Asymptotic Expansions of Smooth Rényi Entropies and Their Applications

Yuta Sakai, Member, IEEE, and Vincent Y. F. Tan, Senior Member, IEEE

Abstract

This study considers the unconditional smooth Rényi entropy, the smooth conditional Rényi entropy proposed by Kuzuoka [IEEE Trans. Inf. Theory, vol. 66, no. 3, pp. 1674–1690, 2020], and a new quantity which we term the conditional smooth Rényi entropy. In particular, we examine asymptotic expansions of these entropies when the underlying source with its side-information are stationary and memoryless. Using these smooth Rényi entropies, we establish one-shot coding theorems of several information-theoretic problems: Campbell’s source coding problems, guessing problems, and task encoding problems, all allowing errors. In each problem, we consider two error formalisms: the average and maximum error criteria, where the averaging and maximization are taken with respect to the side-information of the source. Applying our asymptotic expansions to the one-shot coding theorems, we derive various asymptotic fundamental limits for these problems when their error probabilities are allowed to be non-vanishing. We show that, in non-degenerate settings, the first-order fundamental limits differ under the average and maximum error criteria. This is in contrast to a different but related setting considered by the present authors (for variable-length conditional source coding allowing errors) in which the first-order terms are identical but the second-order terms are different under these criteria.

Index Terms

Smooth Rényi entropy; second-order asymptotics; cumulant generating function of codeword lengths; guessing problems; encoding tasks

I. INTRODUCTION

Rényi’s information measures [1] admit various operational meanings in various information-theoretic problems, e.g., Campbell’s source coding problems [2] which concern with the cumulant generating function of codeword lengths of a prefix-free code (see also [3] for a fixed-to-variable length code without prefix-free constraints), guessing problems [4], [5], task encoding problems [6], and so on. By proposing a new set of axioms, Rényi in [1] generalized the Shannon entropy \( H \) to the Rényi entropy \( H_\alpha \) which is parameterized by \( \alpha \in (0, 1) \cup (1, \infty) \), a quantity which is known as the order.

Renner and Wolf in [7], [8] generalized \( H_\alpha \) by incorporating another parameter \( 0 \leq \varepsilon < 1 \) to form the smooth Rényi entropy \( H_\alpha^\varepsilon \); this parameter \( \varepsilon \) is known as the smoothness parameter (cf. [9]). Note that the two definitions stated in [7, Definition 1.1] and [8, Section 2.1] are slightly different, and the latter is amenable to being generalized to a conditional version of the smooth Rényi entropy defined in [8, Definition 1]. Basic properties of the smooth Rényi entropy \( H_\alpha^\varepsilon \) of [8] were investigated by Renner and Wolf [8] and Koga [10]. Recently, Kuzuoka [11] provided another definition of the smooth conditional Rényi entropy based on Arimoto’s conditional Rényi entropy [12]. He provided a general formula for the smooth conditional Rényi entropy. Moreover, using the smooth conditional Rényi entropy, he [11] established one-shot converse and achievability bounds on both Campbell’s source coding and guessing problems allowing errors in the presence of common side-information.

A. Main Contributions

For a stationary memoryless pair of random vectors \( X^n = (X_1, \ldots, X_n) \) and \( Y^n = (Y_1, \ldots, Y^n) \), this study investigates asymptotic expansions of three information measures: the unconditional version of the smooth Rényi entropy \( H_\alpha^\varepsilon(X^n) \) [8], Kuzuoka’s smooth conditional Rényi entropy \( H_\alpha^\varepsilon(X^n | Y^n) \) [11], and a new quantity that we propose which we term the conditional smooth Rényi entropy \( \tilde{H}_\alpha^\varepsilon(X^n | Y^n) \) which is different from both Renner and Wolf’s and Kuzuoka’s proposals [8], [11]. For \( H_\alpha^\varepsilon(X^n) \), we derive exact first, second, and third-order terms, i.e., coefficients in the \( n, \sqrt{n}, \) and \( \log n \) scales, respectively. More precisely, for fixed real numbers \( 0 < \alpha < 1 \) and \( 0 < \varepsilon < 1 \), we show that

