Identification via Compressed Data

Rudolf Ahlswede, * Enhui Yang, † and Zhen Zhang‡

July 7, 1998

*Rudolf Ahlswede is with Fakultät für Mathematik, Bielefeld Universität, Bielefeld, Germany, Postfach 100131

†En-hui Yang is with Dept. of Mathematics, Nankai University, Tianjin 300071, P.R. China. He is currently visiting Communication Sciences Institute, Dept. of EE-Systems, University of Southern California, Los Angeles, CA 90089-2565.

‡Zhen Zhang is with Communication Sciences Institute, Electrical Engineering–Systems, University of Southern California, Los Angeles, CA. 90089-2565. Research for this author is supported in part by NSF Grant NCR-9205265.

Abstract

We introduce and analyze a new coding problem for a correlated source $(X^n, Y^n)_{n=1}^{\infty}$. The observer of X^n can transmit data depending on X^n at a prescribed rate R. Based on these data the observer of Y^n tries to identify whether for some distortion measure ρ (like the Hamming distance) $\frac{1}{n}\rho(X^n, Y^n) \leq d$, a prescribed fidelity criterion. We investigate as functions of R and d the exponents of two error probabilities, the probabilities for misacceptance and the probabilities for misrejection. Our analysis has led to a new method for proving converses. Its basis is "The Inherently Typical Subset Lemma". It goes considerably beyond the "Entropy Characterisation" of [2], the "Image Size Characterisation" of [3], and its extensions in [5]. It is conceivable that it has a strong impact on Multi-user Information Theory.

Key words:
Correlated source, identification with fidelity, misacceptance and misrejection error probabilities.

I Introduction and Formulation of Problem

A Introduction

In this paper, we consider a new model: identification via compressed data. To put it in perspective, let us first review the traditional problems in source coding theory. Consider the following diagram, where $\{X_n\}_{n=1}^{\infty}$ is an i.i.d. source taking values in a finite alphabet $\mathcal{X}$. The encoder output is a binary sequence which appears at a rate of R bits per symbol. The decoder output is a sequence $\{\hat{X}_n\}_{n=1}^{\infty}$ which takes values in a finite reproduction alphabet $\mathcal{Y}$. In traditional source coding theory, the decoder is required to recover $\{X_n\}_{n=1}^{\infty}$ either completely or with some allowable distortion. That is, the output sequence $\{\hat{X}_n\}_{n=1}^{\infty}$ of the decoder must satisfy

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{E}\rho(X_i, \hat{X}_i) \leq d \;, \tag{1}$$

for sufficiently large n, where $\mathbf{E}$ devotes the expected value,

$$\rho : \mathcal{X} \times \mathcal{Y} \to [0, +\infty) \tag{2}$$

is a distortion measure, and d is the allowable distortion between the source sequence and the reproduction sequence. The problem is then to determine the infimum of the rate R such that the system shown in Fig.1 can operate in such a way that (1) is satisfied. It is known from rate distortion theory [1], that the infimum is given by the rate distortion function of the source $\{X_n\}_1^\infty$. Let us now consider the following system. The sequence $\{Y_n\}_1^\infty$ is a sequence of i.i.d random variables taking values from $\mathcal{Y}$. Knowing Y^n, the decoder is now required to be able to identify whether or not the source sequence X^n and the sequence Y^n have some prescribed relation F in such a way that two kinds of error probabilities satisfy some prescribed conditions. In parallel with rate distortion theory, we consider in this paper the following relation F defined by

$$n^{-1}\sum_{i=1}^{n}\rho(X_i,Y_i)\le d\,. \tag{3}$$

That is, the values X^n and Y^n are said to have relation F if (3) is satisfied. The problem we are interested in is to determine the infimum of the rate R such that the system shown in Fig.2 can operate so that the error probability of misrejection, that is the decoder votes for 0 even though F holds and the error probability of misacceptance, that is the decoder votes for 1 even though F does not hold, satisfy certain constraints. So the goal of the decoder is to identify whether X^n is close to Y^n (in the sense of relation F) or not. The encoder is cooperative.

B Formal Statement of Problem

First, we present some notation used throughout the paper. Script capitals $\mathcal{X}$, $\mathcal{Y}$, $\cdots$, denote finite sets. The cardinality of a set $\mathcal{A}$ is denoted by $|\mathcal{A}|$. The letters P, Q, always stand for probability distributions on finite sets. X, Y, $\cdots$, denote random variables. The distributions of random variables X and Y are denoted by P_X and P_Y, respectively. $\mathcal{P}(\mathcal{X})$ stands for the set of all probability distributions on $\mathcal{X}$. The functions log and exp are understood to be to the base 2. If $\mathcal{A}$ is a finite set, then $\mathcal{A}^n$ denotes the set of all n-tuples $a^n=(a_1,\cdots,a_n)$ from $\mathcal{A}$. If $a=(a_i)$ is a finite or infinite sequence of letters from $\mathcal{A}$, let $a_m^n=(a_m,\cdots,a_n)$ and , for simplicity, write a_1^n as a^n. A similar convention also applies to random variables.

Let $\{(X_n,Y_n)\}_{n=1}^\infty$ be a sequence of independent drawings of a pair (X,Y) of random variables with joint distribution P_{XY} taking values in $\mathcal{X}\times\mathcal{Y}$. Let $\rho:\mathcal{X}\times\mathcal{Y}\to[0,\infty)$ be a distortion measure. Let $\{\rho_n : n=1,2,\cdots\}$ be a single-letter fidelity criterion generated by ρ, where

$$\rho_n:\mathcal{X}^n\times\mathcal{Y}^n\to[0,+\infty)$$

is a mapping defined by

$$\rho_n(x^n, y^n) = \frac{1}{n}\sum_{i=1}^{n} \rho(x_i, y_i),$$

for any $x^n \in \mathcal{X}^n$ and any $y^n \in \mathcal{Y}^n$. Without loss of generality, we assume that the distortion measure ρ satisfies

$$\max_{x\in\mathcal{X}} \min_{y\in\mathcal{Y}} \rho(x, y) = 0 \ . \tag{4}$$

Let $d \geq 0$ satisfy

$$d < \mathbf{E}\rho(X, Y) \ . \tag{5}$$

An n-th order indentification source (IDS) code $\mathcal{C}_n$ is defined as a triple $\mathcal{C}_n = (f_n, B_n, g_n)$ where $B_n \subset \{0,1\}^*$ is a prefix-free set, f_n (called an encoder) is a mapping from $\mathcal{X}^n$ to B_n, and g_n (called a decoder) is a mapping from $\mathcal{Y}^n \times B_n$ to $\{0,1\}$. Note that, in this definition, the encoder f_n can be of variable-length. The correspondence between an identification source code as defined here and the system shown in Figure 2 should be clear. When an identification source code $\mathcal{C}_n = (f_n, B_n, g_n)$ is used in the system shown in Figure 2, the performance can be measured by three quantities: the resulting average rate per symbol $r_n(\mathcal{C}_n)$, the first kind of error probability $P_{e1}(\mathcal{C}_n)$, and the second kind of error probability $P_{e2}(\mathcal{C}_n)$, where

$$r_n(\mathcal{C}_n) = \frac{1}{n}\mathbf{E}(\textit{the length of } f_n(X^n)) \ , \tag{6}$$

$$P_{e1}(\mathcal{C}_n) = \Pr\{g_n(Y^n, f_n(X^n)) = 0 | \rho_n(X^n, Y^n) \leq d\} \ , \tag{7}$$

and

$$P_{e2}(\mathcal{C}_n) = \Pr\{g_n(Y^n, f_n(X^n)) = 1 | \rho_n(X^n, Y^n) > d\} \ . \tag{8}$$

Clearly, $P_{e1}(\mathcal{C}_n)$ and $P_{e2}(\mathcal{C}_n)$ can be interpreted as the probability of misrejection and the probability of misacceptance (or false identification), respectively.

Let $R \in [0, +\infty)$, $\alpha \in (0, +\infty]$, and $\beta \in (0, +\infty]$. A triple (R, α, β) is said to be achievable with respect to a given d, if for any $\epsilon > 0$ there exists a sequence $\{\mathcal{C}_n\}_{n=1}^{\infty}$ n-th order IDS codes $\mathcal{C}_n$, such that for sufficiently large n,

$$r_n(\mathcal{C}_n) \leq R + \epsilon \ , \tag{9}$$

$$P_{e1}(\mathcal{C}_n) \leq 2^{-n(\alpha-\epsilon)} \ , \tag{10}$$

and

$$P_{e2}(\mathcal{C}_n) \leq 2^{-n(\beta-\epsilon)} \ , \tag{11}$$

where as a convention, $\alpha = +\infty(\beta = +\infty$, resp.) means that the probability of misrejection (false identification, resp.) is zero. Let $\mathcal{R}(d)$ be the set of all achievable triples. Let $\bar{\mathcal{R}}(d)$ denote the closure of $\mathcal{R}(d)$ with respect to the usual topology under which $a_n \to +\infty$ means that a_n is equal to ∞ for all but finitely many integers n. In this paper, we are interested in determining the region $\bar{\mathcal{R}}(d)$. Specifically, we define for each pair $(\alpha, \beta) \in [0, +\infty]^2$,

$$R^*_{XY}(\alpha, \beta, d) = \inf\{R : (R, \alpha, \beta) \in \bar{\mathcal{R}}(d)\} \ . \tag{12}$$

Our main problem is the determination of this function.

Note that since $\bar{\mathcal{R}}(d)$ is closed, the infimum in (12) is actually a minimum. It is easy to see that $R^*_{XY}(\alpha, \beta, d) \geq R^*_{XY}(\alpha, 0, d)$ for any $\beta \geq 0$. On the other hand, since $\bar{\mathcal{R}}(d)$ is closed, it follows from (12) that

$$R^*_{XY}(\alpha, 0, d) = \lim_{\beta \to 0} R^*_{XY}(\alpha, \beta, d).$$

Therefore, $R^*_{XY}(\alpha, \beta, d)$ is continuous at $\beta = 0$. A similar result holds for $\alpha = 0$.

C Discussion

In the last subsection we formulated the problem we are interested in as investigating the trade-off between the rate R and the error exponents α and β. A natural question to ask at this point is why the problem should be set up in this way. To answer this question, we first note that since $d < \mathbf{E}\rho(X, Y)$, it follows immediately that $\Pr(\rho_n(X^n, Y^n) \leq d) \to 0$ as n goes to infinity. Therefore, if instead of the two kinds of error probabilities, we use the error probability

$$P_e(\mathcal{C}_n) = \Pr(\rho_n(X^n, Y^n) \leq d) P_{e1}(\mathcal{C}_n) + \Pr(\rho_n(X^n, Y^n) > d) P_{e2}(\mathcal{C}_n)$$

as a criterion, as studied by Ahlswede and Csizar in their 1-Bit Theorem [4], then the present problem becomes trivial and no information needs to be sent. This leads us to consider the two kinds of error probabilities. Second, let us see what happens if the two kinds of error probabilties are only required to vanish as n goes to infinity. The following theorem (which will be proved in the Appendix) tells us that in this case the minimum achievable rate is always equal to zero.

Theorem 1 *For any distribution P_{XY} on $\mathcal{X} \times \mathcal{Y}$,*

$$R^0_{XY} = 0 \ , \tag{13}$$

where R_{XY}^0 is the infimum of all positive real numbers R such that there exists for any $\epsilon > 0$ a sequence $\{\mathcal{C}_n\}$ of IDS codes, where $\mathcal{C}_n$ is an n-th order IDS code, such that for sufficiently large n, $r_n(\mathcal{C}_n) \leq R + \epsilon$, and

$$\lim_{n\to\infty} P_{e1}(\mathcal{C}_n) = 0 \quad \textit{and} \quad \lim_{n\to\infty} P_{e2}(\mathcal{C}_n) = 0 \ .$$

Therefore the only interesting problem left is to investigate the trade-off of the rate R and the two error exponents. Indeed, the results we obtained in this paper show that the problem proposed in the last subsection is really very interesting and even led us to develop a new powerful method for proving converses in information theory.

II Statement and Discussion of Main Results

As before, let (X, Y) be a pair of random variables with probability distribution P_{XY} taking values in $\mathcal{X} \times \mathcal{Y}$. Let $d < \mathbf{E}\rho(X, Y)$ and define

$$\beta(d) = \inf_{\mu\in\mathcal{P}(d)} D(\mu||P_{XY}) \ , \tag{14}$$

where

$$\mathcal{P}(d) = \{\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : \sum_{x\in\mathcal{X},y\in\mathcal{Y}} \mu(x, y)\rho(x, y) \leq d\}$$

and $D(.||.)$ stands for the relative entropy function. It is not hard to see that

$$\lim_{n\to\infty} -\frac{1}{n} \log \Pr\{\rho_n(X^n, Y^n) \leq d\} = \beta(d) \ . \tag{15}$$

We distinguish two cases: (i) X and Y are independent; (ii) X and Y are correlated. To build up ideas we begin with the easier case (i), in which we have conclusive results.

A Indepedent Case

In this subsection, we assume that X and Y are indepedent, that is $P_{XY} = P_X \times P_Y$. Without loss of generality we further assume that $P_X(x) > 0$ and $P_Y(y) > 0$ for every $x \in \mathcal{X}$ and every $y \in \mathcal{Y}$.

Let U be a random variable taking values in some finite set $\mathcal{U}$. Let P_{XU} be the joint distribution of X and U on $\mathcal{X} \times \mathcal{U}$. Define

$$\mathcal{E}(P_{XU}, d) = \inf[D(P_{\tilde{Y}}||P_Y) + I(U \wedge \tilde{Y})] \ , \tag{16}$$

where the infimum is taken over all random variables $\tilde{Y}$ taking values in $\mathcal{Y}$ and being jointly distributed with X, U such that $\mathbf{E}\rho(X, \tilde{Y}) \leq d$. By using the same argument as in the proof of Lemma 2.2. of [5, pp. 124], it is not hard to prove that $\mathcal{E}(P_{XU}, d)$ has the following property.

<u>Lemma</u> 1 *$\mathcal{E}(P_{XU}, d)$ is non-increasing and convex as a function of d and continuous as a function of the pair (P_{XU}, d) where P_{XU} ranges over the set $\mathcal{P}(\mathcal{X} \times \mathcal{U})$.*

For any $\beta > 0$, we next define,

$$R(P_X, P_Y, \beta, d) = \inf\{I(X \wedge U) : U \text{ is a finite valued r.v. with} \mathcal{E}(P_{XU}, d) \geq \beta\} \ . \tag{17}$$

Let

$$R(P_X, P_Y, 0, d) = \lim_{\beta \to 0^+} R(P_X, P_Y, \beta, d) \ . \tag{18}$$

Clearly

$$R(P_X, P_Y, 0, d) =$$

$$\inf\{I(X \wedge U) : U \text{ is a finite valued r.v. with} \mathcal{E}(P_{XU}, d) > 0\} \ . \tag{19}$$

Define

$$\bar{R}(P_X, P_Y, \beta, d) = \lim_{\beta' \to \beta^-} R(P_X, P_Y, \beta', d) \ . \tag{20}$$

This is well defined since $R(P_X, P_Y, \beta, d)$ as a function of β is non-decreasing.

The following theorem gives a formula for $R^*(+\infty, \beta, d)$.

<u>Theorem</u> 2 *Assume X and Y are independent. Then for any $0 < d < \mathbf{E}\rho(X, Y)$ and $0 \leq \beta \leq \beta(d)$,*

$$R^*_{XY}(+\infty, \beta, d) = \bar{R}(P_X, P_Y, \beta, d) \ . \tag{21}$$

<u>Remark 1:</u>

At this point, let us pause to give a few comments on the issue concerning the computation of $R(P_X, P_Y, \beta, d)$. In the following subsection, we shall compute $R(P_X, P_Y, \beta, d)$ in the binary symmetric case. It turns out that in this special case, $R(P_X, P_Y, \beta, d)$ can be expressed in terms of the rate-distortion function of the source X. In general, however, the computation of this function may be very difficult. It seems to the authors that there is no easy way to apply the support lemma([2],[5] Chapter 3) to upper bound the cardinality of the set $\mathcal{U}$. Instead, we shall take an alternative approach to the problem. We define for each integer $k \geq 1$, and any $\beta > 0$

$$R_k(P_X, P_Y, \beta, d) = \inf\{I(X \wedge U) : U \text{ is a r.v. taking} \leq k \text{ values with} \mathcal{E}(P_{XU}, d) \geq \beta\} . \quad (22)$$

For $\beta = 0$, $R_k(P_X, P_Y, 0, d)$ is defined similarly. Clearly, $R_k(P_X, P_Y, \beta, d)$ as a function of k is non-increasing and converges to $R(P_X, P_Y, \beta, d)$ as k goes to infinity. Later on, we shall estimate the rate at which $R_k(P_X, P_Y, \beta, d)$ converges to $R(P_X, P_Y, \beta, d)$ to provide a partial solution to the problem of the computation of $R(P_X, P_Y, \beta, d)$.

To give a general formula for the function $R^*_{XY}(\alpha, \beta, d)$, we next modify the definition of the quantities $\mathcal{E}(P_{XU}, d)$ and $R(P_X, P_Y, \beta, d)$ as follows. For any $\gamma \geq 0$ and any $\alpha \geq 0$, define

$$\mathcal{E}(P_{XU}, \alpha, \gamma, d) = \inf[D(P_{\tilde{Y}}||P_Y) + I(U \wedge \tilde{Y})] , \quad (23)$$

where the infimum is taken over all random variables $\tilde{Y}$ taking values in $\mathcal{Y}$ and being jointly distributed with X, U such that $\mathbf{E}\rho(X, \tilde{Y}) \leq d$ and

$$D(P_{\tilde{Y}}||P_Y) + I(XU \wedge \tilde{Y}) \leq \gamma + \alpha . \quad (24)$$

Here we make use of the convention that the infimum taken over an empty set is $+\infty$. Let

$$\beta(P_X, d) = \inf_{\mathbf{E}\rho(X,\tilde{Y}) \leq d} D(P_{\tilde{Y}}||P_Y) + I(X \wedge \tilde{Y}) , \quad (25)$$

where the infimum is taken over all random variables $\tilde{Y}$ taking values in $\mathcal{Y}$ such that $\mathbf{E}\rho(X, \tilde{Y}) \leq d$. Then it is easy to see that in case $\gamma + \alpha < \beta(P_X, d)$ the following holds

$$\mathcal{E}(P_{XU}, \alpha, \gamma, d) = +\infty \quad (26)$$

for any random variable U. In analogy to Lemma 1, it is not hard to see that $\mathcal{E}(P_{XU}, \alpha, \gamma, d)$ has the following property.

Lemma 2 *$\mathcal{E}(P_{XU}, \alpha, \gamma, d)$ is non-increasing and convex as a function of α (γ or d resp.) and continuous as a function of the quadruple $(P_{XU}, \alpha, \gamma, d)$, where the quadruple $(P_{XU}, \alpha, \gamma, d)$ ranges over all quadruples satisfying $\gamma + \alpha > \beta(P_X, d)$, $\alpha > 0$ and $d > 0$.*

Similarly to (17) and (18), we define for any $\beta > 0$

$$R(P_X, P_Y, \alpha, \gamma, \beta, d) = \inf\{I(X \wedge U) : U \text{ is a finite valued r.v. with} \mathcal{E}(P_{XU}, \alpha, \gamma, d) \geq \beta\} . \tag{27}$$

Let

$$R(P_X, P_Y, \alpha, \gamma, 0, d) = \lim_{\beta \to 0^+} R(P_X, P_Y, \alpha, \gamma, \beta, d) . \tag{28}$$

Define

$$\bar{R}(P_X, P_Y, \alpha, \gamma, \beta, d) = \lim_{\beta' \to \beta^-} R(P_X, P_Y, \alpha, \gamma, \beta', d) . \tag{29}$$

The following theorem gives a general formula for $R^*_{XY}(\alpha, \beta, d)$.

