$ is not empty if $\max\{0, r-k\} \le \left\lfloor \frac{w+r-k}{2} \right\rfloor$. Hence, in order to show the emptiness of $\mathcal{A}^{k(w(x^n),D)}_{\lfloor nD \rfloor}(x^n)$, we show that $0 \le \left\lfloor \frac{w+r-k}{2} \right\rfloor$ and $r - k \le \left\lfloor \frac{w+r-k}{2} \right\rfloor$ for $r = w(x^n)$, $w = \lfloor nD \rfloor$, and $k = k(r, D)$ such that $w < r \le \lfloor (1/D - 1)w \rfloor - 1$. For this setting, we note that $k(r, D) \le w/D \le n$.

First, we show that $0 \le \left\lfloor \frac{w+r-k}{2} \right\rfloor$. To this end, we only have to show that $w + r - k \ge 0$. We have

$$\begin{aligned} w + r - k &\overset{(a)}{\ge} w + r - \left\lfloor \frac{r-w}{1-2D} \right\rfloor - 1 \\ &\ge \frac{(2-2D)w - 2Dr}{1-2D} - 1 \\ &\overset{(b)}{>} \frac{(2-2D)w - 2D(1/D-1)w}{1-2D} - 1 \\ &= -1, \end{aligned}$$

where (a) follows since $k = k(r, D) \le \left\lfloor \frac{r-w}{1-2D} \right\rfloor + 1$, and (b) comes from the fact that $r < (1/D - 1)w$. Since $w + r - k$ is an integer, $w + r - k > -1$ implies that $w + r - k \ge 0$.

Next, we show that $r - k \le \left\lfloor \frac{w+r-k}{2} \right\rfloor$. We have

$$\begin{aligned} \left\lfloor \frac{w+r-k}{2} \right\rfloor - r + k &> \frac{w+r-k}{2} - 1 - r + k \\ &= \frac{w-r+k}{2} - 1. \end{aligned}$$

Since $\left\lfloor \frac{w+r-k}{2} \right\rfloor - r + k$ is an integer, we only have to show that $w - r + k \ge 0$. Now, we have

$$\begin{aligned} w - r + k &\overset{(a)}{\ge} w - r + \left\lfloor \frac{r-w}{1-2D} \right\rfloor \\ &> \frac{2D(r-w)}{1-2D} - 1 \\ &\overset{(b)}{>} -1, \end{aligned}$$

where (a) follows since $k = k(r, D) \ge \left\lfloor \frac{r-w}{1-2D} \right\rfloor$, and (b) comes from the fact that $r > w$. Since $w - r + k$ is an integer, $w - r + k > -1$ implies that $w - r + k \ge 0$.

Therefore, we have $\max\{0, r-k\} \le \left\lfloor \frac{w+r-k}{2} \right\rfloor$, and hence $\mathcal{A}^{k(w(x^n),D)}_{\lfloor nD \rfloor}(x^n)$ is not empty. □

Appendix D:

In this appendix, we prove Lemma 7.

Before we give the proof, we show the next lemma.

Lemma 9. For integers $r > 0$ and $r' > 0$, we have $k(r, D) \ne k(r', D)$ if and only if $r \ne r'$.

Proof. We only show that $r \ne r' \Rightarrow k(r, D) \ne k(r', D)$, because the opposite direction is trivial. Without loss of generality, we assume that $r \ge r' + 1$. Then, we have

$$\begin{aligned} f(r, D) - f(r', D) &> \left(\frac{r - \lfloor nD \rfloor}{1-2D} - 1\right) - \frac{r' - \lfloor nD \rfloor}{1-2D} \\ &\ge r - r' - 1. \end{aligned}$$

Since both sides are integers, this inequality implies

$$f(r, D) \ge f(r', D) + r - r'. \tag{A·14}$$

According to this inequality (A· 14), and the definition of $k(r, D)$, we have

$$k(r, D) \geq k(r', D).$$

Hence, in what follows, we show that $k(r, D) > k(r', D)$ in order to prove that $k(r, D) \neq k(r', D)$.