\[
H_\alpha^\varepsilon(X^n) = n H(X) - \sqrt{n} V(X) \Phi^{-1}(\varepsilon) - \frac{1}{2(1 - \alpha)} \log n + O(1) \quad \text{(as } n \to \infty) \tag{1}
\]

provided that \( V(X) > 0 \) and \( T(X) < \infty \), where these notations are standard in the second-order asymptotics literature (cf. [13]), and will be explicitly defined later. This third-order asymptotic result is derived by refining Polyanskiy, Poor, and Verdú’s technical lemma [14, Lemma 47] whose proof employs the Berry–Esseen theorem. For the conditional versions, we show that the first-order terms of \( H_\alpha^\varepsilon(X^n | Y^n) \) and \( \tilde{H}_\alpha^\varepsilon(X^n | Y^n) \) differ in most non-degenerate cases, and we show that the remainder terms scale as \( +O(\sqrt{n}) \) due to Chebyshev’s inequality.

This work is supported by an NRF Fellowship (R-263-000-D02-281). Y. Sakai is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore, Email: eleyuta@nus.edu.sg. V. Y. F. Tan is with the Department of Electrical and Computer Engineering and the Department of Mathematics, National University of Singapore, Singapore, Email: vtan@nus.edu.sg.
To apply our asymptotic expansions of the smooth Rényi entropies to information-theoretic problems [2]–[6], [11], we establish one-shot coding theorems on the problems but allowing errors. While Kumar et al. [15] studied these problems [2]–[6] for sources \( X \) with finite alphabets \( X \) in the absence of side-information \( Y \), we consider sources with countably infinite alphabets \( X \) in the presence of side-information \( Y \). The extension from finite to countably infinite alphabets \( X \) was mentioned as a direction of future research by Kumar et al. [15, Section V]. Moreover, in each problem, we consider two error formalisms: average and maximum error criteria, where the averaging and maximization are taken with respect to the side-information \( Y \). Our one-shot coding theorems under average and maximum error criteria are formulated by using Kuzuoka’s proposal \( H_\alpha^e(X \mid Y) \) and our proposal \( H_\alpha^e(X \mid Y) \), respectively. We then characterize asymptotic expansions of fundamental limits of these problems. In the presence of the side-information \( Y \), we show that the first-order terms for the average and maximum error formalisms are different in most non-degenerate cases, in contrast to the main result of [16]. In the absence of the side-information \( Y \), we provide third-order asymptotic expansions of the fundamental limits by applying (1) to our one-shot coding theorems. Specifically, for Campbell’s source coding problem, our third-order asymptotic expansion is formulated by the right-hand side of (1). By comparing this asymptotic result to Strassen’s third-order asymptotic result for fixed-length source coding [17], we can quantify the improvement with which “variable-length compression” yields over “fixed-length compression.”

B. Other Related Works

1) Smooth Min- and Max-Entropies: Renner and Wolfe [7], [8] proposed the smooth Rényi entropy \( H_\alpha^e \) to provide operational interpretations of the smooth max-entropy \( H^e_{\max} \) and the smooth min-entropy \( H^e_{\min} \) using two information-theoretic problems, namely, fixed-length source coding and intrinsic randomness [18], [19]; these two entropies are special cases of \( H_\alpha^e \) by taking the limits as \( \alpha \to 0^+ \) and as \( \alpha \to \infty \), respectively. Interestingly, it is immediate from Strassen’s seminal result for fixed-length source coding of independent and identically distributed (i.i.d.) sources \( X^n \) [17] that

\[
H_0^e(X^n) = n H(X) - \sqrt{n V(X)} \Phi^{-1}(\varepsilon) - \frac{1}{2} \log n + O(1) \quad \text{(as } n \to \infty) \tag{2}
\]

for a fixed \( 0 < \varepsilon < 1 \), provided that \( V(X) > 0 \) and \( T(X) < \infty \). This result is consistent with our main result stated in (1).