Theorem 3 *Assume that X and Y are independent, then for any $0 < d < \mathbf{E}\rho(X, Y)$, $0 < \alpha \neq \beta(P_X, d) - \beta(d)$, and $0 \leq \beta \leq \beta(d)$, the following holds*

$$R^*_{XY}(\alpha, \beta, d) = \bar{R}(P_X, P_Y, \alpha, \beta(d), \beta, d) . \tag{30}$$

Remark 2: Obviously, $\beta(d) \leq \beta(P_X, d)$. If $\beta(d) < \beta(P_X, d)$, then it follows from (26) and (27) that for any $\alpha < \beta(P_X, d) - \beta(d)$ and $\beta \geq 0$

$$R(P_X, P_Y, \alpha, \beta(d), \beta, d) = 0 . \tag{31}$$

On the other hand, it is easy to see that in this special case, $R^*_{XY}(\alpha, \beta, d) = 0$ for any $\beta \in [0, +\infty]$. (This will become clear when we prove the direct part of Theorem 3.) Furthermore, it follows from the definition of $R^*_{XY}(\alpha, \beta, d)$ that as a function of α it is left continuous. Thus it will suffice for us to prove Theorem 3 for $\alpha > \beta(P_X, d) - \beta(d)$.

Note that Theorem 2 is actually a special case of Theorem 3, because for $\alpha = +\infty$

$$R(P_X, P_Y, \alpha, \beta(d), \beta, d) = R(P_X, P_Y, \beta, d) . \tag{32}$$

The reasons why we state Theorems 2 and 3 seperately can be seen in the following sections. Similarly to (22), we can also define $R_k(P_X, P_Y, \alpha, \gamma, \beta, d)$ for each $k \geq 1$. We conclude this section with pointing out the following facts on $R_k(P_X, P_Y, \beta, d)$ and $R_k(P_X, P_Y, \alpha, \gamma, \beta, d)$.

Fact 1. $R_k(P_X, P_Y, \beta, d)$ as a function of the triple (P_X, β, d) is lower semi-continuous.

Fact 2. $R_k(P_X, P_Y, \beta, d)$ as a function of β is left continuous.

Fact 3. $R_k(P_X, P_Y, \alpha, \gamma, \beta, d)$ as a function of the quintuple $(P_X, \alpha, \gamma, \beta, d)$ is lower semi-continuous in the range $\gamma + \alpha > \beta(P_X, d)$, $\alpha > 0$ and $d > 0$.

Fact 4. $R_k(P_X, P_Y, \alpha, \gamma, \beta, d)$ as a function of β is left continuous if $\gamma + \alpha > \beta(P_X, d)$.

B Example: The Binary Symmetric Case

In this subsection, we consider the binary symmetric case where $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, X and Y are independent and uniformly distributed over $\{0, 1\}$, and the distortion measure ρ is the Hamming distance.

We first evaluate $R^*_{XY}(+\infty, \beta, d)$ from Theorem 2 and then show as an example how to prove Theorem 2 in this special case. Note that in this case, $\beta(d) = 1 - h(d)$ where $h(\cdot)$ is the binary entropy function.

Theorem 4 *For any* $0 \le d < \frac{1}{2}$ *and* $0 \le \beta \le \beta(d)$,

$$R^*_{XY}(+\infty, \beta, d) = 1 - h(d_\beta - d), \tag{33}$$

where $d_\beta \le \frac{1}{2}$ *satisfies* $h(d_\beta) = 1 - \beta$.

Proof. It is easy to see that $1 - h(d_\beta - d)$ is a continuous function of β. In view of Theorem 2, it suffices to prove that for any $0 < \beta < \beta(d)$

$$R(P_X, P_Y, \beta, d) = 1 - h(d_\beta - d)\ . \tag{34}$$

Let U be a random variable taking values uniformly in $\{0, 1\}$ and such that

$$I(X \wedge U) = 1 - h(d_\beta - d) \quad \text{and} \quad \mathbf{E}\rho(X, U) \le d_\beta - d\ . \tag{35}$$

Since X takes values uniformly in $\{0, 1\}$, such a random variable exists [6]. It is easy to verify that

$$\begin{aligned} \mathcal{E}(P_{XU}, d) &= \inf_{\mathbf{E}\rho(X,\tilde{Y}) \le d} [D(P_{\tilde{Y}} || P_Y) + I(U \wedge \tilde{Y})] \\ &= \inf_{\mathbf{E}\rho(X,\tilde{Y}) \le d} [1 - H(\tilde{Y}|U)] \\ &\ge \inf_{\mathbf{E}\rho(X,\tilde{Y}) \le d} I(U \wedge \tilde{Y}) \end{aligned}$$

$$
\begin{aligned}
&\geq \inf_{\mathbf{E}\rho(U,\tilde{Y})\leq d_\beta} I(U \wedge \tilde{Y}) \\
&= 1 - h(d_\beta) \\
&= \beta
\end{aligned} \tag{36}
$$

where the last inequality is due to (35). Thus it follows from (17) that

$$
R(P_X, P_Y, \beta, d) \leq I(X \wedge U) = 1 - h(d_\beta - d) \; . \tag{37}
$$

To prove the reverse inequality of (37), let U be any random variable taking values in some finite set $\mathcal{U}$ such that $\mathcal{E}(P_{XU}, d) \geq \beta$. Since X takes values uniformly in $\{0, 1\}$, it suffices to prove that

$$
H(X|U) \leq h(d_\beta - d) \; . \tag{38}
$$

To this end, we solve the following optimization problem

$$
\inf_{\mathbf{E}\rho(X,\tilde{Y})\leq d} [1 - H(\tilde{Y}|U)] \; . \tag{39}
$$

For each $u \in \mathcal{U}$, let x_u be an element of $\{0, 1\}$ such that $P_{X|U}(x_u|u) \leq \frac{1}{2}$, where $P_{X|U}(x_u|u)$ denotes the conditional probability of $X = x_u$ given $U = u$. Since

$$
\mathcal{E}(P_{XU}, d) \geq \beta > 0 \; , \tag{40}
$$

it is not hard to see that

$$
\sum_{u\in\mathcal{U}} P_U(u)(\frac{1}{2} - P_{X|U}(x_u|u)) > d \; . \tag{41}
$$

From (41), it follows that the optimization problem (39) is equivalent to the following optimization problem

$$
\inf_{\mathbf{E}\rho(X,\tilde{Y})= d} [1 - H(\tilde{Y}|U)] \; . \tag{42}
$$

Since the objective function $1 - H(\tilde{Y}|U)$ depends only on $P_{\tilde{Y}|U}(x_u|u)$, it is not hard to see that the optimization problem (42) can be reformulated as maximizing

$$
\sum_u P_U(u) h(P_{\tilde{Y}|U}(x_u|u)) \tag{43}
$$

subject to

$$
\sum_u P_U(u) |P_{X|U}(x_u|u)) - P_{\tilde{Y}|U}(x_u|u))| = d \tag{44}
$$

and

$$
0 \leq P_{\tilde{Y}|U}(x_u|u)) \leq 1 \; . \tag{45}
$$

Since $P_{X|U}(x_u|u)) \leq \frac{1}{2}$, conditions (44) and (45) can be replaced by

$$\sum_u P_U(u)(P_{\tilde{Y}|U}(x_u|u)) - P_{X|U}(x_u|u))) = d \tag{46}$$

and

$$P_{X|U}(x_u|u)) \leq P_{\tilde{Y}|U}(x_u|u)) \leq 1 \; . \tag{47}$$

The standard Lagrange Multiplier Method can be used to to show that the maximum of the optimization problem given by (43), (46) and (47) is achieved at the point $\{P_{\tilde{Y}|U}(x_u|u))\}$ for which there exists a $\lambda > 0$ such that

(i) $P_{\tilde{Y}|U}(x_u|u)) = \lambda$ for any $u \in \mathcal{U}$ satisfying $P_{X|U}(x_u|u)) \leq \lambda$;

(ii) $P_{\tilde{Y}|U}(x_u|u)) = P_{X|U}(x_u|u))$ for any $u \in \mathcal{U}$ satisfying $P_{X|U}(x_u|u)) > \lambda$;

(iii)
$$\sum_{u: P_{X|U}(x_u|u) \leq \lambda} P_U(u)[\lambda - P_{X|U}(x_u|u))] = d \; . \tag{48}$$

Clearly, this is something like water-filling. Since $\mathcal{E}(P_{XU}, d) \geq \beta$, it follows that

$$\sum_{u: P_{X|U}(x_u|u) \leq \lambda} P_U(u) h(\lambda) + \sum_{u: P_{X|U}(x_u|u) > \lambda} P_U(u) h(P_{X|U}(x_u|u)) \leq 1 - \beta \; . \tag{49}$$

We now claim that (38) holds. Otherwise, say,

$$H(X|U) = \sum_u P_U(u) h(P_{X|U}(x_u|u)) > h(d_\beta - d) \; . \tag{50}$$

Then, in view of (41), (48) and the fact that the derivative of the function $h(s)$ is strictly decreasing over the interval $s \in (0, \frac{1}{2}]$, we can see that

$$\begin{aligned} \sum_{u: P_{X|U}(x_u|u) \leq \lambda} P_U(u) h(\lambda) + \sum_{u: P_{X|U}(x_u|u) > \lambda} P_U(u) h(P_{X|U}(x_u|u)) &> h(d_\beta) \\ &= 1 - \beta \; . \end{aligned} \tag{51}$$

This contradicts (49). From (37) and (38), (34) follows immediately. This completes the proof of Theorem 4.

Note that $1 - h(d_\beta - d)$ is the rate distortion function of the source X evaluated at the point $d_\beta - d$. In some sense, therefore, Theorem 4 shows that there exists a close relationship between the rate $R^*_{XY}(+\infty, \beta, d)$ and the rate distortion function of X.

Next we outline the proof of Theorem 2 in the binary symmetric case. The direct part is easy. For any $d' < d_\beta - d$, roughly speaking, $2^{n(1-h(d'))}$ balls of radius nd' can almost cover the whole space. For each $x^n \in \{0,1\}^n$, we send simply the center of the ball in which x^n lies. Upon receiving this center, the decoder first calculates the Hamming distance between y^n and the center, and then outputs 1 if the distance is $\leq n(d'+d)$ and 0 otherwise. It is not hard to see that the misrejection probability is guaranteed to be zero, and the misacceptance probability is upper bounded by $2^{-n(1-h(d'+d))}$, which is less than or equal to $2^{-n\beta}$. This implies $R^*_{XY}(+\infty, \beta, d) \leq 1 - h(d_\beta - d)$.

To prove the converse part, let $(R, +\infty, \beta)$ be achievable, where $\beta > 0$. By definition, there exists for any $\epsilon > 0$ and sufficiently large n an nth-order IDS code $\mathcal{C}_n = (f_n, B_n, g_n)$ such that

$$r_n(\mathcal{C}_n) \leq R + \epsilon\ , \quad P_{e1}(\mathcal{C}_n) = 0\ , \quad \text{and} \quad P_{e2}(\mathcal{C}_n) \leq 2^{-n(\beta-\epsilon)}\ . \tag{52}$$

For any $b^n \in B_n$, let

$$S(b^n) = \{x^n \in \mathcal{X}^n : f_n(x^n) = b^n\}$$

and

$$S^d(b^n) = \{y^n \in \mathcal{Y}^n : \rho_n(x^n, y^n) \leq d \ \text{ for some } x^n \in S(b^n)\}\ .$$

From (52) and the Markov inequality, it follows that with very high probability $b^n \in B_n$ satisfies

$$\Pr\{Y^n \in S^d(b^n)\} \leq 2^{-n(\beta-2\epsilon)}\ .$$

To continue our derivation, we use at this point an isoperimetric theorem in combinatorial extremal theory [7] which says roughly that for any subset $A \subset \{0,1\}^n$ with $|A| = \sum_{i=0}^{k} \binom{n}{i}$ for some k, the cardinality of the Hamming l-neighbourhood $\Gamma^l A$ of A is minimized when A is a sphere, where for any $l \geq 0$

$$\Gamma^l A = \{y^n \in \{0,1\}^n : n\rho_n(x^n, y^n) \leq l \text{ for some } x^n \in A\}\ .$$

Using this result, one can prove that with very high probability $b^n \in B_n$ satisfies $|S(b^n)| \leq 2^{nh(d_{\beta,\epsilon}-d)}$, where $d_{\beta,\epsilon} \leq 1/2$ satisfies $h(d_{\beta,\epsilon}) = 1 - \beta + 2\epsilon$. This implies the converse part of Theorem 2.

The above argument is typical. It will be generalized to the general case in the subsequent sections. What makes the proof of the converse part easy is the solution of the isopermetric problem. Unfortunately, the solution of the isopermetric problem is very difficult in general. For the simplest distortion measure – Hamming distance – the solution is known only in the binary case and even for general alphabets, an asymptotically optimal solution can be derived using the image - size characteristization of [3] (see also [5]) In Section III, we develop a new method which yields in particular the asymptotic solution of the general isopermetric problem for arbitrary finite alphabets and arbitrary distortion measures. Although the exact solution of the problem is extremly difficult, the asymptotic solution is good enough for our identification problem at hand.

C Correlated Case

In this subsection, X and Y may be correlated. Unlike the independent case, only partial results on $R^*_{XY}(+\infty, 0, d)$ are obtained in this general case. First note that when X and Y are independent, $R(P_X, P_Y, 0, d)$ given by (19) can be rewritten as

$$R(P_X, P_Y, 0, d) = \inf_U I(X \wedge U) \tag{53}$$

where the infimum is taken over all random variables U taking values in some finite set such that

$$\mathbf{E}\bar{\rho}(P_{X|U}(\cdot|U), P_Y) > d \tag{54}$$

where $\bar{\rho}(P_{X|U}(\cdot|U), P_Y)$ denotes the $\bar{\rho}$- distance between the conditional distribution $P_{X|U}(\cdot|U)$ and the distribution P_Y of Y. (For the definition of $\bar{\rho}$ distance, we refer to [8]). The expression (53) of $R(P_X, P_Y, 0, d)$ will be extended to the general case.

Let $W(\cdot|\cdot) : \mathcal{X} \times \mathcal{Y} \to [0,1]$ be a stochastic matrix such that for any $x \in \mathcal{X}$ and any $y \in \mathcal{Y}$, $W(y|x)$ is the conditional probabiltiy of $Y = y$ given $X = x$. A stochastic matrix $\hat{W}(\cdot|\cdot) : \mathcal{X} \times \mathcal{Y} \to [0,1]$ is said to be absolutely continuous with respect to W if for any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, $W(y|x) = 0$ implies $\hat{W}(y|x) = 0$. Let $X_0(Y_0$, resp.) denote the projection of $\mathcal{X} \times \mathcal{Y}$ onto $\mathcal{X}$ ($\mathcal{Y}$, resp.). For any $P \in \mathcal{P}(\mathcal{X})$, define

$$\bar{\rho}_e(P) = \inf_Q \mathbf{E}_Q\, \rho(X_0, Y_0), \tag{55}$$

where the infimum is taken over all $Q \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ such that

(i) The marginal of Q on $\mathcal{X}$ is P;

(ii) The marginal of Q on $\mathcal{Y}$ is PW, where $PW \in \mathcal{P}(\mathcal{Y})$ is given by

$$PW(y) = \sum_{x \in \mathcal{X}} P(x)W(y|x), \quad y \in \mathcal{Y} \tag{56}$$

(iii) the transition probability matrix from X_0 to Y_0 under the distribution Q is absolutely continuous with respect to W.

Clearly, if $W(y|x) > 0$ for any $x \in \mathcal{X}$ and any $y \in \mathcal{Y}$, then $\bar{\rho}_e(P)$ is just the $\bar{\rho}$-distance between P and PW. For any $0 < d < \mathbf{E}\rho(X,Y)$, we next define

$$\tilde{R}(P_{XY}, 0, d) = \inf_U [I(X \wedge U) - I(Y \wedge U)], \tag{57}$$

where the infimum is taken over all random variables U taking values in some finite set $\mathcal{U}$ such that (i) $U \to X \to Y$ forms a Markov chain, and (ii) $\mathbf{E}\bar{\rho}_e(P_{X|U}(\cdot|U), P_Y) > d$. In the Appendix, we will prove that $\tilde{R}(P_{XY}, 0, d)$ has the following property.

Lemma 3 *$\tilde{R}(P_{XY}, 0, d)$ as a function of d is convex over the interval $0 < d < \mathbf{E}\rho(X,Y)$. Moreover, in evaluating $\tilde{R}(P_{XY}, 0, d)$ from (57), it suffices to consider only sets $\mathcal{U}$ with $|\mathcal{U}| \leq |\mathcal{X}| + 2$.*

Similarly to (57), we define for any $0 < d < \mathbf{E}(\rho(X,Y))$

$$R(P_{XY}, 0, d) = \inf_U I(X \wedge U) \tag{58}$$

where the infimum is taken over all random variables U taking values in some finite set $\mathcal{U}$ such that (i) $U \to X \to Y$ forms a Markov chain, (ii) $\mathbf{E}\bar{\rho}(P_{X|U}(\cdot|U), P_{Y|U}(\cdot|U)) > d$. Obviously, (58) is the extension of (53) to the general case. It is easy to see that a similar result to Lemma 3 holds also for $R(P_{XY}, 0, d)$. The following theorem gives an upper and a lower bound for $R^*_{XY}(+\infty, 0, d)$ in the general case that X and Y may be correlated.

Theorem 5 *For any $0 < d < \mathbf{E}\rho(X,Y)$,*

$$\tilde{R}(P_{XY}, 0, d) \leq R^*_{XY}(+\infty, 0, d) \leq R(P_{XY}, 0, d). \tag{59}$$

Note that when X and Y are independent, the lower and the upper bounds are the same in Theorem 5 and equal to $R(P_X, P_Y, 0, d)$. Considering the expression given by (57), a natural question to ask at this point is whether the lower bound is always tight. Unfortunately, the following example shows that this is not true in general.

Example 1.

Let X, Y, Z be three random variables taking values in finite sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ respectively, such that

(i) Y is independent of X, Z,

(ii) $P_{XZ}(x,z) > 0$ for any pair $(x,z) \in \mathcal{X} \times \mathcal{Z}$.

Assume the decoder in the system shown in Figure 2 now knows (Y^n, Z^n) and wants to identify whether $\rho_n(X^n, Y^n) \le d$. In another word, in addition to Y^n, the decoder knows side information Z^n which is correlated with X^n. Since $P_{XZ}(x,z) > 0$ for all pairs (x,z) and the distortion measure is irrelavent to Z, it is not hard to see that the side information is of no use and the minimum rate in bits per source symbol required to guarantee the zero probability of misrejection is still equal to $R^*_{XY}(+\infty, 0, d) = R(P_{XY}, 0, d)$. On the other hand, if we think of (Y, Z) as one random variable defined on $\mathcal{Y} \times \mathcal{Z}$ and extend ρ from $\mathcal{X} \times \mathcal{Y}$ to $\mathcal{X} \times (\mathcal{Y} \times \mathcal{Z})$ by letting $\rho(x, (y,z)) = \rho(x,y)$ for all triples (x,y,z), then we have

$$\begin{aligned} \tilde{R}(P_{X(YZ)}, 0, d) &= \inf_U [I(X \wedge U) - I(YZ \wedge U))] \\ &= \inf_U [I(X \wedge U) - I(Z \wedge U)], \end{aligned} \tag{60}$$

where the infimum is taken over all random variables U taking values in some finite set $\mathcal{U}$ such that (i) $U \to X \to (YZ)$ forms a Markov chain, or equivalently $U \to X \to Z$ forms a Markov chain, (ii) $\mathbf{E}\bar{\rho}(P_{X|U}(\cdot|U), P_{YZ|U}(\cdot|U)) > d$, or equivalently $\mathbf{E}\bar{\rho}(P_{X|U}(\cdot|U), P_Y) > d$. From (53), (58) and (60), it follows that if X and Z are highly correlated, then in general,

$$\begin{aligned} \tilde{R}(P_{X(YZ)}, 0, d) &< R(P_{XY}, 0, d) \\ &= R(P_{X(YZ)}, 0, d). \end{aligned} \tag{61}$$

This shows that, in this case, the upper bound $R(P_{X(YZ)}, 0, d)$ is tight , but the lower bound $\tilde{R}(P_{X(YZ)}, 0, d)$ is not.