First, we assume that

$$f(r, D) = r - \lfloor nD \rfloor \pmod 2,$$
$$f(r', D) = r' - \lfloor nD \rfloor \pmod 2,$$

or

$$f(r, D) \neq r - \lfloor nD \rfloor \pmod 2,$$
$$f(r', D) \neq r' - \lfloor nD \rfloor \pmod 2.$$

In these cases, according to (A· 14), we have

$$k(r, D) - k(r', D) = f(r, D) - f(r', D) \geq 1. \quad \text{(A· 15)}$$

Secondly, we assume that

$$f(r, D) = r - \lfloor nD \rfloor \pmod 2,$$
$$f(r', D) \neq r' - \lfloor nD \rfloor \pmod 2.$$

Then, we will have

$$f(r, D) \geq f(r', D) + 2.$$

In order to show this inequality, we only consider the case where $r = r' + 1$. This is because if $r \geq r' + 2$, from (A· 14), we immediately have

$$f(r, D) \geq f(r', D) + r - r' \geq f(r', D) + 2.$$

Since $r = r' + 1$ and $f(r, D) = r - \lfloor nD \rfloor \pmod 2$, both of $f(r, D)$ and $r - \lfloor nD \rfloor$ are even or odd. If we assume that $r - \lfloor nD \rfloor$ is even (resp. odd), then $r' - \lfloor nD \rfloor$ is odd (resp. even), and hence $f(r', D)$ is even (resp. odd). Thus, both $f(r, D)$ and $f(r', D)$ are even (resp. odd). On the other hand, according to (A· 14), $f(r, D) > f(r', D)$. Thus, we have

$$f(r, D) \geq f(r', D) + 2.$$

Now, we have

$$k(r, D) - k(r', D) = f(r, D) - f(r', D) - 1 \geq 1. \quad \text{(A· 16)}$$

Finally, we assume that

$$f(r, D) \neq r - \lfloor nD \rfloor \pmod 2,$$
$$f(r', D) = r' - \lfloor nD \rfloor \pmod 2.$$

Then, we have

$$k(r, D) - k(r', D) = f(r, D) + 1 - f(r', D) \geq 2. \quad \text{(A· 17)}$$

Therefore, combining (A· 15)–(A· 17), we have $k(r, D) > k(r', D)$. □

By using this lemma, we prove Lemma 7.

Proof of Lemma 7. First, we consider the case where $y^n \in \{0^n\} \cup \mathcal{S}^{r_\delta}_{w_\delta+1}$. Then, by the definition (29), we have

$$\sup_{x^n \in \mathcal{X}^n} \tilde{P}^{[\mathrm{s}]}_{Y^n|X^n}(y^n|x^n) = 1. \quad \text{(A· 18)}$$

Secondly, we consider the case where $y^n \in \mathcal{S}_{k(r,D)} \cap \left[\{0^n\} \cup \mathcal{S}^{r_\delta}_{w_\delta+1}\right]^c$ for some $r \in \{\lfloor nD \rfloor + 1, \cdots, w_\delta\}$. Then, for any $x^n \in \mathcal{A}^r_{\lfloor nD \rfloor}(y^n)$, we have $y^n \in \mathcal{A}^{k(r,D)}_{\lfloor nD \rfloor}(x^n) = \mathcal{A}^{k(w(x^n),D)}_{\lfloor nD \rfloor}(x^n)$. Thus, by the definition (29), for any $x^n \in \mathcal{A}^r_{\lfloor nD \rfloor}(y^n)$, we have

$$\tilde{P}^{[\mathrm{s}]}_{Y^n|X^n}(y^n|x^n) = \frac{1}{A^{k(r,D)}_{\lfloor nD \rfloor}(r, n)}. \quad \text{(A· 19)}$$