In subsequent works, several applications of the smooth max- and min-entropies \( H_0^e \) and \( H_\infty^e \), respectively, were studied. Operational characterizations of \( H_0^e \) and \( H_\infty^e \), in various quantum information-theoretic problems were discussed by König, Renner, and Schaffner [9]. Using \( H_0^e \) and \( H_\infty^e \), Tomamichel, Colbeck, and Renner [20] formulated a quantum version of the asymptotic equipartition property (AEP) in the presence of quantum mechanical side-information. They [21] also discussed the duality between \( H_0^e \) and \( H_\infty^e \) in the context of quantum information theory. Uyematsu [22], [23] provided general formulas for fixed-length source coding problems and resolvability problems in terms of \( H_0^e \). Uyematsu and Kunimatsu provided a general formula for intrinsic randomness problems [24] in terms of \( H_\infty^e \). Finally, Saito and Matsushima [25] provided a general formula of the threshold of overflow probabilities for variable-length compressions in terms of \( H_0^e \).

2) Smooth Rényi Entropy of Order \( 0 < \alpha < 1 \): Operational characterizations of the smooth Rényi entropy \( H_\alpha^e \) defined in [8, Section 2.1] for \( 0 < \alpha < 1 \) were initiated by Kuzuoka [26]. He established one-shot bounds and a general formula for Campbell’s source coding problem [2] allowing errors. Sason and Verdú [27, Theorem 17] also provided a converse bound for Campbell’s source coding problem in the absence of prefix-free constraints. Yagi and Han [28] provided a general formula for the variable-length resolvability problem in terms of the smooth Rényi entropy of order one [7].

3) Unified One-Shot Coding Theorems: Recently, Kumar, Sunny, Thakre, and Kumar [15] proved unified one-shot coding theorems that can be specialized to various information-theoretic problems [2]–[6], [11] in the error-free regime. We will contrast our work to that of Kumar et al.; see Section III-A. Note that in [3, Lemma 2], Courtade and Verdú provided a unified lemma which can be specialized to Campbell’s source coding problem [2] in the absence of prefix-free constraints. Sason and Verdú [27, Lemma 7] established inequalities between the guessing moment and the moment generating function of codeword lengths for a variable-length source code without prefix-free constraints; see [16, Lemma 6] for a similar result to [27, Lemma 7] in the almost lossless regime.

4) Variable-Length Compression Allowing Errors: While this study examines the cumulant generating function of codeword lengths of a variable-length source code [2], [3] allowing errors as in Kuzuoka’s works [11], [26], the (ordinary) expectation of codeword lengths of a variable-length source code allowing errors has been investigated by several researchers [16], [29]–[31]. Specifically, the present authors [16] derived second-order asymptotic expansions of the fundamental limits of variable-length conditional source coding problems under both average and maximum error criteria. We then showed that the difference between the average and maximum error criteria is manifested in the second-order terms in these asymptotic expansions, and this difference can be quantified by the law of total variance for the information variance. However, in this work, the difference between the two error criteria is manifested in the first-order term.

C. Paper Organization

The rest of this paper is organized as follows: Section II introduces various definitions of the smooth Rényi entropies, and establishes various asymptotic expansions of these quantities for i.i.d. sources. Section III provides one-shot coding theorems
for various information-theoretic problems [2]–[6], [11] in the regime in which the error probabilities are allowed to be non-vanishing. In this section, we also derive asymptotic expansions of the fundamental limits of these problems by applying the results in Section II. Section IV concludes this study and discusses several directions for future works. Technical proofs are relegated to the appendices.

II. ASYMPTOTICS OF SMOOTH RÉNYI ENTROPIES

A. Unconditional Version of Smooth Rényi Entropy

Let $X$ be a countably infinite alphabet, and $X$ an $X$-valued random variable (r.v.). Denote by $P_X := \mathbb{P} \circ X^{-1}$ the probability law of $X$. Throughout this paper, denote by log the logarithm to the base 2. Given $\alpha \in (0, 1) \cup (1, \infty)$ and $0 \leq \varepsilon < 1$, Renner and Wolf [8, Section 2.1] defined the $\varepsilon$-smooth Rényi entropy of $X$ by

$$H^\varepsilon_\alpha(X) = H^\varepsilon_\alpha(P_X) := \frac{1}{1 - \alpha} \log \left( \inf_{Q \in \mathcal{B}^\varepsilon_\alpha(P_X)} \sum_{x \in X} Q(x)^\alpha \right),$$

where the infimum is taken over the collection $\mathcal{B}^\varepsilon_\alpha(P_X)$ of sub-probability distributions $Q$ on $X$ given as

$$\mathcal{B}^\varepsilon_\alpha(P_X) := \left\{ Q \left| \sum_{x \in X} Q(x) \geq 1 - \varepsilon \text{ and } 0 \leq Q(a) \leq P_X(a) \text{ for all } a \in X \right. \right\}. \quad (4)$$