Example 1 shows a case where side information is of no use to reduce the transmission rate R in the system shown in Figure 2. Let us now look at a case where side information does help in reducing the transmission rate R.

Let X, Y, Z be three random variables taking values on finite sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ respectively, such that $X \to Z \to Y$ form a Markov chain. Let $\{X^n, Z^n, Y^n\}$ be n independent drawings of the triple X, Z, Y. Assume that both the encoder and the decoder now know the side information Z^n. The decoder is still required to identify whether $\rho_n(X^n, Y^n) \le d$ with zero probability of misrejection. Clearly, this is a special case of the situation we considered in Theorem 5, if we think of (X, Z) and (Y, Z) as two random variables, and extend $\rho(x, y)$ to $\rho((x, z), (y, z'))$ accordingly. Interestingly enough, in this special case the side information does help in reducing the transmission rate.

Theorem 6 *If $X \to Z \to Y$ form a Markov chain, then for any $0 < d < \mathbf{E}\rho(X, Y)$,*

$$R^*_{(XZ),(YZ)}(+\infty, 0, d) = \tilde{R}(P_{(XZ)(YZ)}, 0, d). \tag{62}$$

In contrast to Example 1, Theorem 6 gives us another example for which the lower bound of (59) is tight, but the correponding upper bound is not.

We conclude this subsection with pointing out that if $X \to Z \to Y$ forms a Markov chain, then $\tilde{R}(P_{(XZ)(YZ)}, 0, d)$ can be rewritten as

$$\tilde{R}(P_{(XZ)(YZ)}, 0, d) = \inf_{U} I(X \wedge U|Z) \ , \tag{63}$$

where the infimum is taken over all random variables U taking values in some finite set $\mathcal{U}$ such that (i) $U \to (X, Z) \to Y$ forms a Markov chain; and (ii) $\mathbf{E}\bar{\rho}(P_{X|UZ}(\cdot|UZ), P_{Y|Z}(\cdot|Z)) > d$.

III Inherently Typical Subset Lemma

This section is devoted to develop a new method for proving converses, which can be used to prove the converse parts of Theorem 2 and 3 and to solve the general isoperimetric problem (a subject to which we intend to return in another paper). The main idea of this method is embodied in what we call inherently typical subset lemma.

For each integer $m > 0$, let $\mathcal{P}_m(\mathcal{X})$ denote the set of all m-types on $\mathcal{X}$, that is

$$\mathcal{P}_m(\mathcal{X}) = \{P \in \mathcal{P}(\mathcal{X}) : P(x) \in \{0, \frac{1}{m}, \frac{2}{m}, \cdots, \frac{m-1}{m}, 1\} \ \forall \ x \in \mathcal{X}\}. \tag{64}$$

Let $\mathcal{U}_m = \{u_1, \cdots, u_{|\mathcal{P}_m(\mathcal{X})|}\}$ be an arbitrary set. Since $|\mathcal{U}_m| = |\mathcal{P}_m(\mathcal{X})|$, we can associate with each $P \in \mathcal{P}_m(\mathcal{X})$ an element $u \in \mathcal{U}_m$ so that elememnts of $\mathcal{U}_m$ associated with distinct m-types are distinct. If $u \in \mathcal{U}_m$ is associated with $P \in \mathcal{P}_m(\mathcal{X})$, for convenience, we shall write P as

$P(\cdot|u)$. In terms of this notation, we have

$$\mathcal{P}_m(\mathcal{X}) = \{P(\cdot|u) : u \in \mathcal{U}_m\}. \tag{65}$$

Let A be any subset of $\mathcal{X}^n$. For any $0 \leq i \leq n-1$, define

$$A_i = \{x^i \in \mathcal{X}^i : x^i \text{is a prefix of some element of} A\}. \tag{66}$$

Here, we make use of the convention that $A_0 = \{\Lambda\}$, where Λ is the empty string. Assume that the integer m is greater than or equal to $2^{16|\mathcal{X}|^2}$.

<u>Definition</u> 1 *$A \subset \mathcal{X}^n$ is called m-inherently typical if there exists a mapping*

$$\phi : \cup_{i=0}^{n-1} A_i \to \mathcal{U}_m \tag{67}$$

such that the following hold:

(i) There exists an n-type $Q \in \mathcal{P}_n(\mathcal{X} \times \mathcal{U}_m)$ such that for any $x^n \in A$,

$$P_{x^n u^n}(x, u) = Q(x, u), \quad x \in \mathcal{X}, \quad u \in \mathcal{U}_m \tag{68}$$

where $u^n = (u_1, u_2, \cdots, u_n) \in \mathcal{U}_m^n$ is a sequence defined by $u_i = \phi(x^{i-1})$ for all $i : 1 \leq i \leq n$, (Such a sequence is called a sequence associated with x^n through ϕ) and for any $x \in \mathcal{X}$ and any $u \in \mathcal{U}_m$,

$$P_{x^n u^n}(x, u) = \frac{1}{n}|\{i : (x_i, u_i) = (x, u)\}|. \tag{69}$$

(ii) If $(\hat{X}, \hat{U})$ is a pair of random variables taking values in $\mathcal{X} \times \mathcal{U}_m$ with joint distribution Q, then

$$\frac{1}{n}\log|A| \leq H(\hat{X}|\hat{U}) \leq \frac{1}{n}\log|A| + \frac{\log^2 m}{m}. \tag{70}$$

Let $A \subset \mathcal{X}^n$ be m-inherently typical. Let ϕ be the corresponding mapping such that (68) and (70) are satisfied. For any random vector $\tilde{X}^n = (\tilde{X}_1, \cdots, \tilde{X}_n)$ taking values in A, we define another random vector $\tilde{U}^n = (\tilde{U}_1, \cdots, \tilde{U}_n)$ by letting $\tilde{U}_i = \phi(\tilde{X}^{i-1})$ for all $i : 1 \leq i \leq n$. Clearly (68) implies that with probability one, the following holds:

$$P_{\tilde{X}^n \tilde{U}^n}(x, u) = \frac{1}{n}\sum_{i=1}^{n} \Pr\{\tilde{X}_i = x, \tilde{U}_i = u\} \quad x \in \mathcal{X}, \quad u \in \mathcal{U}_m. \tag{71}$$

Note that the left hand side of (71) is the frequency, i.e. the average over time, and the right hand side is the average probability over ensemble. Intuitively, therefore, (71) just says that with probability one, the average over time is equal to the average over the ensemble. This is where the word "inherently typical" comes from. In typical applications, (see the following sections), the random vector $\tilde{X}^n$ is often assumed to be uniformly distributed on A. In this case

$$\begin{aligned}\frac{1}{n}\log|A| &= \frac{1}{n}H(\tilde{X}^n) \\ &= \frac{1}{n}\sum_{i=1}^{n} H(\tilde{X}_i|\tilde{X}^{i-1}).\end{aligned} \tag{72}$$

Let I be a random variable taking values uniformly in $\{1, \cdots, n\}$ and independent of $\tilde{X}^n$. Let $\tilde{X} = \tilde{X}_I$ and $U = (\tilde{X}^{I-1}, I)$, then

$$\frac{1}{n}\log|A| = H(\tilde{X}|U). \tag{73}$$

If we extend the mapping ϕ in the obvious way so that $\phi(U) = \phi(\tilde{X}^{I-1})$ whenever $U = (\tilde{X}^{I-1}, I)$, then it is not hard to see that $\tilde{X}$ and $\tilde{U}$ have the joint distribution $P_{\tilde{X}\tilde{U}} = Q$ where $\tilde{U} = \phi(U)$. Therefore, (70) just says that

$$H(\tilde{X}|U) \leq H(\tilde{X}|\tilde{U}) \leq H(\tilde{X}|U) + \frac{\log^2 m}{m}\ . \tag{74}$$

Lemma 4 (Inherently Typical Subset Lemma) *For any $m \geq 2^{16|\mathcal{X}|^2}$, n satisfying $((m+1)^{5|\mathcal{X}|+4}\ln(n+1))/n \leq 1$, and any $A \subset \mathcal{X}^n$, there exists an m-inherently typical subset $\tilde{A} \subset A$ such that*

$$\frac{1}{n}\log\frac{|A|}{|\tilde{A}|} \leq |\mathcal{X}|(m+1)^{|\mathcal{X}|}\frac{\log(n+1)}{n}\ . \tag{75}$$

Before proving Lemma 4, we remind the reader of the following two basic inequalities.

Lemma 5 (Pinsker Inequality [9], [5]) *For any two distributions $P_1,\ P_2 \in \mathcal{P}(\mathcal{X})$,*

$$D(P_1||P_2) \geq \frac{1}{2\ln 2}||P_1 - P_2||^2\ .$$

Lemma 6 (Folklore, Lemma 1.2.7 of [5]) *If P_1 and P_2 are two distributions on $\mathcal{X}$ such* *that*

$$||P_1 - P_2|| \leq \Theta \leq \frac{1}{2} ,$$

then

$$|H(P_1) - H(P_2)| \leq -\Theta \log \frac{\Theta}{|\mathcal{X}|} .$$

<u>Proof of Lemma 4.</u>

Let A be any subset of $\mathcal{X}^n$. Let $\tilde{X}^n = (\tilde{X}_1, \cdots, \tilde{X}_n)$ be a random vector taking values uniformly in A. Let p denote the distribution of $\tilde{X}^n$ on A. Define a mapping ϕ from $\cup_{i=0}^{n-1} A_i$ to $\mathcal{U}_m$ so that for any $x^i \in A_i$, $0 \leq i \leq n-1$,

$$||p(\cdot|x^i) - P(\cdot|\phi(x^i))|| \leq \frac{2|\mathcal{X}|}{m} , \tag{76}$$

where $p(\cdot|x^i)$ is the conditional distribution of $\tilde{X}_{i+1}$ given $\tilde{X}^i = x^i$, $P(\cdot|\phi(x^i))$ is the distribution in $\mathcal{P}_m(\mathcal{X})$ corresponding to $u = \phi(x^i)$(see (65)), and $||\cdot||$ denotes the variational distance between distributions. It is easy to see that such a mapping exists. (Essentially, this says that we use m-types to quantize distributions $p(\cdot|x^i)$, $x^i \in A_i$ and $0 \leq i \leq n-1$.) For each n-type $\bar{Q} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{U}_m)$, let $A_{\bar{Q}} \subset A$ consist of all sequences $x^n \in A$ such that

$$P_{x^n u^n}(x, u) = \bar{Q}(x, u), \quad x \in \mathcal{X} \quad \textit{and} \quad u \in \mathcal{U}_m ,$$

where $u^n \in \mathcal{U}_m^n$ is the sequence associated with x^n through the mapping ϕ and

$$P_{x^n u^n}(x, u) = \frac{1}{n} |\{i : (x_i, u_i) = (x, u)\}|.$$

Clearly, $\{A_{\bar{Q}} : \bar{Q} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{U}_m)\}$ is a partition of A. Let $Q \in \mathcal{P}_n(\mathcal{X} \times \mathcal{U}_m)$ satisfy

$$|A_Q| = \max\{|A_{\bar{Q}}| : \bar{Q} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{U}_m)\} . \tag{77}$$

We claim that $\tilde{A} = A_Q$ is the desired subset in Lemma 4. That is, that $\tilde{A}$ satisfies (75) and is an m-inherently typical subset under the mapping ϕ. To see this, first note that

$$|A| = \sum_{\bar{Q} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{U}_m)} |A_{\bar{Q}}| \leq |\mathcal{P}_n(\mathcal{X} \times \mathcal{U}_m)| |A_Q| \leq (n+1)^{|\mathcal{X}||\mathcal{U}_m|} |\tilde{A}| . \tag{78}$$

This, together with $|\mathcal{U}_m| \leq (m+1)^{|\mathcal{X}|}$, implies immediately (75). On the other hand, by the definition of A_Q, it follows that for any $x^n \in \tilde{A}$

$$P_{x^n u^n}(x, u) = Q(x, u), \quad x \in \mathcal{X}, \quad u \in \mathcal{U}_m, \tag{79}$$

where u^n is the sequence associated with x^n through ϕ. Therefore, all remaining to be proved is that if $(\hat{X}, \hat{U})$ is a random vector taking values on $\mathcal{X} \times \mathcal{U}_m$ with joint distribution $P_{\hat{X}\hat{U}} = Q$, then

$$\frac{1}{n} \log |\tilde{A}| \leq H(\hat{X}|\hat{U}) \leq \frac{1}{n} \log |\tilde{A}| + \frac{\log^2 m}{m} . \tag{80}$$

To prove (80), let $\bar{X}^n = (\bar{X}_1, \cdots, \bar{X}_n)$ be a random vector taking values uniformly in $\tilde{A}$. Let $\tilde{p}$ denotes the distribution of $\bar{X}^n$ on $\tilde{A}$. As in the analysis following definition 1, let I be a random variable taking values uniformly on $\{1, \cdots, n\}$. Let $\bar{X} = \bar{X}_I$, $U = (\bar{X}^{I-1}, I)$ and $\bar{U} = \phi(U)$ where $\phi(U) = \phi(\bar{X}^{I-1})$ whenever $U = (\bar{X}^{I-1}, I)$. Then

$$\frac{1}{n} \log |\tilde{A}| = H(\bar{X}|U) \leq H(\bar{X}|\bar{U}). \tag{81}$$

In view of (79), it is easy to see that $\bar{X}$ and $\bar{U}$ have the joint distribution $P_{\bar{X}\bar{U}} = Q$. Consequently, in the following it suffices to prove

$$H(\bar{X}|\bar{U}) \leq H(\bar{X}|U) + \frac{\log^2 m}{m} . \tag{82}$$

To this end, note that for any $x \in \mathcal{X}$ and $u \in \mathcal{U}_m$,

$$\begin{aligned} Q(x, u) &= P_{\bar{X}\bar{U}}(x, u) \\ &= \frac{1}{n} \sum_{i=1}^{n} \Pr(\bar{X}_i = x, \phi(\bar{X}^{i-1}) = u) \\ &= \sum_{x^n \in \tilde{A}} \tilde{p}(x^n) \frac{1}{n} \sum_{i=1}^{n} \delta(x_i, x) \delta(\phi(x^{i-1}), u) \\ &= \sum_{x^n \in \tilde{A}} \tilde{p}(x^n) \frac{1}{n} \sum_{i=1}^{n} \tilde{p}(x|x^{i-1}) \delta(\phi(x^{i-1}), u) , \end{aligned} \tag{83}$$

where $\delta(\cdot, \cdot)$ is the Kronecker Delta function, that is

$$\delta(z, z') = \begin{cases} 1 & \text{if } z = z' \\ 0 & \text{otherwise} \end{cases}$$

and $\tilde{p}(x|x^{i-1})$ is the conditional probability of $\bar{X}_i = x$ given $\bar{X}^{i-1} = x^{i-1}$. Since for any $x^n \in \tilde{A}$,

$$P_{\bar{U}}(u) = \frac{1}{n} \sum_{i=1}^{n} \delta(\phi(x^{i-1}), u), \quad u \in \mathcal{U}_m$$

it follows from (83) that

$$P_{\bar{X}\bar{U}}(x,u) - P_{\bar{U}}(u)P(x|u) = \sum_{x^n \in \tilde{A}} \tilde{p}(x^n)\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{p}(x|x^{i-1}) - P(x|u)\right)\delta(\phi(x^{i-1},u). \tag{84}$$

This implies

$$\sum_{x\in\mathcal{X}} |P_{\bar{X}\bar{U}}(x,u) - P_{\bar{U}}(u)P(x|u)| \le \sum_{x^n \in \tilde{A}} \tilde{p}(x^n)\frac{1}{n}\sum_{i=1}^{n}||\tilde{p}(\cdot|x^{i-1}) - P(\cdot|u)||\delta(\phi(x^{i-1}),u). \tag{85}$$

On the other hand, from (75) it follows that

$$\begin{aligned}
\frac{1}{n}\sum_{x^n\in\tilde{A}}\tilde{p}(x^n)\log\frac{\tilde{p}(x^n)}{p(x^n)} &= \sum_{x^n\in\tilde{A}}\tilde{p}(x^n)\frac{1}{n}\sum_{i=1}^{n}\log\frac{\tilde{p}(x_i|x^{i-1})}{p(x_i|x^{i-1})} \\
&= \sum_{x^n\in\tilde{A}}\tilde{p}(x^n)\frac{1}{n}\sum_{i=1}^{n}D(\tilde{p}(\cdot|x^{i-1})||p(\cdot|x^{i-1})) \\
&\le |\mathcal{X}|(m+1)^{|\mathcal{X}|}\frac{\log(n+1)}{n}\ .
\end{aligned} \tag{86}$$

Using Pinsker's inequality, we get

$$\sum_{x^n\in\tilde{A}}\tilde{p}(x^n)\frac{1}{n}\sum_{i=1}^{n}||\tilde{p}(\cdot|x^{i-1}) - p(\cdot|x^{i-1})||^2 \le 2|\mathcal{X}|(m+1)^{|\mathcal{X}|}\frac{\ln(n+1)}{n}\ . \tag{87}$$

Applying Schwartz inequality to (87) yields

$$\sum_{x^n\in\tilde{A}}\tilde{p}(x^n)\frac{1}{n}\sum_{i=1}^{n}||\tilde{p}(\cdot|x^{i-1}) - p(\cdot|x^{i-1})|| \le \left[2|\mathcal{X}|(m+1)^{|\mathcal{X}|}\frac{\ln(n+1)}{n}\right]^{\frac{1}{2}}\ . \tag{88}$$