On the other hand, for any $x^n \notin \mathcal{A}^r_{\lfloor nD \rfloor}(y^n)$, we will have

$$\tilde{P}^{[\mathrm{s}]}_{Y^n|X^n}(y^n|x^n) = 0. \quad \text{(A· 20)}$$

In order to show this equality (A· 20), we separately consider two cases: One is $x^n \notin \mathcal{B}_{\lfloor nD \rfloor}(y^n)$ and the other is $x^n \notin \mathcal{S}_r$. If $x^n \notin \mathcal{B}_{\lfloor nD \rfloor}(y^n)$, then $y^n \notin \mathcal{B}_{\lfloor nD \rfloor}(x^n)$, and hence obviously $y^n \notin \mathcal{A}^{k(w(x^n),D)}_{\lfloor nD \rfloor}(x^n)$. Thus, $\tilde{P}^{[\mathrm{s}]}_{Y^n|X^n}(y^n|x^n) = 0$ holds in this case. If $x^n \notin \mathcal{S}_r$, then $x^n \in \mathcal{S}_{r'}$ for some $r' \neq r$. According to Lemma 9, we have $k(r, D) \neq k(r', D)$. Then, by recalling the fact that $y^n \in \mathcal{S}_{k(r,D)}$, it holds that $y^n \notin \mathcal{S}_{k(r',D)} = \mathcal{S}_{k(w(x^n),D)}$, and hence we have $y^n \notin \mathcal{A}^{k(w(x^n),D)}_{\lfloor nD \rfloor}(x^n)$. Since $y^n \notin \{0^n\} \cup \mathcal{S}^{r_\delta}_{w_\delta+1}$, $\tilde{P}^{[\mathrm{s}]}_{Y^n|X^n}(y^n|x^n) = 0$ also holds in this case. Hence, from (A· 19) and (A· 20), for any $y^n \in \mathcal{S}_{k(r,D)} \cap \left[\{0^n\} \cup \mathcal{S}^{r_\delta}_{w_\delta+1}\right]^c$, we have

$$\sup_{x^n \in \mathcal{X}^n} \tilde{P}^{[\mathrm{s}]}_{Y^n|X^n}(y^n|x^n) = \frac{1}{A^{k(r,D)}_{\lfloor nD \rfloor}(r, n)}. \quad \text{(A· 21)}$$

Finally, we consider the case where $y^n \notin \{0^n\} \cup \mathcal{S}^{r_\delta}_{w_\delta+1} \cup \bigcup_{r \in \{\lfloor nD \rfloor + 1, \cdots, w_\delta\}} \mathcal{S}_{k(r,D)}$. Then, by the definition, we have

$$\sup_{x^n \in \mathcal{X}^n} \tilde{P}^{[\mathrm{s}]}_{Y^n|X^n}(y^n|x^n) = 0. \quad \text{(A· 22)}$$

By combining (A· 18), (A· 21) and (A· 22), we have the lemma. □

Tetsunao Matsuta was born in Fukui, Japan, on March 17, 1985. He received the B.E. degree from Utsunomiya University in 2007. He received the M.E. degree, and the Dr.Eng. degree from Tokyo Institute of Technology, Japan, in 2009, 2012, respectively. Since 2012, he has been an assistant professor at Tokyo Institute of Technology, and currently he is in the Department of Information and Communications Engineering of Tokyo Institute of Technology. He received the Best Paper Award in 2016 from IEICE. His current research interests are in the area of multi-terminal information theory and non-asymptotic analysis of coding problems.

Tomohiko Uyematsu received the B.E., M.E. and Dr.Eng. 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 Tokyo Institute of Technology as associate professor, and currently he is with the Department of Information and Communications Engineering as professor. In 1992 and 1996, he was a visiting researcher at the Supelec (Ecole superieure d'electricite), France and Delft University of Technology, Netherlands, respectively. He was an associate editor of IEEE Trans. Information Theory from 2010 to 2013. He received the Achievement Award in 2008, and the Best Paper Award in 1993, 1996, 2002, 2007, 2011, 2014 and 2016 both from IEICE. His current research interests are in the areas of information theory, especially Shannon theory and multi-terminal information theory. Dr. Uyematsu is a senior member of IEEE.