Note that $H^\varepsilon_\alpha(P)$ coincides with the Rényi entropy $H_\alpha(X)$ [1] if $\varepsilon = 0$. In other words, one has

$$H^\varepsilon_\alpha(X)\big|_{\varepsilon = 0} = H_\alpha(X) := \frac{1}{1 - \alpha} \log \left( \sum_{x \in X} P_X(x)^\alpha \right). \quad (5)$$

Given an $X$-valued r.v. $X$ and a real number $0 \leq \varepsilon < 1$, we define $\mathcal{A}^\varepsilon_X \subset X$ as any proper subset that satisfies the following properties:

$$x_1 \in \mathcal{A}^\varepsilon_X \text{ and } x_2 \in X \setminus \mathcal{A}^\varepsilon_X \implies P_X(x_1) \geq P_X(x_2)$$

and

$$P_X(\mathcal{A}^\varepsilon_X) < 1 - \varepsilon \leq P_X(\mathcal{A}^\varepsilon_X) + \max_{x \in X \setminus \mathcal{A}^\varepsilon_X} P_X(x). \quad (7)$$

Note that $\mathcal{A}^\varepsilon_X$ is the empty set $\emptyset$ if and only if

$$\max_{x \in X} P_X(x) \geq 1 - \varepsilon. \quad (8)$$

The following lemma states a formula of $H^\varepsilon_\alpha(X)$ without the infimum operation used in the right-hand side of (3).

**Lemma 1** (Koga [10, Theorem 2]). For every $0 < \alpha < 1$ and $0 \leq \varepsilon < 1$, it holds that

$$H^\varepsilon_\alpha(X) = \frac{1}{1 - \alpha} \log \left( \sum_{x \in \mathcal{A}^\varepsilon_X} P_X(x)^\alpha + \left( 1 - \varepsilon - P_X(\mathcal{A}^\varepsilon_X) \right)^\alpha \right). \quad (9)$$

**Remark 1.** The collection of all $\mathcal{A}^\varepsilon_X \subset X$ satisfying (6) and (7) is, in general, not a singleton. However, this diversity of choices is irrelevant in this study, because the sub-probability distribution $Q_X^\varepsilon$ given as

$$Q_X^\varepsilon(\mathcal{B}) = P_X(\mathcal{B} \cap \mathcal{A}^\varepsilon_X) \quad (\text{for } \mathcal{B} \subset X) \quad (10)$$

is unique. In fact, instead of $\mathcal{A}^\varepsilon_X$, the original statement [10, Theorem 2] of Lemma 1 is stated in terms of a decreasing rearrangement of the probability masses of $P_X(\cdot)$. Assume that $X = \{1, 2, \ldots\}$ and

$$P_X(1) \geq P_X(2) \geq P_X(3) \geq P_X(4) \geq P_X(5) \geq \cdots. \quad (11)$$

Then, the subset $\mathcal{A}^\varepsilon_X$ can be written as

$$\mathcal{A}^\varepsilon_X = \begin{cases} \emptyset & \text{if } P_X(1) \geq 1 - \varepsilon, \\ \{1, 2, \ldots, J\} & \text{otherwise}, \end{cases} \quad (12)$$

where the positive integer $J$ is chosen so that

$$J = \sup \left\{ j \geq 0 \left| \sum_{k=1}^j P_X(k) < 1 - \varepsilon \right. \right\}. \quad (13)$$

1In [7, Definition I.1], Renner and Wolf also proposed another definition of the $\varepsilon$-smooth Rényi entropy.
In this case, it is clear that the $P_X$-probability of $\mathcal{A}_X^\varepsilon$ can be written as

$$P_X(\mathcal{A}_X^\varepsilon) = \sum_{k=1}^{J} P_X(k).$$

(14)