Going back to (85), we get

$$\begin{aligned}
&\sum_{x\in\mathcal{X}}|P_{\bar{X}\bar{U}}(x,u) - P_{\bar{U}}(u)P(x|u)| \\
&\le \sum_{x^n\in\tilde{A}}\tilde{p}(x^n)\frac{1}{n}\sum_{i=1}^{n}||\tilde{p}(\cdot|x^{i-1}) - p(\cdot|x^{i-1})||\delta(\phi(x^{i-1}),u) \\
&\quad + \sum_{x^n\in\tilde{A}}\tilde{p}(x^n)\frac{1}{n}\sum_{i=1}^{n}||p(\cdot|x^{i-1}) - P(\cdot|u)||\delta(\phi(x^{i-1}),u) \\
&\le \left[2|\mathcal{X}|(m+1)^{|\mathcal{X}|}\frac{\ln(n+1)}{n}\right]^{\frac{1}{2}} + \frac{2|\mathcal{X}|}{m}P_{\bar{U}}(u)\ ,
\end{aligned} \tag{89}$$

where the last inequality is due to (88) and (76). Therefore, if

$$P_{\bar{U}}(u) \geq \left[2|\mathcal{X}|(m+1)^{|\mathcal{X}|}\frac{\ln(n+1)}{n}\right]^{\frac{1}{4}},$$

then

$$\begin{aligned}
||P_{\bar{X}|\bar{U}}(\cdot|u) - P(\cdot|u)|| &\leq \left[2|\mathcal{X}|(m+1)^{|\mathcal{X}|}\frac{\ln(n+1)}{n}\right]^{\frac{1}{4}} + \frac{2|\mathcal{X}|}{m} \\
&\leq \frac{4|\mathcal{X}|}{m}, \qquad (90)
\end{aligned}$$

where $P_{\bar{X}\bar{U}}(\cdot|u)$ is the conditional probability distribution of $\bar{X}$ given $\bar{U} = u$ and the last inequality is due to the assumption that $[(m+1)^{5|\mathcal{X}|+4}\ln(n+1)]/n \leq 1$. From (90) and Lemma 6, we have

$$\begin{aligned}
H(\bar{X}|\bar{U}) &= \sum_{x^n \in \tilde{A}} \tilde{p}(x^n)\frac{1}{n}\sum_{i=1}^{n} H(P_{\bar{X}|\bar{U}}(\cdot|\phi(x^{i-1}))) \\
&\leq \sum_{x^n \in \tilde{A}} \tilde{p}(x^n)\frac{1}{n}\sum_{i=1}^{n} H(P(\cdot|\phi(x^{i-1}))) + 4|\mathcal{X}|\frac{\log m}{m} \\
&\quad + |\mathcal{U}_m|\left[2|\mathcal{X}|(m+1)^{|\mathcal{X}|}\frac{\log(n+1)}{n}\right]^{\frac{1}{4}}\log|\mathcal{X}| \\
&\leq^{1)} \sum_{x^n \in \tilde{A}} \tilde{p}(x^n)\frac{1}{n}\sum_{i=1}^{n} H(P(\cdot|\phi(x^{i-1}))) + (4|\mathcal{X}|+1)\frac{\log m}{m} \\
&\leq^{2)} \sum_{x^n \in \tilde{A}} \tilde{p}(x^n)\frac{1}{n}\sum_{i=1}^{n} H(p(\cdot|x^{i-1})) + (6|\mathcal{X}|+1)\frac{\log m}{m}, \qquad (91)
\end{aligned}$$

where the inequality 1) is due to the assumption that $[(m+1)^{5|\mathcal{X}|+4}\ln(n+1)]/n \leq 1$ and $m \geq 2^{16|\mathcal{X}|^2}$, and inequality 2) is due to (76) and Lemma 6. To continue (91), we next compare $H(p(\cdot|x^{i-1}))$ with $H(\tilde{p}(\cdot|x^{i-1}))$. Let

$$F = \left\{(i, x^{i-1}) : 1 \leq i \leq n, x^n \in \tilde{A}, ||\tilde{p}(\cdot|x^{i-1}) - p(\cdot|x^{i-1})|| \geq \left[2|\mathcal{X}|(m+1)^{|\mathcal{X}|}\frac{\ln(n+1)}{n}\right]^{\frac{1}{4}}\right\}.$$

From (88) and the Markov inequality,

$$\sum_{x^n \in \tilde{A}} \tilde{p}(x^n)\frac{1}{n}\sum_{i=1}^{n} 1_F(i, x^n) \leq [2|\mathcal{X}|(m+1)^{|\mathcal{X}|}\frac{\log(n+1)}{n}]^{\frac{1}{4}}, \qquad (92)$$

where 1_F denotes the indicator function of F. From (92) and Lemma 6, it is not hard to verify that

$$\begin{aligned}\sum_{x^n\in\tilde{A}}\tilde{p}(x^n)\frac{1}{n}\sum_{i=1}^{n}H(p(\cdot|x^{i-1})) &\leq \sum_{x^n\in\tilde{A}}\tilde{p}(x^n)\frac{1}{n}\sum_{i=1}^{n}H(\tilde{p}(\cdot|x^{i-1})) \\ &\quad+\frac{2|\mathcal{X}|^2}{m}+2|\mathcal{X}|\frac{\log m}{m} \\ &= H(\bar{X}|U)+\frac{2|\mathcal{X}|^2}{m}+2|\mathcal{X}|\frac{\log m}{m} \\ &\leq H(\bar{X}|U)+(2|\mathcal{X}|+1)\frac{\log m}{m}\ . \end{aligned} \tag{93}$$

Combining (93) with (91) yields

$$\begin{aligned}H(\bar{X}|\bar{U}) &\leq H(\bar{X}|U)+(8|\mathcal{X}|+2)\frac{\log m}{m} \\ &\leq H(\bar{X}|U)+\frac{\log^2 m}{m}\ , \end{aligned} \tag{94}$$

where the last inequality is due to the assumption that $m > 2^{16|\mathcal{X}|^2}$. This completes the proof of (82) and hence the proof of Lemma 4.

Note that Lemma 4 is proved by estimating the variational distance between the distribution $P_{\bar{X}|\bar{U}}(\cdot|\phi(x^{i-1}))$ and $\tilde{p}(\cdot|x^{i-1})$ where $x^{i-1}\in\tilde{A}_{i-1}$. Roughly speaking, the attempt we have made in the proof of Lemma 4 is to show that the variational distance between these two distributions is roughly upper bounded by $\frac{\log^2 m}{m}$. In fact, this is just what the second condition of the definition of m-inherently typical subset implies. To see this, let us go back to (74), where $A\subset\mathcal{X}^n$ is assumed to be m-inherently typical. It is not hard to see that (74) can be rewritten as

$$\begin{aligned}H(\tilde{X}|\tilde{U})-H(\tilde{X}|U) &= \sum_{x^n\in A}p(x^n)\frac{1}{n}\sum_{i=1}^{n}D(p(\cdot|x^{i-1})||P_{\tilde{X}|\tilde{U}}(\cdot|\phi(x^{i-1}))) \\ &\leq \frac{\log^2}{m}\ . \end{aligned} \tag{95}$$

Using Pinsker's inequality once again, we get

$$\sum_{x^n\in A}p(x^n)\frac{1}{n}\sum_{i=1}^{n}||p(\cdot|x^{i-1})-P_{\tilde{X}|\tilde{U}}(\cdot|\phi(x^{i-1})))||^2\leq 2\frac{\log^2 m}{m} \tag{96}$$

which, together with Schwartz inequality, implies

$$\sum_{x^n \in A} p(x^n) \frac{1}{n} \sum_{i=1}^{n} ||p(\cdot|x^{i-1}) - P_{\tilde{X}|\tilde{U}}(\cdot|\phi(x^{i-1}))|| \leq \sqrt{2\frac{\log^2 m}{m}} . \tag{97}$$

This means that the average variational distance between $P_{\tilde{X}|\tilde{U}}(\cdot|\phi(x^{i-1}))$ and $p(\cdot|x^{i-1})$ is upper bounded by $\sqrt{2\frac{\log^2 m}{m}}$. Therefore, the second condition in the definition of m-inherently typical subsets is also stringent.

IV Proofs of Theorems 2 and 3

In this section, we assume X and Y are independent. First of all, let us review some basic facts about types and typical sequences. Let U be a random variable taking values on some finite set $\mathcal{U}$. Let $\{\gamma_n\}$ be a sequence of positive numbers such that $\gamma_n \to 0$ and $\sqrt{n}\gamma_n \to \infty$ as $n \to +\infty$. Recall that $\mathcal{P}_n(\mathcal{U})$ denotes the set of all n-types on $\mathcal{U}$. For each $u^n \in \mathcal{U}^n$, the type P_{u^n} of u^n is defined by

$$P_{u^n}(u) = \frac{1}{n}|\{i : u_i = u\}|, \quad u \in \mathcal{U}.$$

For each $P \in \mathcal{P}_n(\mathcal{U})$, let

$$T_P^n(\mathcal{U}) = \{u^n \in \mathcal{U}^n : P_{u^n} = P\}$$

and denote by $\mathcal{V}_n(P, \mathcal{U} \times \mathcal{X})$ the set of all stochastic matrices $V = (V(x|u))_{x\in\mathcal{X}, u\in\mathcal{U}}$ such that

$$V(x|u) \in \{0, \frac{1}{nP(u)}, \frac{2}{nP(u)}, \cdots, 1\} \text{ for all } x \in \mathcal{X} \text{ and } u \in \mathcal{U}.$$

Given $u^n \in \mathcal{U}^n$ and $V \in \mathcal{V}_n(P_{u^n}, \mathcal{U} \times \mathcal{X})$, a sequence x^n is said to be V-generated by u^n if for all $x \in \mathcal{X}$ and all $u \in \mathcal{U}$,

$$P_{u^n x^n}(u, x) = P_{u^n}(u)V(x|u).$$

Denote by $T_V^n(u^n, \mathcal{X})$ the set of all sequences x^n V-generated by u^n.

An n-type $P \in \mathcal{P}_n(\mathcal{U})$ is said to be (U, γ_n)-essential if

$$|P(u) - P_U(u)| \leq \gamma_n$$

and $P(u) = 0$ whenever $P_U(u) = 0$. A sequence $u^n \in \mathcal{U}^n$ is called (U, γ_n)-typical if P_{u^n} is (U, γ_n)-essential. Denote by T_{U,γ_n}^n the set of all (U, γ_n)-typical sequences. Similarly, for

$u^n \in \mathcal{U}^n$, we call $V \in \mathcal{V}_n(P_{u^n}, \mathcal{U} \times \mathcal{X})$ $(u^n, X|U, \gamma_n)$-essential if

$$|P_{u^n}(u)V(x|u) - P_{u^n}(u)P_{X|U}(x|u)| \leq \gamma_n$$

and $V(x|u) = 0$ whenever $P_{X|U}(x|u) = 0$ where $P_{X|U}(x|u)$ is the conditional probability of $X = x$ given $U = u$. A sequence $x^n \in \mathcal{X}^n$ is called $(u^n, X|U, \gamma_n)$-typical if there esists a $(u^n, X|U, \gamma_n)$-essential stochastic matrix $V \in \mathcal{V}_n(P_{u^n}, \mathcal{U} \times \mathcal{X})$ such that x^n is V-generated by u^n. Denote by $T^n_{X|U,\gamma_n}(u^n)$ the set of all $(u^n, X|U, \gamma_n)$-typical sequences x^n.

Although the above notation is introduced for random variables X and U and for finite sets $\mathcal{X}$ and $\mathcal{U}$, in the following we shall use freely these notation and terminology for other random variables and finite sets. Note that if u^n is (U, γ_n)-typical, and x^n is $(u^n, X|U, \gamma_n)$-typical, then $u^n x^n$ is $(UX, 2\gamma_n)$-typical and x^n is $(X, 2|\mathcal{U}|\gamma_n)$-typical. The following facts will be used.

$$|\mathcal{V}_n(P, \mathcal{U} \times \mathcal{X})| \leq (n+1)^{|\mathcal{U}||\mathcal{X}|}, \text{for any } P \in \mathcal{P}_n(\mathcal{U}). \tag{98}$$

$$(n+1)^{-|\mathcal{U}|} 2^{nH(P)} \leq |T^n_P(\mathcal{U})| \leq 2^{nH(P)}, \quad P \in \mathcal{P}_n(\mathcal{U}). \tag{99}$$

$$(n+1)^{-|\mathcal{U}||\mathcal{X}|} 2^{nH(V|P)} \leq |T^n_V(u^n, \mathcal{X})| \leq 2^{nH(V|P)}, \quad u^n \in T^n_P, \ P \in \mathcal{P}_n(\mathcal{U}) \textit{ and } V \in \mathcal{V}_n(P, \mathcal{U} \times \mathcal{X}), \tag{100}$$

where

$$\begin{aligned} H(V|P) &= \sum_{u \in \mathcal{U}} P(u) H(V(\cdot|u)) \\ &= \sum_{u \in \mathcal{U}} \sum_{x \in \mathcal{X}} -P(u)V(x|u) \log V(x|u) \ . \end{aligned} \tag{101}$$

Furthermore,

$$\Pr(U^n \in T^n_{U,\gamma_n}) \geq 1 - \frac{|\mathcal{U}|}{4n\gamma_n^2} \tag{102}$$

and if $\Pr(U^n = u^n) > 0$, then

$$\Pr(X^n \in T^n_{X|U,\gamma_n}(u^n) | U^n = u^n) \geq 1 - \frac{|\mathcal{X}||\mathcal{U}|}{4n\gamma_n^2} \ , \tag{103}$$

where (X^n, U^n) are n independent drawings of (X, U).

A Proof of Theorem 2

In view of the fact that $R^*_{XY}(\alpha, \beta, d)$ is continuous at $\beta = 0$, it suffices to prove Theorems 2 and 3 for $0 < \beta < \beta(d)$.

Proof of Theorem 2. We first prove the direct part, that is,

$$R^*_{XY}(+\infty, \beta, d) \le \bar{R}(P_X, P_Y, \beta, d).$$

By the definition of $R^*_{XY}(+\infty, \beta, d)$, it suffices to prove that for any $R > \bar{R}(P_X, P_Y, \beta, d)$, $(R, +\infty, \beta)$ is achievable. To this end, let us fix $R > \bar{R}(P_X, P_Y, \beta, d)$ below and prove $(R, +\infty, \beta)$ is achievable.

In view of the definition of $\bar{R}(P_X, P_Y, \beta, d)$ and $R(P_X, P_Y, \beta, d)$, it is not hard to see that for any $\delta > 0$, there exists a random variable U taking values on some finite set $\mathcal{U}$ such that

$$I(X \wedge U) < R, \quad \textit{and} \quad \mathcal{E}(P_{XU}, d) \ge \beta - \delta \, . \tag{104}$$

Based on the pair (X, U), the standard technique of [2] (see also [5]) can be used to show that there exists for sufficiently large n a system $\{(u^n(i), \mathcal{S}_i) : 1 \le i \le M\}$ which has the following properties:

Property (i) $\log M \le n(I(X \wedge U) + \delta)$,

Property (ii) For $1 \le i \le M$, $u^n(i) \in T^n_{U,\gamma_n}$, $\mathcal{S}_i \subset T^n_{X|U,\gamma_n}(u^n(i))$ and

$$\Pr(X^n \in \mathcal{S}_i | U^n = u^n(i)) \ge \frac{\delta}{2},$$

where (X^n, U^n) are n-independent drawings of (X, U).

Property (iii) $\mathcal{S}_i : 1 \le i \le M$ are disjoint and

$$\Pr(X^n \in \cup_{i=1}^M \mathcal{S}_i) \ge 1 - \delta.$$

Based on this system, we construct an nth-order ID source code $\mathcal{C}_n = (f_n, B_n, g_n)$ as follows. For each $x^n \notin \cup_{i=1}^M \mathcal{S}_i$, the encoder simply sends the sequence x^n itself to the decoder. After receiving x^n, the decoder outputs 1 if $\rho_n(x^n, y^n) \le d$ and 0 otherwise. The number of bits needed for the lossless transmission of x^n is $\lceil n \log |\mathcal{X}| \rceil$ plus one bit flag indicating $x^n \notin \cup_{i=1}^M \mathcal{S}_i$. For each $x^n \in \cup_{i=1}^M \mathcal{S}_i$, the encoder first finds the integer i such that $x^n \in \mathcal{S}_i$ and then transmitts

i to the decoder. Upon receiving i, the decoder outputs 1 if $\rho_n(\mathcal{S}_i, y^n) \le d$, and 0 otherwise, where

$$\rho_n(\mathcal{S}_i, y^n) = \min\{\rho_n(\tilde{x}^n, y^n) : \tilde{x}^n \in \mathcal{S}_i\}.$$

The number of bits needed for the transmission of the integer i is $\lceil \log M \rceil$ plus one bit flag indicating $x^n \in \cup_{i=1}^M \mathcal{S}_i$. Therefore the average rate of the IDS code described above is upper bounded by

$$\begin{aligned} r_n(\mathcal{C}_n) &= \frac{1}{n} \Pr(X^n \in \cup_{i=1}^M \mathcal{S}_i)(\lceil \log M \rceil + 1) + \frac{1}{n} \Pr(X^n \notin \cup_{i=1}^M \mathcal{S}_i)(\lceil n \log |\mathcal{X}| \rceil + 1) \\ &\le I(X \wedge U) + \delta + \delta \log |\mathcal{X}| + \frac{2}{n} \\ &\le R + (1 + \log |\mathcal{X}|)\delta + \frac{2}{n}, \end{aligned} \tag{105}$$

where the last inequality is due to (104). From the construction of $\mathcal{C}_n$, it is clear that the probability of misrejection is zero and the probability of false identification is upper bounded by

$$\begin{aligned} P_{e2}(\mathcal{C}_n) &\le \frac{1}{\Pr(\rho_n(X^n, Y^n) > d)} \sum_{i=1}^M \Pr(X^n \in \mathcal{S}_i) \Pr(Y^n \in \mathcal{S}_i^d) \\ &\le 2 \sum_{i=1}^M \Pr(X^n \in \mathcal{S}_i) \Pr(Y^n \in \mathcal{S}_i^d) \end{aligned} \tag{106}$$

for sufficiently large n, where

$$\mathcal{S}_i^d = \{y^n \in \mathcal{Y}^n : \rho_n(\mathcal{S}_i, y^n) \le d\} \tag{107}$$

and the last inequality is due to the fact that $d < \mathbf{E}\rho(X, Y)$. To continue (106), note that for $1 \le i \le M$ and $x^n \in \mathcal{S}_i$, $(u^n(i), x^n)$ is $(UX, 2\gamma_n)$-typical. Since $\mathcal{E}(P_{XU}, d) \ge \beta - \delta$, it follows from Lemma 1 that for sufficiently large n and any $x^n \in \mathcal{S}_i$,

$$\mathcal{E}(P_{x^n u^n(i)}, d) \ge \beta - 2\delta . \tag{108}$$

Clearly, for each $1 \le i \le M$,

$$\mathcal{S}_i^d = \bigcup_{V \in \mathcal{V}_n(P_{u^n(i)}, \mathcal{U} \times \mathcal{Y})} \mathcal{S}_i^d \cap T_V^n(u^n(i), \mathcal{Y}). \tag{109}$$

It is easy to see that if $\mathcal{S}_i^d \cap T_V^n(u^n(i), \mathcal{Y})$ is not empty, then there exists $x^n \in \mathcal{S}_i$ and $Q \in \mathcal{P}_n(\mathcal{U} \times \mathcal{X} \times \mathcal{Y})$ such that

(i) the marginal of Q on $\mathcal{U} \times \mathcal{X}$ is $P_{u^n(i)x^n}$;

(ii) the marginal of Q on $\mathcal{U} \times \mathcal{Y}$ is given by

$$P_{u^n(i)}(u)V(y|u), \quad u \in \mathcal{U} \ \ \textit{and} \ \ y \in \mathcal{Y};$$

(iii) under the distribution Q, $\mathbf{E}\rho(X_0, Y_0) \leq d$.

In view of (16) and (108), this implies

$$\begin{aligned}\sum_{u\in\mathcal{U}} P_{u^n(i)}(u)D(V(\cdot|u)||P_Y) &= \sum_{u\in\mathcal{U},y\in\mathcal{Y}} P_{u^n(i)}(u)V(y|u)\log\frac{V(y|u)}{P_Y(y)} \\ &> \beta - 2\delta\ . \end{aligned} \tag{110}$$

Therefore, if $\mathcal{S}_i^d \cap T_V^n(u^n(i), \mathcal{Y})$ is not empty, then

$$\begin{aligned}\Pr(Y^n \in \mathcal{S}_i^d \cap T_V^n(u^n(i), \mathcal{Y})) &\leq \Pr\{Y^n \in T_V^n(u^n(i), \mathcal{Y})\} \\ &= |T_V^n(u^n(i), \mathcal{Y})| 2^{-n[H(V|P_{u^n(i)}) + \sum_{u\in\mathcal{U}} P_{u^n(i)}(u)D(V(\cdot|u)||P_Y)]} \\ &\stackrel{1)}{\leq} 2^{-n\sum_{u\in\mathcal{U}} P_{u^n(i)}(u)D(V(\cdot|u)||P_Y)} \\ &\leq 2^{-n(\beta-2\delta)}, \end{aligned} \tag{111}$$

where the inequality 1) is due to (100) and the last inequality is due to (110). Going back to (109), we get for sufficiently large n

$$\begin{aligned}\Pr(Y^n \in \mathcal{S}_i^d) &\leq |\mathcal{V}_n(P_{u^n(i)}, \mathcal{U}\times\mathcal{Y})| 2^{-n(\beta-2\delta)} \\ &\leq 2^{-n(\beta-3\delta)}, \end{aligned} \tag{112}$$

where the last inequality is due to (98). Substituting (112) into (106) yields

$$P_{e2}(\mathcal{C}_n) \leq 2^{-n(\beta-3\delta)}. \tag{113}$$

Since $\delta > 0$ is arbitrary, by definition, (105) and (113) imply that $(R, +\infty, \beta)$ is achievable. This completes the proof of the direct part of Theorem 2.

We next turn to the converse part, that is

$$R^*_{XY}(+\infty, \beta, d) \geq \bar{R}(P_X, P_Y, \beta, d).$$

Clearly, it is enough to prove that for any achievable triple $(R, +\infty, \beta)$

$$R \geq \bar{R}(P_X, P_Y, \beta, d).$$

To this end, let us below fix an achievable triple $(R, +\infty, \beta)$. By definition, there exists for any $\epsilon > 0$ a sequence of ID source codes $\mathcal{C}_n = (f_n, B_n, g_n)$ such that for sufficiently large n,

$$r_n(\mathcal{C}_n) \leq R + \epsilon, P_{e1}(\mathcal{C}_n) = 0, \quad and \quad P_{e2}(\mathcal{C}_n) \leq 2^{-n(\beta-\epsilon)}. \tag{114}$$

As in the binary symmetric case, we define for each $b^n \in B_n$

$$\mathcal{S}(b^n) = \{x^n \in \mathcal{X}^n : f_n(x^n) = b^n\}$$

and

$$\mathcal{S}^d(b^n) = \{y^n \in \mathcal{Y}^n : \rho_n(x^n, y^n) \leq d \text{ for some } x^n \in S(b^n)\} .$$

Since $P_{e1}(\mathcal{C}_n) = 0$, we must have

$$\mathcal{S}^d(b^n) \subset \{y^n \in \mathcal{Y}^n : g_n(y^n, b^n) = 1\}.$$

Therefore, the inequality $P_{e2}(\mathcal{C}_n) \leq 2^{-n(\beta-\epsilon)}$ implies

$$\sum_{b^n \in B_n} \Pr(X^n \in \mathcal{S}(b^n)) \Pr(Y^n \in \mathcal{S}^d(b^n)) \leq 2^{-n(\beta-\epsilon)} + \Pr(\rho_n(X^n, Y^n) \leq d).$$

In view of (15) and the fact that $\beta < \beta(d)$, it follows that for sufficiently large n

$$\sum_{b^n \in B_n} \Pr(X^n \in \mathcal{S}(b^n)) \Pr(Y^n \in \mathcal{S}^d(b^n)) \leq 2^{-n(\beta-\epsilon)}.$$

From Markov' s inequality, one gets that

$$\sum_{b^n \in B'_n} Pr(X^n \in \mathcal{S}(b^n)) \geq 1 - \epsilon, \tag{115}$$

where

$$B'_n = \{b^n \in B_n : \Pr\{Y^n \in S^d(b^n)\} \leq 2^{-n(\beta-\epsilon-n^{-1}\log(2/\epsilon))}\} .$$

Let m be a sufficiently large positive integer to be specified later. Fix a $b^n \in B'_n$. Applying the inherently typical subset lemma(i.e., Lemma 4) to $S(b^n) \cap T^n_{X,r_n}$, where T^n_{X,r_n} is the set of all (X, r_n)-typical sequences x^n, we get an m-inherently typical subset $A \subset S(b^n) \cap T^n_{X,r_n}$ such that

$$\frac{1}{n} \log \frac{|S(b^n) \cap T^n_{X,r_n}|}{|A|} \leq |\mathcal{X}|(m+1)^{|\mathcal{X}|} \frac{\log(n+1)}{n} . \tag{116}$$

By the definition of m-inherently typical subsets, there exists a mapping ϕ from $\cup_{i=0}^{n-1} A_i$ to $\mathcal{U}_m$, where $A_i = \{x^i \in \mathcal{X}^i : x^i \text{ is a prefix of some element in } A\}$, such that the following hold:

(i) There exists an n-type $Q \in \mathcal{P}_n(\mathcal{X} \times \mathcal{U}_m)$ such that for any $x^n \in A$,

$$P_{x^n u^n}(x,u) = Q(x,u) \ , \qquad x \in \mathcal{X} \text{ and } u \in \mathcal{U}_m \ , \tag{117}$$

where $u^n \in \mathcal{U}_m^n$ is the sequence associated with x^n through ϕ.

(ii) If $(\hat{X}, \hat{U})$ is a random vector taking values on $\mathcal{X} \times \mathcal{U}_m$ with joint distribution Q, then

$$\frac{1}{n} \log |A| \le H(\hat{X}|\hat{U}) \le \frac{1}{n} \log |A| + \frac{\log^2 m}{m} \ . \tag{118}$$

As what we did in the analysis following Definition 1, let $\tilde{X}^n = (\tilde{X}_1, \cdots, \tilde{X}_n)$ be a random vector taking values uniformly on A. Define a random vector $\tilde{U}^n = (\tilde{U}_1, \cdots, \tilde{U}_n)$ by letting $\tilde{U}_i = \phi(\tilde{X}^{i-1})$, $1 \le i \le n$. Let I be a random variable taking values uniformly on $\{1, 2, \cdots, n\}$ and independent of $\tilde{X}^n$. Let

$$\tilde{X} = \tilde{X}_I \ , \quad \tilde{U} = \tilde{U}_I \ , \quad \text{and } U = (\tilde{X}^{I-1}, I) \ . \tag{119}$$

Clearly, if we extend the mapping ϕ in the obvious way so that $\phi(U) = \phi(\tilde{X}^{I-1})$ whenever $U = (\tilde{X}^{I-1}, I)$, then $\tilde{U} = \phi(U)$. As pointed out in the analysis following Definition 1, $\tilde{X}$ and $\tilde{U}$ have the joint distribution $P_{\tilde{X}\tilde{U}} = Q$. Furthermore, $n^{-1} \log |A| = H(\tilde{X}|U)$ and (118) can be rewritten as

$$H(\tilde{X}|U) \le H(\tilde{X}|\tilde{U}) \le H(\tilde{X}|U) + \frac{\log^2 m}{m}. \tag{120}$$

Having defined the random pair $(\tilde{X}, \tilde{U})$ taking values on $\mathcal{X} \times \mathcal{U}_m$, we next lower bound $\Pr(Y^n \in A^d)$ by a function of $\mathcal{E}(P_{\tilde{X}\tilde{U}}, d - \epsilon)$, where

$$A^d = \{y^n \in \mathcal{Y}^n : \rho_n(x^n, y^n) \le d \text{ for some } x^n \in A\}.$$

In view of the definition of $\mathcal{E}(P_{\tilde{X}\tilde{U}}, d - \epsilon)$, let $\tilde{Y}$ be a random variable taking values on $\mathcal{Y}$ such that $\mathbf{E}\rho(\tilde{X}, \tilde{Y}) \le d - \epsilon$. Let $V = (V(y|xu))_{x \in \mathcal{X}, u \in \mathcal{U}_m, y \in \mathcal{Y}}$ be a stochastic matrix so that $V(y|xu)$ is the conditional probability of $\tilde{Y} = y$ given $\tilde{X} = x$ and $\tilde{U} = u$. Let $\tilde{Y}^n = (\tilde{Y}_1, \cdots, \tilde{Y}_n)$ be the random vector resulting from passing $(\tilde{X}^n, \tilde{U}^n)$ through the channel V. From (103), it follows that for any $x^n \in A$,

$$\Pr(\tilde{Y}^n \in T^n_{\tilde{Y}|\tilde{X}\tilde{U}, \gamma_n}(x^n u^n) | \tilde{X}^n = x^n, \tilde{U}^n = u^n) > 1 - \frac{|\mathcal{X}||\mathcal{Y}||\mathcal{U}_m|}{4n\gamma_n^2}, \tag{121}$$

where $u^n \in \mathcal{U}_m^n$ is the sequence associated with x^n through ϕ. From (117), it is not hard to see that for any $x^n \in A$,

$$T^n_{\tilde{Y}|\tilde{X}\tilde{U},\gamma_n}(x^n, u^n) \subset T^n_{\tilde{Y},|\tilde{X}||\tilde{U}|\gamma_n}. \tag{122}$$

Furthermore, since $\mathbf{E}\rho(\tilde{X}, \tilde{Y}) \leq d - \epsilon$, it follows that for sufficiently large n and any $y^n \in T^n_{\tilde{Y}|\tilde{X}\tilde{U},\gamma_n}(x^n u^n)$,

$$\rho_n(x^n, y^n) \leq d.$$

Therefore, if we let

$$F = \cup_{x^n \in A} T^n_{\tilde{Y}|\tilde{X}\tilde{U},\gamma_n}(x^n u^n),$$

where $u^n \in \mathcal{U}_m^n$ is the sequence associated with x^n through ϕ, then $F \subset A^d$ and for any $x^n \in A$,

$$\Pr(\tilde{Y}^n \in F|\tilde{X}^n = x^n, \tilde{U}^n = u^n) > 1 - \frac{|\mathcal{X}||\mathcal{Y}||\mathcal{U}_m|}{4n\gamma_n^2}.$$

This implies

$$\Pr(\tilde{Y}^n \in F) > 1 - \frac{|\mathcal{X}||\mathcal{Y}||\mathcal{U}_m|}{4n\gamma_n^2}. \tag{123}$$

For convenience, let

$$\delta_n = \frac{|\mathcal{X}||\mathcal{Y}||\mathcal{U}_m|}{4n\gamma_n^2}.$$

From (123), we have

$$H(\tilde{Y}^n) \leq h(\delta_n) + \log|F| + n\delta_n \log|\mathcal{Y}|,$$

where $h(\cdot)$ represents the binary entropy function. From this, it is not hard to verify that

$$\begin{aligned}
\log|F| &\geq H(\tilde{Y}^n) - h(\delta_n) - n\delta_n \log|\mathcal{Y}| \\
&= \sum_{i=1}^n H(\tilde{Y}_i|\tilde{Y}^{i-1}) - h(\delta_n) - n\delta_n \log|\mathcal{Y}| \\
&\geq \sum_{i=1}^n H(\tilde{Y}_i|\tilde{Y}^{i-1}\tilde{X}^{i-1}\tilde{U}^{i-1}) - h(\delta_n) - n\delta_n \log|\mathcal{Y}| \\
&=^{1)} \sum_{i=1}^n H(\tilde{Y}_i|\tilde{X}^{i-1}\tilde{U}^{i-1}) - h(\delta_n) - n\delta_n \log|\mathcal{Y}| \\
&=^{2)} \sum_{i=1}^n H(\tilde{Y}_i|\tilde{X}^{i-1}) - h(\delta_n) - n\delta_n \log|\mathcal{Y}|,
\end{aligned} \tag{124}$$

where equality 1) is due to the fact that given $(\tilde{X}^{i-1}, \tilde{U}^{i-1})$, $\tilde{Y}^{i-1}$ and $\tilde{Y}_i$ are conditionally independent, and the equality 2) follows from the fact that $\tilde{U}^{i-1} = \phi(\tilde{X}^{i-2})$. Recall that I is the random variable which takes values uniformly on $\{1, \cdots, n\}$ and is independent of $\tilde{X}^n$

and $\tilde{Y}^n$. Let $\bar{Y} = \tilde{Y}_I$. In view of (119), (124) continues as follows

$$\log|F| \geq nH(\bar{Y}|U)) - h(\delta_n) - n\delta_n \log|\mathcal{Y}|$$

$$= nH(\bar{Y}|\tilde{U}) + n(H(\bar{Y}|U) - H(\bar{Y}|\tilde{U})) - h(\delta_n) - n\delta_n \log|\mathcal{Y}|$$

$$= nH(\tilde{Y}|\tilde{U}) + n(H(\bar{Y}|U) - H(\bar{Y}|\tilde{U})) - h(\delta_n) - n\delta_n \log|\mathcal{Y}| \tag{125}$$

where the last equality follows from the observation that $(\tilde{X}, \tilde{U}, \bar{Y})$ has the same joint distribution as that of $(\tilde{X}, \tilde{U}, \tilde{Y})$. To continue (125) further, we next estimate the difference $H(\bar{Y}|\tilde{U}) - H(\bar{Y}|U)$. Since $\tilde{U} = \phi(U)$, it is not hard to verify that

$$H(\bar{Y}|\tilde{U}) - H(\bar{Y}|U) =$$

$$= \frac{1}{n}\sum_{i=1}^{n} \sum_{x^{i-1}\in A_{i-1}} \Pr(\tilde{X}^{i-1} = x^{i-1}) \sum_{y\in\mathcal{Y}} \Pr(\tilde{Y}_i = y|\tilde{X}^{i-1} = x^{i-1}) \log \frac{\Pr(\tilde{Y}_i = y|\tilde{X}^{i-1} = x^{i-1})}{P_{\bar{Y}|\tilde{U}}(y|\phi(x^{i-1}))}, \tag{126}$$

where $P_{\bar{Y}|\tilde{U}}(y|\phi(x^{i-1}))$ is the conditional probability of $\bar{Y} = y$ given $\tilde{U} = \phi(x^{i-1})$. By construction, it is not hard to see that

$$\Pr(\tilde{Y}_i = y|\tilde{X}^{i-1} = x^{i-1}) = \sum_{x\in\mathcal{X}} \Pr(\tilde{X}_i = x|\tilde{X}^{i-1} = x^{i-1}) V(y|x\phi(x^{i-1}))$$

and

$$P_{\bar{Y}|\tilde{U}}(y|\phi(x^{i-1})) = \sum_{x\in\mathcal{X}} P_{\tilde{X}|\tilde{U}}(x|\phi(x^{i-1})) V(y|x\phi(x^{i-1})).$$

Using the log-sum inequality, one gets that

$$\sum_{y\in\mathcal{Y}} \Pr(\tilde{Y}_i = y|\tilde{X}^{i-1} = x^{i-1}) \log \frac{\Pr(\tilde{Y}_i = y|\tilde{X}^{i-1} = x^{i-1})}{P_{\bar{Y}|\tilde{U}}(y|\phi(x^{i-1}))}$$

$$\leq \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} \Pr(\tilde{X}_i = x|\tilde{X}^{i-1} = x^{i-1}) V(y|x\phi(x^{i-1})) \log \frac{\Pr(\tilde{X}_i = x|\tilde{X}^{i-1} = x^{i-1})}{P_{\tilde{X}|\tilde{U}}(x|\phi(x^{i-1}))}$$

$$\leq \sum_{x\in\mathcal{X}} \Pr(\tilde{X}_i = x|\tilde{X}^{i-1} = x^{i-1}) \log \frac{\Pr(\tilde{X}_i = x|\tilde{X}^{i-1} = x^{i-1})}{P_{\tilde{X}|\tilde{U}}(x|\phi(x^{i-1}))}. \tag{127}$$

Substituting (127) into (126) yields

$$H(\bar{Y}|\tilde{U}) - H(\bar{Y}|U) \leq H(\tilde{X}|\tilde{U}) - H(\tilde{X}|U)$$

$$\leq \frac{\log^2 m}{m}, \tag{128}$$

where the last inequality is due to (120). Combining (125) and (128) yields

$$\log |F| \geq nH(\tilde{Y}|\tilde{U}) - n\frac{\log^2 m}{m} - -h(\delta_n) - n\delta_n \log |\mathcal{Y}|.$$

From (122),

$$F \subset T^n_{\tilde{Y},|\mathcal{X}||\mathcal{U}_m|\gamma_n}.$$

Thus,

$$\Pr(Y^n \in F) \geq |F| 2^{-n(H(\tilde{Y})+D(P_{\tilde{Y}}||P_Y)+o(1))}$$

$$\geq \exp\{-n(I(\tilde{U} \wedge \tilde{Y}) + D(P_{\tilde{Y}}||P_Y) + \frac{\log^2 m}{m} + \epsilon_n)\}\ , \tag{129}$$

where $\epsilon_n \to 0$ as n goes to infinity. Since $F \subset A^d$, (129) implies

$$\Pr(Y^n \in A^d) \geq \exp\{-n(I(\tilde{U} \wedge \tilde{Y}) + D(P_{\tilde{Y}}||P_Y) + \frac{\log^2 m}{m} + \epsilon_n)\}. \tag{130}$$

Note that (130) holds for any random variable $\tilde{Y}$ taking values on $\mathcal{Y}$ such that $\mathbf{E}\rho(\tilde{X},\tilde{Y}) \leq d - \epsilon$. This, together with the definition of $\mathcal{E}(P_{\tilde{X}\tilde{U}}, d-\epsilon)$, implies

$$\Pr(Y^n \in A^d) \geq \exp\{-n(\mathcal{E}(P_{\tilde{X}\tilde{U}}, d-\epsilon) + \frac{\log^2 m}{m} + \epsilon_n)\}. \tag{131}$$

Let us go back to (115) and (116). We next want to estimate the probability $\Pr\{X^n \in \mathcal{S}(b^n) \cap T^n_{X,\gamma_n}\}$, where $b^n \in B'_n$. Since $A \subset \mathcal{S}(b^n) \cap T^n_{X,\gamma_n}$, it is easy to see from (117) and (119) that $P_{\tilde{X}}$ is (X, γ_n)-essential, that is

$$|P_{\tilde{X}}(x) - P_X(x)| \leq \gamma_n, \quad x \in \mathcal{X} \tag{132}$$

and $P_{\tilde{X}}(x)$ whenever $P_X(x) = 0$. For convenience, let

$$a_n = |\mathcal{X}|(m+1)^{|\mathcal{X}|}\frac{\log(n+1)}{n}.$$

It is not hard to see that

$$\begin{aligned}
\Pr(X^n \in \mathcal{S}(b^n) \cap T^n_{X,\gamma_n}) &\leq |\mathcal{S}(b^n) \cap T^n_{X,\gamma_n}| 2^{-n(H(\tilde{X})-o(1))} \\
&\leq^{1)} 2^{-n(H(\tilde{X}) - n^{-1}\log|A| - a_n - o(1))} \\
&\leq^{2)} 2^{-n(H(\tilde{X}) - H(\tilde{X}|\tilde{U}) - a_n - o(1))} \\
&= 2^{-n(I(\tilde{X}\wedge\tilde{U}) - \epsilon'_n)}\ ,
\end{aligned} \tag{133}$$

where the inequality 1) is due to (116), the inequality 2) is due to the fact that $n^{-1}\log|A| \le H(\tilde{X}|\tilde{U})$ and ϵ'_n goes to zero as n goes to infinity. Since $b^n \in B'_n$, it follows that

$$\begin{aligned}\Pr(Y^n \in A^d) &\le \Pr(Y^n \in \mathcal{S}^d(b^n)) \\ &\le 2^{-n(\beta-\epsilon-\frac{1}{n}\log\frac{2}{\epsilon})}. \qquad (134)\end{aligned}$$

Comparing (134) with (131) yields

$$\mathcal{E}(P_{\tilde{X}\tilde{U}}, d-\epsilon) \ge \beta - \epsilon - \frac{\log^2 m}{m} - \epsilon_n - \frac{1}{n}\log\frac{2}{\epsilon}.$$

Note that $\tilde{U}$ takes values on $\mathcal{U}_m$. From the definition of $R_k(P_X, P_Y, \beta, d)$, it follows that

$$I(\tilde{X}\wedge\tilde{U}) \ge R_{|\mathcal{U}_m|}(P_{\tilde{X}}, P_Y, \beta-\epsilon-\frac{\log^2 m}{m}-\epsilon_n-\frac{1}{n}\log\frac{2}{\epsilon}, d-\epsilon)$$

which, combined with (133), implies

$$-\frac{1}{n}\log\Pr(X^n \in \mathcal{S}(b^n)\cap T^n_{X,\gamma_n}) \ge R_{|\mathcal{U}_m|}(P_{\tilde{X}}, P_Y, \beta-\epsilon-\frac{\log^2 m}{m}-\epsilon_n-\frac{1}{n}\log\frac{2}{\epsilon}, d-\epsilon)-\epsilon'_n. \quad (135)$$

In view of Fact 1 in Subsection A and (132), (135) continues as follows.

$$-\frac{1}{n}\log\Pr(X^n \in \mathcal{S}(b^n)\cap T^n_{X,\gamma_n}) \ge R_{|\mathcal{U}_m|}(P_X, P_Y, \beta-\epsilon-\frac{\log^2 m}{m}, d-\epsilon)-\epsilon''_n\,, \quad (136)$$

where ϵ''_n goes to zero as n goes to infinity.