Now, we shall investigate asymptotic expansions of $H_\alpha^\varepsilon(X^n)$ as $n \to \infty$ for fixed real parameters $0 < \alpha < 1$ and $0 < \varepsilon < 1$, where $X^n = (X_1, \ldots, X_n)$ denotes $n$ i.i.d. copies of $X$. Define the following three information quantities:

$$H(X) = H(P_X) : = \sum_{x \in X} P_X(x) \log \frac{1}{P_X(x)},$$

(15)

$$V(X) = V(P_X) : = \sum_{x \in X} P_X(x) \left( \log \frac{1}{P_X(x)} - H(X) \right)^2,$$

(16)

$$T(X) = T(P_X) : = \sum_{x \in X} P_X(x) \left| \log \frac{1}{P_X(x)} - H(X) \right|^3.$$

(17)

In addition, denote by $\Phi^{-1} : (0, 1) \to \mathbb{R}$ the inverse of the Gaussian cumulative distribution function

$$\Phi(u) := \int_{-\infty}^{u} \varphi(t) \, dt,$$

(18)

where

$$\varphi(t) := \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$  

(19)

The following theorem states an asymptotic expansion of the smooth Rényi entropy up to the third-order term.

**Theorem 1.** Fix two real numbers $0 < \alpha < 1$ and $0 < \varepsilon < 1$. If $V(X)$ is zero, then

$$H_\alpha^\varepsilon(X^n) = n H(X) + O(1) \quad \text{as } n \to \infty.$$  

(20)

On the other hand, if $V(X)$ is positive and finite, then

$$H_\alpha^\varepsilon(X^n) = n H(X) + O(\sqrt{n}) \quad \text{as } n \to \infty.$$  

(21)

In particular, if $V(X)$ is positive and $T(X)$ is finite, then

$$H_\alpha^\varepsilon(X^n) = n H(X) - \sqrt{n V(X)} \Phi^{-1}(\varepsilon) - \frac{1}{2(1 - \alpha)} \log n + O(1) \quad \text{as } n \to \infty.$$  

(22)

**Proof of Theorem 1:** Since (20) and (21) are special cases of Theorem 2 stated later when $Y$ is almost surely constant; we defer their proofs to the proof of Theorem 2. It remains to prove the asymptotic expansion stated in (22).

Similar to [31, Equation (13)], define the $\varepsilon$-cutoff random transformation action on a real-valued r.v. $Z$ by

$$\langle Z \rangle_\varepsilon := \begin{cases} Z & \text{if } Z < \eta, \\ B Z & \text{if } Z = \eta, \\ 0 & \text{if } Z > \eta, \end{cases}$$

(23)

where $B$ is the Bernoulli r.v. with parameter $1 - \beta$ in which $B \perp\!\perp Z$, and $\eta \in \mathbb{R}$ and $0 \leq \beta < 1$ are chosen so that

$$\mathbb{P}(Z > \eta) + \beta \mathbb{P}(Z = \eta) = \varepsilon.$$  

(24)

Then, we assert the following lemma.

**Lemma 2.** Let $Z_1, Z_2, \ldots$ be a sequence of independent and real-valued r.v.’s. For each positive integer $n$, define the following three quantities:

$$E_n = \sum_{i=1}^{n} \mathbb{E}[Z_i],$$

(25)

$$V_n = \sum_{i=1}^{n} \mathbb{E}[(Z_i - \mathbb{E}[Z_i])^2],$$

(26)

$$T_n = \sum_{i=1}^{n} \mathbb{E}|Z_i - \mathbb{E}[Z_i]|^3.$$  

(27)

Suppose that
• there exists a positive constant $c_1$ such that $n c_1 < V_n < n / c_1$ for sufficiently large $n$, and
• there exists a positive constant $c_2$ such that $T_n < c_2 V_n$ for sufficiently large $n$.