Note that (136) holds for any $b^n \in B'_n$. Now it is not hard to verify that

$$\begin{aligned}R+\epsilon &\ge r_n(\mathcal{C}_n) \\ &\ge \frac{1}{n}H(f_n(X^n)) \\ &\ge \sum_{b^n\in B'_n} -\frac{1}{n}\Pr(X^n \in \mathcal{S}(b^n)\cap T^n_{X,\gamma_n})\log\Pr(X^n \in \mathcal{S}(b^n)\cap T^n_{X,\gamma_n}) - \frac{1}{n} \\ &\ge (1-\epsilon-\frac{|\mathcal{X}|}{4n\gamma_n^2})(R_{|\mathcal{U}_m|}(P_X, P_Y, \beta-\epsilon-\frac{\log^2 m}{m}, d-\epsilon)-\epsilon''_n)-\frac{1}{n}, \qquad (137)\end{aligned}$$

where the last inequality is due to (115) and the following inequality

$$\Pr(X^n \in T^n_{X,\gamma_n}) \ge 1-\frac{|\mathcal{X}|}{4n\gamma_n^2}.$$

In view of Fact 1, letting $n \to +\infty$ and then letting $\epsilon \to 0$ in (137) yield

$$R \geq R_{|\mathcal{U}_m|}(P_X, P_Y, \beta - \frac{\log^2 m}{m}, d).$$

Since $(R, +\infty, \beta)$ is an arbitrary achievable triple, this implies

$$R^*_{XY}(+\infty, \beta, d) \geq R_{|\mathcal{U}_m|}(P_X, P_Y, \beta - \frac{\log^2 m}{m}, d). \tag{138}$$

Thus, for sufficiently large m,

$$R^*_{XY}(+\infty, \beta, d) \geq R(P_X, P_Y, \beta - \frac{\log^2 m}{m}, d).$$

Letting m go to infinity yields

$$R^*_{XY}(+\infty, \beta, d) \geq \bar{R}(P_X, P_Y, \beta, d).$$

This completes the proof of the converse part and hence the proof of Theorem 2.

Remark 3. In response to Remark 1, let us note that in view of Lemma 4, the inequality (138) actually holds for any $m \geq 2^{16|\mathcal{X}|^2}$. From the proof of Theorem 2, therefore we obtain that for any $m \geq 2^{16|\mathcal{X}|^2}$,

$$\begin{aligned} R_{|\mathcal{U}_m|}(P_X, P_Y, \beta - \frac{\log^2 m}{m}, d) &\leq R^*_{XY}(+\infty, \beta, d) \\ &= \bar{R}(P_X, P_Y, \beta, d) \\ &\leq R(P_X, P_Y, \beta, d) \\ &\leq R_{|\mathcal{U}_m|}(P_X, P_Y, \beta, d). \end{aligned} \tag{139}$$

This gives us in a sense how accurate the value obtained could be if we approximate $\bar{R}(P_X, P_Y, \beta, d)$ by $R_{|\mathcal{U}_m|}(P_X, P_Y, \beta, d)$. If some regular conditions are satisfied, hopefully this approximation could be as accurate as $O(\log^2 m/m)$.

B Proof of Theorem 3

We next turn to the proof of Theorem 3. Although the proof of Theorem 3 is more complicated than that of Theorem 2, the basic idea is the same and in fact, most parts of the proof are just the translation of the corresponding parts in the proof of Theorem 2 to the present case. This is why we stated separately Theorems 2 and 3. Hope this will help the reader understand the proofs more easily.

Proof of Theorem 3. In view of Remark 2, it suffices to prove Theorem 3 for $\alpha > \beta(P_X, d) - \beta(d)$ and $0 < \beta < \beta(d)$. We first prove the direct part, that is

$$R^*_{XY}(\alpha, \beta, d) \leq \bar{R}(P_X, P_Y, \alpha, \beta(d), \beta, d).$$

By the definition of $R^*_{XY}(\alpha, \beta, d)$, it is enough to prove that for any R satisfying

$$R > \bar{R}(P_X, P_Y, \alpha, \beta(d), \beta, d),$$

(R, α, β) is achievable. To this end, let us fix below $R > \bar{R}(P_X, P_Y, \alpha, \beta(d), \beta, d)$. As in the proof of Theorem 2, it is not hard to see that for any $\delta > 0$, there exists a random variable U taking values on some finite set $\mathcal{U}$ such that

$$I(X \wedge U) < R \text{ and } \mathcal{E}(P_{XU}, \alpha, \beta(d), d) \geq \beta - \delta. \tag{140}$$

Corresponding to the random pair (X, U), there exists for sufficiently large n a system $\{(u^n(i), \mathcal{S}_i) | 1 \leq i \leq M\}$ which satisfies properties (i)–(iii). Based on this system, we construct an n-th order ID source code $\mathcal{C}_n = (f_n, B_n, g_n)$ as follows. For each $x^n \notin \cup_{i=1}^M \mathcal{S}_i$, the encoder simply sends the sequence x^n itself to the decoder. After receiving x^n, the decoder outputs 1 if $\rho_n(x^n, y^n) \leq d$ and 0 otherwise. For each $x^n \in \cup_{i=1}^M \mathcal{S}_i$, the encoder first finds the integer i such that $x^n \in \mathcal{S}_i$ and then transmitts i to the decoder. Upon receiving i, the decoder outputs 1 if $y^n \in \mathcal{Y}^n$ satisfies that there exists some $x^n \in S_i$ such that $\rho_n(x^n, y^n) \leq d$ and

$$\sum_{x \in \mathcal{X}, u \in \mathcal{U}} P_{x^n u^n(i)}(x, u) D(P_{y^n | x^n u^n(i)}(\cdot | xu) || P_Y) \leq \beta(d) + \alpha, \tag{141}$$

where

$$P_{y^n | x^n u^n(i)}(\cdot | xu) \in \mathcal{V}_n(P_{x^n u^n(i)}, (\mathcal{X} \times \mathcal{U}) \times \mathcal{Y})$$

is defined by $P_{x^n u^n(i)}(x, u) P_{y^n | x^n u^n(i)}(y | xu) = P_{x^n u^n(i) y^n}(x, u, y)$ for all $y \in \mathcal{Y}$; and otherwise outputs 0. Clearly, the encoder f_n defined here is the same as in the proof of Theorem 2. From (105), therefore, the average rate $r_n(\mathcal{C}_n)$ is also upper bounded by

$$r_n(\mathcal{C}_n) \leq R + (1 + \log |\mathcal{X}|)\delta + \frac{2}{n}. \tag{142}$$

For each $1 \leq i \leq M$, let

$$\mathcal{S}_i^d = \{y^n \in \mathcal{Y}^n : \rho_n(x^n, y^n) \leq d \text{ for some } x^n \in S_i\} \tag{143}$$

$$\hat{\mathcal{S}}_i^d = \{y^n \in \mathcal{Y}^n : g_n(y^n, i) = 1\} \tag{144}$$

and

$$\bar{\mathcal{S}}_i^d = \mathcal{Y}^n - \hat{\mathcal{S}}_i^d. \tag{145}$$

Obviously, $\hat{S}_i^d \subset S_i^d$ for $1 \le i \le M$. For each $x^n \in \mathcal{S}_i$, let $B^i(x^n)$ denote the set of all $y^n \in \mathcal{Y}^n$ such that $\rho_n(x^n, y^n) \le d$ and

$$\sum_{x\in\mathcal{X}, u\in\mathcal{U}} P_{x^n u^n(i)}(x,u) D(P_{y^n|x^n u^n(i)}(\cdot|xu)||P_Y) > \beta(d) + \alpha. \tag{146}$$

From (146), it is not hard to see that

$$\Pr(Y^n \in B^i(x^n)) \le (n+1)^{|\mathcal{X}||\mathcal{U}||\mathcal{Y}|} 2^{-n(\beta(d)+\alpha)} . \tag{147}$$

By the construction of the ID source code $\mathcal{C}_n$, we can now verify that

$$\begin{aligned} P_{e1}(\mathcal{C}_n) &= \frac{\Pr\{(X^n, Y^n) \in \cup_{i=1}^M \mathcal{S}_i \times \bar{\mathcal{S}}_i^d, \& \rho_n(X^n, Y^n) \le d\}}{\Pr\{\rho_n(X^n, Y^n) \le d\}} \\ &\le \frac{1}{\Pr\{\rho_n(X^n, Y^n) \le d\}} \sum_{i=1}^M \sum_{x^n \in \mathcal{S}_i} \Pr(X^n = x^n) \Pr(Y^n \in B^i(x^n)) \\ &\le^{1)} \frac{(n+1)^{|\mathcal{X}||\mathcal{U}||\mathcal{Y}|}}{\Pr(\rho_n(X^n, Y^n) \le d)} 2^{-n(\beta(d)+\alpha)} \\ &\le^{2)} 2^{-n(\alpha-\delta)} \end{aligned} \tag{148}$$

for sufficiently large n, where the inequality 1) follows from (147) and the inequality 2) follows from (15). As in the proof of the direct part of Theorem 2, it is clear that the probability $P_{e2}(\mathcal{C}_n)$ of false identification of the ID source code $\mathcal{C}_n$ is upper bounded by

$$\begin{aligned} P_{e2}(\mathcal{C}_n) &\le \frac{1}{\Pr\{\rho_n(X^n, Y^n) > d\}} \sum_{i=1}^M \Pr(X^n \in \mathcal{S}_i) \; Pr(Y^n \in \hat{\mathcal{S}}_i^d) \\ &\le 2 \sum_{i=1}^M \Pr(X^n \in \mathcal{S}_i) \Pr(Y^n \in \hat{\mathcal{S}}_i^d) \end{aligned} \tag{149}$$

for sufficiently large n. To continue (149), we do the same thing as we did before. First note that for any $1 \le i \le M$ and $x^n \in \mathcal{S}_i$, $(u^n(i), x^n)$ is $(UX, 2\gamma_n)$-typical. In view of Lemma 2 and (140), therefore, it follows that for sufficiently large n and any $x^n \in \mathcal{S}_i$,

$$\mathcal{E}(P_{x^n u^n(i)}, \alpha, \beta(d), d) \ge \beta - 2\delta . \tag{150}$$

Let us now look at $\hat{\mathcal{S}}_i^d \cap T_V^n(u^n(i), \mathcal{Y})$, where $V \in \mathcal{V}_n(P_{u^n(i)}, \mathcal{U} \times \mathcal{Y})$. Clearly, if $\hat{\mathcal{S}}_i^d \cap T_V^n(u^n(i), \mathcal{Y})$ is not empty, then from the definition of $\hat{\mathcal{S}}_i^d$, there exist a $x^n \in \mathcal{S}_i$ and a $Q \in \mathcal{P}_n(\mathcal{U} \times \mathcal{X} \times \mathcal{Y})$ such that

(i) the marginal of Q on $\mathcal{U} \times \mathcal{X}$ is $P_{u^n(i)x^n}$;

(ii) the marginal of Q on $\mathcal{U} \times \mathcal{Y}$ is given by

$$P_{u^n(i)}(u)V(y|u), \quad u \in \mathcal{U} \quad and \quad y \in \mathcal{Y};$$

(iii) if $(\tilde{U}, \tilde{X}, \tilde{Y})$ is a random vector taking values on $\mathcal{U} \times \mathcal{X} \times \mathcal{Y}$ with joint distribution $P_{\tilde{X}\tilde{U}\tilde{Y}} = Q$, then $\mathbf{E}\rho(X_0, Y_0) \le d$ and $D(P_{\tilde{Y}}||P_Y) + I(\tilde{X}\tilde{U} \wedge \tilde{Y}) \le \beta(d) + \alpha$.

In view of the definition of $\mathcal{E}(P_{x^n u^n(i)}, \alpha, \beta(d), d)$ and the inequality (150), this implies

$$\sum_{u \in \mathcal{U}} P_{u^n(i)}(u) D(V(\cdot|u)||P_Y) > \beta - 2\delta \ . \tag{151}$$

Therefore, if $\hat{\mathcal{S}}_i^d \cap T_V^n(u^n(i), \mathcal{Y})$ is not empty, then

$$\begin{aligned} \Pr\{Y^n \in \hat{\mathcal{S}}_i^d \cap T_V^n(u^n(i), \mathcal{Y})\} &\le \Pr\{Y^n \in T_V^n(u^n(i), \mathcal{Y})\} \\ &\le 2^{-n \sum_{u \in \mathcal{U}} P_{u^n(i)}(u) D(V(\cdot|u)||P_Y))} \\ &\le 2^{-n(\beta - 2\delta)}, \end{aligned} \tag{152}$$

which implies

$$\begin{aligned} Pr(Y^n \in \hat{\mathcal{S}}_i^d) &\le |\mathcal{V}_n(P_{u^n(i)}, \mathcal{U} \times \mathcal{Y})| 2^{-n(\beta - 2\delta)} \\ &\le 2^{-n(\beta - 3\delta)} \end{aligned} \tag{153}$$

for sufficiently large n. Substituting (153) into (149) yields

$$P_{e2}(\mathcal{C}_n) \le 22^{-n(\beta - 3\delta)} \ . \tag{154}$$

Since $\delta > 0$ is arbitrary, (142), (148) and (154) implies that (R, α, β) is achievable. This completes the proof of the direct part of Theorem 3.

We next turn to the converse part. Clearly, it is enough to prove that for any achievable triple

(R, α, β)

$$R \geq \bar{R}(P_X, P_Y, \alpha, \beta(d)\beta, d).$$

To this end, let us below fix an achievable triple (R, α, β). By definition, there exists for any $\epsilon > 0$ a sequence of ID source codes $\mathcal{C}_n = (f_n, B_n, g_n)$ such that for sufficiently large n,

$$r_n(\mathcal{C}_n) \leq R + \epsilon, P_{e1}(\mathcal{C}_n) \leq 2^{-n(\alpha-\epsilon)}, \quad and \quad P_{e2}(\mathcal{C}_n) \leq 2^{-n(\beta-\epsilon)} . \tag{155}$$

For each $b^n \in B_n$, let

$$\mathcal{S}(b^n) = \{x^n \in \mathcal{X}^n : f_n(x^n) = b^n\}$$

and

$$G(b^n) = \{y^n \in \mathcal{Y}^n : g_n(y^n, b^n) = 1\}.$$

For each $x^n \in \mathcal{X}^n$, denote by $B(x^n)$ the set of sequences $y^n \in \mathcal{Y}^n$ such that $\rho_n(x^n, y^n) \leq d$ and $y^n \notin G(f_n(x^n))$. It is not hard to see that

$$P_{e1}(\mathcal{C}_n) = \frac{1}{\Pr(\rho_n(X^n, Y^n) \leq d)} \sum_{x^n \in \mathcal{X}^n} \Pr(X^n = x^n) \Pr(Y^n \in B(x^n)).$$

By virtue of (15) and (155), we have for sufficiently large n

$$\sum_{x^n \in \mathcal{X}^n} \Pr(X^n = x^n) \Pr(Y^n \in B(x^n)) \leq 2^{-n(\alpha - 2\epsilon + \beta(d))} . \tag{156}$$

Let

$$F_n = \{x^n \in \mathcal{X}^n : \Pr(Y^n \in B(x^n)) \leq \epsilon^{-1} 2^{-n(\alpha-2\epsilon+\beta(d))}\}.$$

From (156) and the Markov inequality,

$$\Pr(X^n \in F_n) \geq 1 - \epsilon . \tag{157}$$

As in the proof of the converse part of Theorem 2, it is not hard to prove that the inequality $P_{e2}(\mathcal{C}_n) \leq 2^{-n(\beta-\epsilon)}$ implies that for sufficiently large n,

$$\sum_{b^n \in B_n} \Pr(X^n \in \mathcal{S}(b^n)) \Pr(Y^n \in G(b^n)) \leq 2 2^{-n(\beta-\epsilon)}.$$

Usin Markov inequality once again, one gets that

$$\sum_{b^n \in B_n'} \Pr(X^n \in \mathcal{S}(b^n)) \geq 1 - \epsilon,$$

where

$$B_n' = \{b^n \in B_n | \Pr(Y^n \in G(b^n)) \leq 2^{-n(\beta - \epsilon - \frac{1}{n}\log\frac{2}{\epsilon})}\}.$$

We are now in a position to apply the inherently typical subset lemma (Lemma 4). Fix a $b^n \in B'_n$. Applying the lemma to $\mathcal{S}(b^n) \cap T^n_{X,\gamma_n} \cap F_n$, we get an m-inherently typical subset $A \subset \mathcal{S}(b^n) \cap T^n_{X,\gamma_n} \cap F_n$ such that

$$\frac{1}{n} \log \frac{|\mathcal{S}(b^n) \cap F_n \cap T^n_{X,\gamma_n}|}{|A|} \leq |\mathcal{X}|(m+1)^{|\mathcal{X}|} \frac{\log(n+1)}{n} . \tag{158}$$

The remaining proof is much the same as that shown in the proof of the converse part of Theorem 2. In what follows, therefore, we only point out places where changes are needed. (Unless otherwise specified, all notation below is the same as in the proof of the converse part of Theorem 2).

Having defined the random pair $(\tilde{X}, \tilde{U})$ taking values on $\mathcal{X} \times \mathcal{U}_m$, we, instead of lower bounding $\Pr(Y^n \in A^d)$ by a function of $\mathcal{E}(P_{\tilde{X}\tilde{U}}, d-\epsilon)$, lower bound $\Pr(Y^n \in G(b^n))$ by a function of $\mathcal{E}(P_{\tilde{X}\tilde{U}},, \alpha - 4\epsilon, \beta(d), d-\epsilon)$. In view of the definition of $\mathcal{E}(P_{\tilde{X}\tilde{U}},, \alpha - 4\epsilon, \beta(d), d-\epsilon)$, let $\tilde{Y}$ be a random variable taking values on $\mathcal{Y}$ such that

$$\mathbf{E}\rho(\tilde{X}, \tilde{Y}) \leq d - \epsilon \text{ and } D(P_{\tilde{Y}}||P_Y) + I(\tilde{X}\tilde{U} \wedge \tilde{Y}) \leq \beta(d) + \alpha - 4\epsilon . \tag{159}$$

Let $V = (V(y|xu))_{x\in\mathcal{X}, u\in\mathcal{U}_m, y\in\mathcal{Y}}$ be a stochastic matrix so that $V(y|xu)$ is the conditional probability of $\tilde{Y} = y$ given $\tilde{X} = x, \tilde{U} = u$. Let $\tilde{Y}^n = (\tilde{Y}_1, \cdots, \tilde{Y}_n)$ be a random vector resulting from passing $(\tilde{X}^n, \tilde{U}^n)$ through the channel V. From each $x^n \in A$, consider $T^n_{\tilde{Y}|\tilde{X}\tilde{U},r_n}(x^n, u^n)$, where $u^n \in \mathcal{U}^n_m$ is the sequence associated with x^n through ϕ. In view of (159), it is not hard to see that for sufficiently large n and for any $y^n \in T^n_{\tilde{Y}|\tilde{X}\tilde{U},r_n}(x^n, u^n)$,

$$\rho_n(x^n, y^n) \leq d \tag{160}$$

and

$$\sum_{x\in\mathcal{X}, u\in\mathcal{U}_m} P_{x^n u^n}(x, u) D(P_{y^n|x^n u^n}(\cdot|xu)||P_Y) \leq \beta(d) + \alpha - 3\epsilon. \tag{161}$$

Let $\hat{V} \in \mathcal{V}_n(P_{x^n u^n}, (\mathcal{X} \times \mathcal{U}_m) \times \mathcal{Y})$ be $(x^n u^n, \tilde{Y}|\tilde{X}\tilde{U}, \gamma_n)$-essential, then (161) implies

$$\sum_{x\in\mathcal{X}, u\in\mathcal{U}_m} P_{x^n u^n}(x, u) D(\hat{V}(\cdot|xu)||P_Y) \leq \beta(d) + \alpha - 3\epsilon. \tag{162}$$

Since $A \subset \mathcal{S}(b^n) \cap F_n \cap T^n_{X,\gamma_n}$, $x^n \in A$ implies $x^n \in F_n$. By definition of F_n, therefore, it follows that

$$\Pr(Y^n \in B(x^n)) \leq \epsilon^{-1} 2^{-n(\alpha+\beta(d)-2\epsilon)}. \tag{163}$$

By comparing (163) with (162), we can obtain that

$$|T^n_{\hat{V}}(x^n u^n, \mathcal{Y}) \cap B(x^n)| \leq^{1)} \epsilon^{-1} 2^{-n\epsilon}(n+1)^{|\mathcal{X}||\mathcal{Y}||\mathcal{U}_m|}|T^n_{\hat{V}}(x^n u^n, \mathcal{Y})|$$
$$\leq 2^{-\frac{n\epsilon}{2}}|T^n_{\hat{V}}(x^n u^n, \mathcal{Y})| \tag{164}$$

for sufficiently large n, where in derivation of 1), the following inequality was used:

$$|T^n_{\hat{V}}(x^n u^n, \mathcal{Y})| \geq (n+1)^{-|\mathcal{X}||\mathcal{Y}||\mathcal{U}_m|} \exp\{n \sum_{x \in \mathcal{X}, u \in \mathcal{U}_m} P_{x^n u^n}(x, u) H(\hat{V}(\cdot|x, u))\}.$$

From (164), it is now easy to check that

$$\Pr(\tilde{Y}^n \in T^n_{\tilde{Y}|\tilde{X}\tilde{U},\gamma_n}(x^n u^n) \cap B(x^n)|\tilde{X}^n = x^n, \tilde{U}^n = u^n)$$
$$= \sum_{\hat{V}} \Pr(\tilde{Y} \in T^n_{\hat{V}}(x^n u^n, \mathcal{Y}) \cap B(x^n)|\tilde{X}^n = x^n, \tilde{U}^n = u^n)$$
$$\leq 2^{-\frac{n\epsilon}{2}} \sum_{\hat{V}} \Pr(\tilde{Y} \in T^n_{\hat{V}}(x^n u^n, \mathcal{Y})|\tilde{X}^n = x^n, \tilde{U}^n = u^n)$$
$$= 2^{-\frac{n\epsilon}{2}} \Pr(\tilde{Y} \in T^n_{\tilde{Y}|\tilde{X}\tilde{U},\gamma_n}(x^n u^n)|\tilde{X}^n = x^n, \tilde{U}^n = u^n)$$
$$\leq 2^{-\frac{n\epsilon}{2}}, \tag{165}$$

where the summation is taken over all $\hat{V}$ such that $\hat{V}$ is $(x^n u^n, \tilde{Y}|\tilde{X}\tilde{U}, \gamma_n)$-essential. In view of (121), (160) and (165), it follows that

$$\Pr(\tilde{Y}^n \in T^n_{\tilde{Y}|\tilde{X}\tilde{U},\gamma_n}(x^n u^n) \cap G(b^n)|\tilde{X}^n = x^n, \tilde{U}^n = u^n)$$
$$\geq 1 - \frac{|\mathcal{U}_m||\mathcal{X}||\mathcal{Y}|}{4n\gamma_n^2} - 2^{-\frac{n\epsilon}{2}}$$
$$\geq 1 - \delta'_n, \tag{166}$$

where $\delta'_n \to 0$ as $n \to \infty$. Let

$$G = \cup_{x^n \in A} T^n_{\tilde{Y}|\tilde{X}\tilde{U},\gamma_n}(x^n u^n) \cap G(b^n).$$

(166) implies that for any $x^n \in A$,

$$\Pr(\tilde{Y}^n \in G|\tilde{X}^n = x^n, \tilde{U}^n = u^n) \geq 1 - \delta'_n,$$

which in turn implies

$$\Pr(\tilde{Y}^n \in G) \geq 1 - \delta'_n. \tag{167}$$

Note that (167) is in paralell with (123). A similar argument to the derivation of (129) can be used to show that

$$\Pr(Y^n \in G) \geq \exp\{-n(I(\tilde{U} \wedge \tilde{Y}) + D(P_{\tilde{Y}}||P_Y) + \frac{\log^2 m}{m} + \epsilon_n)\}, \tag{168}$$

where $\epsilon_n \to 0$ as n goes to infinity. Since $G \subset G(b^n)$, (168) implies

$$\Pr(Y^n \in G(b^n)) \geq \exp\{-n(I(\tilde{U} \wedge \tilde{Y}) + D(P_{\tilde{Y}}||P_Y) + \frac{\log^2 m}{m} + \epsilon_n)\}. \tag{169}$$

Note that (169) holds for any random variable $\tilde{Y}$ taking values on $\mathcal{Y}$ such that (159) is satisfied. This, together with the definition of $\mathcal{E}(P_{\tilde{X}\tilde{U}}, \alpha - 4\epsilon, \beta(d), d - \epsilon)$, implies that

$$\Pr(Y^n \in G(b^n)) \geq \exp\{-n(\mathcal{E}(P_{\tilde{X}\tilde{U}}, \alpha - 4\epsilon, \beta(d), d - \epsilon) + \frac{\log^2 m}{m} + \epsilon_n)\}. \tag{170}$$

On the other hand, since $b^n \in B'_n$,

$$\Pr(Y^n \in G(b^n)) \leq 2^{-n(\beta - \epsilon - \frac{1}{n}\log\frac{2}{\epsilon})}$$

which, combined with (170), yields

$$\mathcal{E}(P_{\tilde{X}\tilde{U}}, \alpha - 4\epsilon, \beta(d), d - \epsilon) \geq \beta - \epsilon - \frac{\log^2 m}{m} - \epsilon'_n \tag{171}$$

where ϵ'_n goes to zero as n goes to infinity. A similar argument to the derivation of (133) can be used to show that

$$\Pr(X^n \in \mathcal{S}(b^n) \cap F_n \cap T^n_{X,\gamma_n}) \leq 2^{-n(I(\tilde{X}\wedge\tilde{U}) - \epsilon''_n)},$$

that is

$$-\frac{1}{n}\log\Pr(X^n \in \mathcal{S}(b^n) \cap F_n \cap T^n_{X,\gamma_n}) \geq I(\tilde{X} \wedge \tilde{U}) - \epsilon''_n, \tag{172}$$

where ϵ''_n goes to zero as n goes to infinity. In view of (171) and the definition of $R_{|\mathcal{U}_m|}(P_{\tilde{X}}, P_Y, \alpha, \gamma, \beta, d)$, (172) continues as follows

$$-\frac{1}{n}\log\Pr(X^n \in \mathcal{S}(b^n) \cap F_n \cap T^n_{X,\gamma_n}) \geq R_{|\mathcal{U}_m|}(P_{\tilde{X}}, P_Y, \alpha - 4\epsilon, \beta(d), \beta - \epsilon - \frac{\log^2 m}{m} - \epsilon'_n, d - \epsilon) - \epsilon''_n. \tag{173}$$

Using Fact 3 in Subsetion A, we get

$$-\frac{1}{n}\log\Pr(X^n \in \mathcal{S}(b^n)\cap F_n \cap T^n_{X,\gamma_n}) \geq R_{|\mathcal{U}_m|}(P_X, P_Y, \alpha - 4\epsilon, \beta(d), \beta - \epsilon - \frac{\log^2 m}{m}, d-\epsilon) - \bar{\epsilon}_n \ , \tag{174}$$

where $\bar{\epsilon}_n$ goes to zero as n goes to infinity. Note that (174) holds for any $b^n \in B'_n$. In paralell with (137), we now have

$$R+\epsilon \geq r_n(\mathcal{C}_n) \geq \frac{1}{n}H(f_n(X^n))$$

$$\geq \sum_{b^n\in B'_n} -\frac{1}{n}\Pr(X^n \in \mathcal{S}(b^n)\cap F_n\cap T^n_{X,\gamma_n})\log\Pr(X^n\in\mathcal{S}(b^n)\cap F_n\cap T^n_{X,\gamma_n}) - \frac{2}{n}$$

$$\geq (1-2\epsilon-\frac{|\mathcal{X}|}{4n\gamma_n^2})(R_{|\mathcal{U}_m|}(P_X, P_Y, \alpha-4\epsilon, \beta(d), \beta-\epsilon-\frac{\log^2 m}{m}, d-\epsilon) - \bar{\epsilon}_n) - \frac{2}{n}. \tag{175}$$

In view of Fact 3 in Subsection A once again, letting $n\to\infty$ and then letting $\epsilon \to 0$ yields

$$R \geq R_{|\mathcal{U}_m|}(P_X, P_Y, \alpha, \beta(d), \beta - \frac{\log^2 m}{m}, d),$$

which implies

$$R^*_{XY}(\alpha,\beta,d) \geq R_{|\mathcal{U}_m|}(P_X, P_Y, \alpha, \beta(d), \beta - \frac{\log^2 m}{m}, d). \tag{176}$$

Letting $m\to\infty$ in(176) yields

$$R^*_{XY}(\alpha,\beta,d) \geq \bar{R}(P_X, P_Y, \alpha, \beta(d), \beta, d),$$

which completes the proof of the converse part and hence the proof of Theorem 3.

V Proofs of Thereoms 5 and 6

In this section, X and Y may be correlated. As in Subsection C, let $W = (W(y|x))_{x\in\mathcal{X},y\in\mathcal{Y}}$ denote the transition probability matrix from X to Y.

Proof of Theorem 5 We begin with proving

$$R^*_{XY}(+\infty, 0, d) \leq R(P_{XY}, 0, d). \tag{177}$$

To prove (177), it suffices to prove that for any $R > R(P_{XY}, 0, d)$, there exists a $\delta_0 > 0$ such that $(R, +\infty, \delta_0)$ is achievable. To this end, we fix below $R > R(P_{XY}, 0, d)$. By the definition of $R(P_{XY}, 0, d)$, there exists a random variable U taking values on some finite set $\mathcal{U}$ such that

(i) $U \to X \to Y$ form a Markov chain;

(ii) $I(X \wedge U) < R$ and $\mathbf{E}\bar{\rho}(P_{X|U}(\cdot|U), P_{Y|U}(\cdot|U)) > d$.

Without loss of generality, in what follows, we shall assume $P_U(u) > 0$ for any $u \in \mathcal{U}$. Let δ be a positive real to be specified later. As in the proof of the direct part of Theorem 2, corresponding to the random pair (X, U), there exists for sufficiently large n a system $\{(u^n(i), \mathcal{S}_i) | 1 \leq i \leq M\}$ which has the properties (i) to (iii). Let $\mathcal{C}_n = (f_n, B_n, g_n)$ be the n-th order IDS code which is based on the system we just defined and constructed as in the proof of the direct part of Theorem 2. From the proof of the direct part of Theorem 2, the probability of misidentification of $\mathcal{C}_n$ is zero and the average rate in bits per symbol of $\mathcal{C}_n$ is upper bounded by

$$r_n(\mathcal{C}_n) \leq I(X \wedge U) + (1 + \log |\mathcal{X}|)\delta + \frac{2}{n}. \tag{178}$$

Furthermore, the probability of false identification of $\mathcal{C}_n$ is now upper bounded by

$$\begin{aligned} P_{e2}(\mathcal{C}_n) &\leq \frac{1}{\Pr(\rho_n(X^n, Y^n) > d)} \sum_{i=1}^{M} \Pr((X^n, Y^n) \in \mathcal{S}_i \times \mathcal{S}_i^d) \\ &\leq 2 \sum_{i=1}^{M} \sum_{y^n \in \mathcal{S}_i^d} \sum_{x^n \in \mathcal{S}_i} \Pr\{X^n = x^n\} \Pr\{Y^n = y^n | X^n = x^n\} \end{aligned} \tag{179}$$

for sufficiently large n. Since $u^n(i) \in T^n_{U,\gamma_n}$ and $\mathcal{S}_i \subset T^n_{X|U,\gamma_n}(u^n(i))$ for each $1 \leq i \leq M$, it follows that for sufficiently large n and for any $x^n \in \mathcal{S}_i$,

$$\Pr(X^n = x^n) = 2^{-n(I(X \wedge U) + o(1))} \Pr(X^n = x^n | U^n = u^n(i))$$

which, together with (179) and the fact that $U \to X \to Y$ form a Markov chain, implies

$$\begin{aligned} P_{e2}(\mathcal{C}_n) &\leq 2 2^{-n(I(X \wedge U) + o(1))} \sum_{i=1}^{M} \sum_{y^n \in \mathcal{S}_i^d} \Pr(Y^n = y^n | U^n = u^n(i)) \\ &= 2^{-n(I(X \wedge U) + o(1))} \sum_{i=1}^{M} \Pr(Y^n \in \mathcal{S}_i^d | U^n = u^n(i)) \,, \end{aligned} \tag{180}$$

where (U^n, X^n, Y^n) is n independent drawings of (U, X, Y). Let

$$d_0 = \mathbf{E}\bar{\rho}(P_{X|U}(\cdot|U), P_{Y|U}(\cdot|U)).$$

For convenience, we think of $\mathbf{E}\bar{\rho}(P_{X|U}(\cdot|U), P_{Y|U}(\cdot|U))$ as a function of $(P_U, P_{X|U}, P_{Y|U})$ which is denoted by $F(P_U, P_{X|U}, P_{Y|U})$. It is not hard to prove that this function is continuous. Since $d_0 > d$, there exists a $\sigma > 0$ such that for any $P \in \mathcal{P}(\mathcal{U})$ and any stochastic matrix $V = V(x|u))_{u\in\mathcal{U},x\in\mathcal{X}}$,

$$||P - P_u|| \leq \sigma, ||V - P_{X|U}|| \leq \sigma \Longrightarrow F(P, V, P_{Y|U}) > \frac{d_0 + d}{2}, \tag{181}$$

where

$$||V - P_{X|U}|| = \sum_{u\in\mathcal{U}} ||V(\cdot|u) - P_{X|U}(\cdot|u)||.$$

Particularly, for sufficiently large n and for any $x^n \in \mathcal{S}_i$,

$$F(P_{u^n(i)}, P_{x^n|u^n(i)}, P_{Y|U}) > \frac{d_0 + d}{2}, \tag{182}$$

where $P_{x^n|u^n(i)} \in \mathcal{V}_n(P_{u^n(i)}, \mathcal{U} \times \mathcal{X})$ is the stochastic matrix so that x^n is $P_{x^n|u^n(i)}$-generated by $u^n(i)$, since $(u^n(i), x^n)$ is $(UX, 2\gamma_n)$-typical. To continue (180), let us note that if $\mathcal{S}_i^d \cap T_V^n(u^n(i), \mathcal{Y})$ is not empty, where $V \in \mathcal{V}_n(P_{u^n(i)}, \mathcal{U} \times \mathcal{Y})$, then there exists $x^n \in \mathcal{S}_i$ and $Q \in \mathcal{P}_n(\mathcal{U} \times \mathcal{X} \times \mathcal{Y})$ such that

(i) the marginal of Q on $\mathcal{U} \times \mathcal{X}$ is $P_{u^n(i)x^n}$;

(ii) the marginal of Q on $\mathcal{U} \times \mathcal{Y}$ is given by

$$P_{u^n(i)}(u)V(y|u), \quad u \in \mathcal{U} \quad and \quad y \in \mathcal{Y};$$

(iii) under the distribution Q, $\mathbf{E}\rho(X_0, Y_0) \leq d$.

This implies

$$F(P_{u^n(i)}, P_{x^n|u^n(i)}, V) \leq d. \tag{183}$$

In view of (181) and (182), (183) implies

$$\sum_{u\in\mathcal{U}} P_{u^n(i)}(u) D(V(\cdot|u)||P_{Y|U}(\cdot|u)) > 3\delta_0, \tag{184}$$

where $\delta_0 > 0$ is a constant independent of i, $u^n(i)$ and V. Therefore, if $\mathcal{S}_i^d \cap T_V^n(u^n(i), \mathcal{Y})$ is not empty, then

$$\Pr\{Y^n \in \mathcal{S}_i^d \cap T_V^n(u^n(i), \mathcal{Y}) | U^n = u^n(i)\} \leq \Pr\{Y^n \in T_V^n(u^n(i), \mathcal{Y}) | U^n = u^n(i)\}$$
$$\leq 2^{-3n\delta_0} \tag{185}$$

which in turn implies

$$\Pr\{Y^n \in \mathcal{S}_i^d | U^n = u^n(i)\} \leq |\mathcal{V}_n(P_{u^n(i)}, \mathcal{U} \times \mathcal{Y})| 2^{-3n\delta_0}$$
$$\leq 2^{-2n\delta_0} \tag{186}$$

for sufficiently large n. Substituting (186) into (180) yields,

$$P_{e2}(\mathcal{C}_n) \leq 2^{-n(2\delta_0 - \delta + o(1))}.$$

Selecting $\delta < \delta_0$ so small that the right hand side of (178) is less than R. Accordingly,

$$P_{e2}(\mathcal{C}_n) \leq 2^{-n\delta_0} \tag{187}$$

for sufficiently large n. This shows that $(R, +\infty, \delta_0)$ is achievable and hence completes the proof of (177).

We next turn to proving

$$R_{XY}^*(+\infty, 0, d) \geq \tilde{R}(P_{XY}, 0, d). \tag{188}$$

By the definition of $R_{XY}^*(+\infty, 0, d)$, it suffices to prove that for any achievable triple $(R, +\infty, \beta)$,

$$R \geq \tilde{R}(P_{XY}, 0, d). \tag{189}$$

To this end, let us fix below an achievable triple $(R, +\infty, \beta)$. By definition, there exists for any $\epsilon > 0$ a sequence $\{\mathcal{C}_n\}$ of IDS codes, where $\mathcal{C}_n = (f_n, B_n, g_n)$ is an n-th order IDS code, such that for sufficiently large n,

$$r_n(\mathcal{C}_n) \leq R + \epsilon, P_{e1}(\mathcal{C}_n) = 0, \quad and \quad P_{e2}(\mathcal{C}_n) \leq 2^{-n(\beta - \epsilon)}. \tag{190}$$

As what we did before, for each $b^n \in B_n$, let

$$\mathcal{S}(b^n) = \{x^n \in \mathcal{X}^n : f_n(x^n) = b^n\}.$$

Let $\hat{\mathcal{S}}^d(b^n)$ denote the set of all sequences $y^n \in \mathcal{Y}^n$ such that

$$\Pr(X^n \in \mathcal{S}(b^n) \cap B_d(y^n) | Y^n = y^n) > 0. \tag{191}$$

where

$$B_d(y^n) = \{x^n \in \mathcal{X}^n : \rho_n(x^n, y^n) \le d\}.$$

Clearly, $P_{e1}(\mathcal{C}_n) = 0$ implies

$$\hat{\mathcal{S}}^d(b^n) \subset \{y^n \in \mathcal{Y}^n : g_n(y^n, b^n) = 1\}.$$

From (190), therefore, it is not hard to see that

$$\sum_{b^n \in B_n} \Pr\{(X^n, Y^n) \in \mathcal{S}(b^n) \times \hat{\mathcal{S}}^d(b^n)\} \le 2^{-n(\beta-\epsilon)} + \Pr\{\rho_n(X^n, Y^n) \le d\}$$

which in turn implies

$$\sum_{b^n \in B_n} \Pr\{X^n \in \mathcal{S}(b^n)\} \sum_{x^n \in \mathcal{S}(b^n)} \frac{\Pr\{X^n = x^n\}}{\Pr\{X^n \in \mathcal{S}(b^n)\}} \Pr\{Y^n \in \hat{\mathcal{S}}^d(b^n) | X^n = x^n\} \to 0 \qquad (192)$$

as n goes to infinity. Let B'_n consist of all $b^n \in B_n$ such that

$$\sum_{x^n \in \mathcal{S}(b^n)} \frac{\Pr(X^n = x^n)}{\Pr(X^n \in \mathcal{S}(b^n))} \Pr(Y^n \in \hat{\mathcal{S}}^d(b^n) | X^n = x^n) < \frac{\epsilon}{d}. \qquad (193)$$

From (192) and the Markov inequation, for sufficiently large,

$$\sum_{b^n \in B'_n} \Pr(X^n \in \mathcal{S}(b^n)) > 1 - \epsilon.$$

Fix $b^n \in B'_n$ and consider $\mathcal{S}(b^n) \cap T^n_{X,\gamma_n}$. It is easy to see that there exists $A \subset \mathcal{S}(b^n) \cap T^n_{X,\gamma_n}$ such that $A \subset T^n_P(\mathcal{X})$ for some (X, γ_n)-essential P and

$$\frac{1}{n} \log \frac{|\mathcal{S}(b^n) \cap T^n_{X,\gamma_n}|}{|A|} \le |\mathcal{X}| \frac{\log n + 1}{n}. \qquad (194)$$

From (193),

$$\sum_{x^n \in A} \frac{1}{|A|} \Pr(Y^n \in \hat{A}^d | X^n = x^n) \le \frac{\epsilon}{d} \qquad (195)$$

where $\hat{A}^d$ is defined in the same way as $\hat{\mathcal{S}}^d(b^n)$was. Focusing on A, we define a random vector $\tilde{X}^n = (\tilde{X}_1, \cdots, \tilde{X}_n)$ taking values uniformly on A. Let $\tilde{Y}^n = (\tilde{Y}_1, \cdots, \tilde{Y}_n)$ be the output of the memoryless channel W resulting from passing $\tilde{X}^n$ through W. It is easy to verify that

$$\frac{1}{n} \log |A| = \frac{1}{n} H(\tilde{X}^n) \;\; = \;\; \frac{1}{n} H(\tilde{X}^n | \tilde{Y}^n) + \frac{1}{n} I(\tilde{X}^n; \tilde{Y}^n)$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n} H(\tilde{X}_i|\tilde{X}^{i-1},\tilde{Y}^n) + \frac{1}{n} I(\tilde{X}^n;\tilde{Y}^n) \\
&\leq \frac{1}{n}\sum_{i=1}^{n} H(\tilde{X}_i|\tilde{X}^{i-1},\tilde{Y}_i,\tilde{Y}_{i+1}^n) + \frac{1}{n} H(\tilde{Y}^n) - \frac{1}{n}\sum_{i=1}^{n} H(\tilde{Y}_i|\tilde{X}_i).
\end{aligned}
\tag{196}
$$

Let I be a random variable taking values uniformly on $\{1,\cdots,n\}$ and independent of $\tilde{X}^n$ and $\tilde{Y}^n$. Let

$$
\tilde{X} = \tilde{X}_I, \tilde{Y} = \tilde{Y}_I \text{ and } U = (\tilde{X}^{I-1}, \tilde{Y}_{I+1}^n, I). \tag{197}
$$

Then (196) continues as follows

$$
\begin{aligned}
\frac{1}{n}\log|A| &\leq H(\tilde{X}|\tilde{Y},U) + I(\tilde{X}\wedge\tilde{Y}) \\
&= H(\tilde{X}|U) + I(U\wedge\tilde{Y}),
\end{aligned}
\tag{198}
$$

where the last step follows from the fact that $U \to \tilde{X} \to \tilde{Y}$ form a Markov chain. From (194) and (198), we now have

$$
\begin{aligned}
\Pr(X^n \in \mathcal{S}(b^n)\cap T^n_{X,\gamma_n}) &\leq |\mathcal{S}(b^n)\cap T^n_{X,\gamma_n}|2^{-n(H(\tilde{X})+o(1))} \\
&\leq \exp\{-n[I(\tilde{X}\wedge U) - I(\tilde{Y}\wedge U) - o(1)]\}.
\end{aligned}
\tag{199}
$$

Next we show that $U,\tilde{X},\tilde{Y}$ satisfy

$$
\mathbf{E}\bar{\rho}_e(P_{\tilde{X}|U}(\cdot|U)) > d-\epsilon. \tag{200}
$$

To this end, Let $\mathcal{U}$ be the finite set on which U takes values, that is

$$
\mathcal{U} = \{(x^{i-1}, y_{i+1}^n, i) : x^{i-1}\in\mathcal{X}^{i-1}, y_{i+1}^n \in \mathcal{Y}^{n-i}, 1\leq i\leq n\}.
$$

For each $u\in\mathcal{U}$, let $\hat{W}_u = (\hat{W}_u(y|x))$ be a stochastic matrix such that

(i) $P_{\tilde{X}|U}(\cdot|u)\hat{W}_u = P_{\tilde{X}|U}(\cdot|u)W$;

(ii) $\hat{W}_u$ is absolutely continuous with respect to W;

(iii)

$$
\bar{\rho}_e(P_{\tilde{X}|U}(\cdot|u)) = \sum_{x\in\mathcal{X},y\in\mathcal{Y}} P_{\tilde{X}|U}(x|u)\hat{W}_u(y|x)\rho(x,y).
$$

Therefore,

$$\mathbf{E}\bar{\rho}_e(P_{\tilde{X}|U}(\cdot|U)) = \sum_{u\in\mathcal{U},x\in\mathcal{X},y\in\mathcal{Y}} P_U(u)P_{\tilde{X}|U}(x|u)\hat{W}_u(y|x)\rho(x,y).$$

We now write $\hat{W}_u(y|x)$ as $\hat{W}_i(y|x^{i-1}, y_{i+1}^n, x)$ whenever $u = (x^{i-1}, y_{i+1}^n, i)$. Think of $\hat{W}_i(\cdot|\cdot)$ as a channel $\mathcal{X}^{i-1} \times \mathcal{Y}^{n-i} \times \mathcal{X} \to \mathcal{Y}$ and construct a random vector $\hat{Y}^n = (\hat{Y}_1, \cdots, \hat{Y}_n)$ as follows:

Step 1 . For $i = n$, $\hat{W}_n$ is from $\mathcal{X}^{n-1} \times \mathcal{X}$ to $\mathcal{Y}$. Pass $\tilde{X}^n$ (viewed as $(\tilde{X}^{n-1}, \tilde{X}_n)$) through $\hat{W}_n$ and denote the output by $\hat{Y}_n$;

Step 2 . Pass $(\tilde{X}^{n-2}, \hat{Y}_n, \tilde{X}_{n-1})$ through $\hat{W}_{n-1}$, and denote the output by $\hat{Y}_{n-1}$;

Step i . So far, $\hat{Y}_{n-j}$ for $j = 0, \cdots, i-2$ have been constructed. Pass $(\tilde{X}_{n-i}, \hat{Y}_{n-i+2}^n, \tilde{X}_{n-i+1})$ through channel $\hat{W}_{n-i+1}$ and denote the output by $\hat{Y}_{n-i+1}$.

Continue this procedure until

Step n . Pass $(\hat{Y}_2^n, \tilde{X}_1)$ through the channel $\hat{W}_1$ and denote the output by $\hat{Y}_1$.

Since $P_{\tilde{X}|U}(\cdot|u)\hat{W}_u = P_{\tilde{X}|U}(\cdot|u)W$, from the above construction, we can see that for any $i : 1 \leq i \leq n$, $(\tilde{X}^{n-i}, \hat{Y}_{n-i+1}^n)$ has the same distribution as that of $(\tilde{X}^{n-i}, \tilde{Y}_{n-i+1}^n)$. From this, we obtain

$$\begin{aligned}\mathbf{E}\bar{\rho}_e(P_{\tilde{X}|U}(\cdot|U)) &= \mathbf{E}\rho_n(\tilde{X}^n, \hat{Y}^n) \\ &= \mathbf{E}[\mathbf{E}(\rho_n(\tilde{X}^n, \hat{Y}^n)|\hat{Y}^n)], \end{aligned} \tag{201}$$

where $\mathbf{E}(\cdot|\hat{Y}^n)$ denote the conditional expectation with respect to $\hat{Y}^n$. Since $\hat{W}_u$ is absolutely continuous with respect to W for any $u \in \mathcal{U}$, it follows from the construction of $\hat{Y}^n$ that $P_{\tilde{X}^n\hat{Y}^n}$ is also absolutely continuous with respect to $P_{\tilde{X}^n\tilde{Y}^n}$. Therefore, for any $y^n \notin \hat{A}^d$, if $\Pr(\hat{Y}^n = y^n) > 0$, or equivalently, $\Pr(\tilde{Y}^n = y^n) > 0$, then from the definition of $\hat{A}^d$, we have

$$\Pr(\rho_n(\tilde{X}^n, y^n) \leq d|\hat{Y}^n = y^n) = 0$$

which implies

$$\mathbf{E}(\rho_n(\tilde{X}^n, \hat{Y}^n)|\hat{Y}^n = y^n) \geq d. \tag{202}$$

Note that (195) can be rewritten as

$$\Pr(\tilde{Y}^n \in \hat{A}^d) < \frac{\epsilon}{d}.$$

This, combined with (201) and (202), yields

$$\mathbf{E}\bar{\rho}_e(P_{\tilde{X}|U}(\cdot|U)) = \mathbf{E}[\mathbf{E}(\rho_n(\tilde{X}^n, \hat{Y}^n)|\hat{Y}^n)]$$

$$\geq d \Pr(\hat{Y}^n \notin \hat{A}^d)$$
$$= d \Pr(\tilde{Y}^n \notin \hat{A}^d)$$
$$> d - \epsilon.$$

Finally, let us go back to (199). In view of the definition of $\tilde{R}(P_{\tilde{X}\tilde{Y}}, 0, d)$, we have

$$\begin{aligned} \Pr(X^n \in \mathcal{S}(b^n) \cap T_{X,\gamma_n}) &\leq \exp\{-n\tilde{R}(P_{\tilde{X}\tilde{Y}}, 0, d-\epsilon) + o(n)\} \\ &\leq \exp\{-n\tilde{R}(P_{XY}, 0, d-\epsilon) + o(n)\}\ , \end{aligned} \tag{203}$$

where the last inequality is due to the fact that $P_{\tilde{X}}$ is (X, γ_n)-essential. Note that (203) holds for any $b^n \in B'_n$. In view of (190), we have

$$\begin{aligned} R + \epsilon &\geq r_n(\mathcal{C}_n) \\ &\geq \sum_{b^n \in B'_n} -\frac{1}{n} \Pr(X^n \in \mathcal{S}(b^n) \cap T^n_{X,\gamma}) \log \Pr(X^n \in \mathcal{S}(b^n) \cap T^n_{X,\gamma}) - \frac{1}{n} \\ &\geq (1 - \epsilon - \frac{|\mathcal{X}|}{4n\gamma_n^2})(\tilde{R}(P_{XY}, 0, d-\epsilon) - o(1)) - \frac{1}{n}. \end{aligned}$$

In view of Lemma 3, letting $n \to \infty$ and then letting $\epsilon \to 0$ yield

$$R \geq \tilde{R}(P_{XY}, 0, d).$$

This completes the proof of (189) and hence the proof of Theorem 5.

Remark 4. At this point, we point out the reason why the method used in Section IV to prove Theorems 2 and 3 can not be generalized to the general case in which X and Y may be correlated. The main difficulty lies in the fact that even in the simplest case of $\alpha = +\infty$ and $\beta = 0$, the auxilliary random variable U introduced in the proof of the lower bound of Theorem 5 involves both sets $\mathcal{X}$ and $\mathcal{Y}$.

Proof of Theorem 6.

Clearly, we need to prove only

$$R^*_{(XZ)(YZ)}(+\infty, 0, d) \leq \tilde{R}(P_{(XZ)(YZ)}, 0, d). \tag{204}$$

By using the formula (63) for $\tilde{R}(P_{(XZ)(YZ)}, 0, d)$, an argument similar to the derivation of (177) can be used to show (204).

VI Open Problems

The following problems remain open:

Problem 1 . When X and Y are independent, Theorem 3 gives $R^*_{XY}(\alpha, \beta, d)$ for $0 \le \beta < \beta(d)$. What happens if $\beta \ge \beta(d)$?

Problem 2 . What is the counterpart of Theorem 3 in the general case in which X and Y may be correlated?

Problem 3 . Problem 2 may be too difficult to solve. An easier problem is this:what is $R^*_{XY}(+\infty, 0, d)$ in the general case?

Problem 4 . In this paper, we considered the case when $d < \mathbf{E}\rho(X,Y)$. What happens if $d > \mathbf{E}\rho(X,Y)$? In the binary symmetric case, of course, the problem associated with $d > \mathbf{E}\rho(X,Y)$ is equivalent to that associated with $d < \mathbf{E}\rho(X,Y)$. In general, however, this is not true.

VII Appendix

<u>A. Proof of Theorem 1</u>.

To prove $R^o_{XY} = 0$, we construct, for sufficiently large n, an n-th order ID source code $\mathcal{C}_n = (f_n, B_n, g_n)$ as follows. For each $x^n \in \mathcal{X}^n$, the encoder sends x^k completely to the decoder. This needs $n^{-1}\lceil k \log |\mathcal{X}| \rceil$ bits per source symbol. Observing $y^n \in \mathcal{Y}^n$ and receiving x^k, the decoder outputs 1 if $\rho_k(x^k, y^k) \le d + \delta$ and 0 otherwise, where $\delta > 0$ is selected so that $d + \delta < \mathbf{E}\rho(X,Y)$. The probability of false identification is given by

$$P_{e2}(\mathcal{C}_n) = \Pr(\rho_k(X^k, Y^k) \le d + \delta | \rho_n(X^n, Y^n) > d).$$

Since $d < \mathbf{E}\rho(X,Y)$, it is easy to see that for sufficiently large n

$$P_{e2}(\mathcal{C}_n) \le 2 \Pr(\rho_k(X^k, Y^k) \le d + \delta).$$

Let $\mathcal{P}^d$ denote the set of all $Q \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ such that

$$\mathbf{E}_Q \rho(X_0, Y_0) \le d$$

Clearly, $\mathcal{P}^d$ is a convex and closed set. Let Q^* be the unique element of $\mathcal{P}^d$ such that

$$D(Q^*||P_{XY}) = \min_{Q \in \mathcal{P}^d} D(Q||P_{XY}).$$

For any $\epsilon > 0$, select k so large that

$$P_{e2}(\mathcal{C}_n) \leq \epsilon$$

and

$$\Pr(\rho_k(\hat{X}^k, \hat{Y}^k) > d + \delta) < \epsilon.,$$

where $(\hat{X}^k, \hat{Y}^k)$ is the sequence of k-times independent drawings of a pair of random variables $(\hat{X}, \hat{Y})$ taking values on $\mathcal{X} \times \mathcal{Y}$ with joint distribution of $P_{\hat{X}\hat{Y}} = Q^*$. Fix such a k. All remaining is to prove that for sufficiently large n, the probability of misidentification $P_{e1}(\mathcal{C}_n)$ will be less than ϵ. To see this is true, note that

$$P_{e1}(\mathcal{C}_n) = \Pr(\rho_k(X^k, Y^k) > d + \delta | \rho_n(X^n, Y^n) \leq d).$$

By virtue of the conditional limit theorem [6][10], it is not hard to prove that

$$\lim_{n \to \infty} P_{e1}(\mathcal{C}_n) = \Pr(\rho_k(\hat{X}^k, \hat{Y}^k) > d + \delta) < \epsilon.$$

This completes the proof of Theorem 1.

B. Proof of Lemma 3.
For $i = 1, 2$, let (U_i, X_i, Y_i) be a random vector such that

(i) $P_{X_i Y_i} = P_{XY}$;

(ii) $U_i \to X_i \to Y_i$ forms a Markov chain;

(iii) $\mathbf{E}\bar{\rho}_e(P_{X_i|U_i}(\cdot|U_i)) > d_i$.

Let I be a random variable taking values in $\{1, 2\}$ with $\Pr(I = 1) = \lambda$. I is assumed to be indepedent of (U_i, X_i, Y_i) for $i = 1, 2$. Define

$$\tilde{X} = X_I, \tilde{Y} = Y_I, \text{ and } U = (U_I, I).$$

Clearly, $P_{\tilde{X}\tilde{Y}} = P_{XY}$ and $U \to \tilde{X} \to \tilde{Y}$ forms a Markov chain. Furthermore, it is not hard to see that

$$\begin{aligned}\mathbf{E}\bar{\rho}_e(P_{\tilde{X}|U}(\cdot|U)) &= \lambda \mathbf{E}\bar{\rho}_e(P_{X_1|U_1}(\cdot|U_1)) + (1-\lambda)\mathbf{E}\bar{\rho}_e(P_{X_2|U_2}(\cdot|U_2)) \\ &> \lambda d_1 + (1-\lambda) d_2\end{aligned}$$

and

$$I(\tilde{X} \wedge U) - I(\tilde{Y} \wedge U) = \lambda(I(X_1 \wedge U_1) - I(Y_1 \wedge U_1)) + (1-\lambda)(I(X_2 \wedge U_2) - I(Y_2 \wedge U_2)).$$

From this and the definition of $\tilde{R}(P_{XY}, 0, d)$, it follows that $\tilde{R}(P_{XY}, 0, d)$ as a function of d is convex.

To prove the second part of Lemma 3, first note that $\bar{\rho}_e(P)$ is convex as a function of P over $\mathcal{P}(\mathcal{X})$. Since $\mathcal{P}(\mathcal{X})$ is a convex polytope, it follows from [11] that $\bar{\rho}_e(P)$ is upper semicontinuous on $\mathcal{P}(\mathcal{X})$. On the other hand, from the definition of $\bar{\rho}_e(P)$, it is easy to prove that $\bar{\rho}_e(P)$ is lower semicontinuous on $\mathcal{P}(\mathcal{X})$. Therefore, $\bar{\rho}_e(P)$ is continuous on $\mathcal{P}(\mathcal{X})$. Applying the support lemma to the definition of $\tilde{R}(P_{XY}, 0, d)$ yields immediately the second result of Lemma 3.

References

[1] T. Berger, *Rate Distortion Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1971.

[2] R. Ahlswede and J. Körner, Source coding with side information and a converse for degrades broadcast channels, IEEE Trans. Inf. Th., Vol. IT-21,629-637, 1975

[3] R. Ahlswede, P. Gács, and J. Körner, Bounds on conditional probabilities with applications in multiuser communication, Z. Wahrscheinlichkeitstheorie und verw. Geb. 34, 157-177, 1976.

[4] R. Ahlswede and I. Csiszar, To get a bit of information may be as hard as to get full information, IEEE Trans. Inf. Theory, IT-27, 389-408, 1981.

[5] I. Csiszar and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, Inc., N.Y., 1981.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[7] L. H. Harper, "Optimal numbering and isoperimetric problems on graphs," *Jounal. Comb. Th.*, no.1, pp. 385–394, 1966.

[8] R. M. Gray, D. L. Neuhoff, and P. C. Shields, " A generalization of Ornstein's $\bar{d}$-distance with application to information theory," *Ann. Probab.*, Vol. 3, pp. 315–328, 1975.

[9] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes.* Holden-Day, San Francisco, 1964.

[10] I. Csiszar, "Sanov property, generalized I-projection and conditional limit theorem," *Ann. Probab.*, no. 12, pp.768–793, 1984.

[11] B. Grunbaum, *Convex Polytopes.* New York: Intersciences, 1967.