For any fixed real numbers $0 < \varepsilon < 1$ and $s > 0$, it holds that
\[
\frac{1}{s} \log \mathbb{E} \left[ \left( \exp \left( s \sum_{i=1}^{n} Z_i \right) \right) \right] = E_n - \sqrt{V_n} \Phi^{-1}(\varepsilon) - \frac{1}{2} s \log n + O(1) \quad \text{(as $n \to \infty$).} \tag{28}
\]

**Proof of Lemma 2:** We prove Lemma 2 by applying Polyanskiy, Poor, and Verdú’s upper bound [14, Lemma 47] on the antilogarithm of the left-hand side of (28), and by establishing a corresponding inequality in the opposite direction. See Appendix A for details.

Let $X^n = (X_1, \ldots, X_n)$ be $n$ i.i.d. copies of $X$. For each $x \in X^n$, define the information density $\imath_n(x)$ by
\[
\imath_n(x) \coloneqq \log \frac{1}{P_{X^n}(x)}.
\tag{29}
\]
For the sake of brevity, white $\mathcal{A}_n = \mathcal{A}_n^c$. Choose an $x^* \in X^n \setminus \mathcal{A}_n$ so that
\[
x^* \in \arg \max_{x \in X^n \setminus \mathcal{A}_n} P_{X^n}(x).
\tag{30}
\]
Then, it follows from Lemma 1 that
\[
H_n^c(X^n) \leq \frac{1}{1 - \alpha} \log \left( \sum_{x \in \mathcal{A}_n} P_{X^n}(x)^{\alpha} + \left( 1 - \varepsilon - P_{X^n}(\mathcal{A}_n) \right) \right)
\leq \frac{1}{1 - \alpha} \log \left( \sum_{x \in \mathcal{A}_n} P_{X^n}(x) \exp \left( (1 - \alpha) \log \frac{1}{P_{X^n}(x)} \right) + \left( 1 - \varepsilon - P_{X^n}(\mathcal{A}_n) \right) \exp \left( (1 - \alpha) \log \frac{1}{1 - \varepsilon - P_{X^n}(\mathcal{A}_n)} \right) \right)
\leq \frac{1}{1 - \alpha} \log \left( \mathbb{E} \left[ \exp \left( (1 - \alpha) \log \frac{1}{P_{X^n}(x)} \right) \right] \right)
\leq \frac{1}{1 - \alpha} \log \left( \mathbb{E} \left[ \exp \left( (1 - \alpha) \log \frac{1}{P_{X^n}(x)} \right) \right] \right) + 1
\tag{33}
\]
where the last equality follows by the definition of $\langle \cdot \rangle_e$ stated in (23). Noting that
\[
0 < 1 - \varepsilon - P_{X^n}(\mathcal{A}_n) \leq P_{X^n}(x^*),
\tag{32}
\]
it follows from (31) that
\[
\frac{1}{1 - \alpha} \log \left( \mathbb{E} \left[ \exp \left( (1 - \alpha) \log \frac{1}{P_{X^n}(x)} \right) \right] \right) \leq H_n^c(X^n) \leq \frac{1}{1 - \alpha} \log \left( \mathbb{E} \left[ \exp \left( (1 - \alpha) \log \frac{1}{P_{X^n}(x)} \right) \right] \right) + 1
\tag{33}
\]
for every $n \geq 1$. Since $\imath_n(x) \geq n H_{\omega}(P)$, we see that
\[
\mathbb{E} \left[ \exp \left( (1 - \alpha) \imath_n(X^n) \right) \right] \geq (1 - \varepsilon) \exp \left( n (1 - \alpha) H_{\omega}(X) \right),
\tag{34}
\]
where the min-entropy $H_{\omega}(X)$ is defined as
\[
H_{\omega}(X) \coloneqq \lim_{\alpha \to \infty} H_\alpha(X) = \log \left( \frac{1}{\max_{x \in X} P(X)} \right).
\tag{35}
\]
In addition, since $V(P) > 0$ implies that $H_{\omega}(P) > 0$, we can find an integer $n_0$ satisfying
\[
(1 - \varepsilon) \exp \left( n (1 - \alpha) H_{\omega}(X) \right) > 1
\tag{36}
\]
for every $n \geq n_0$. Hence, it follows from (33), (34), and (36) that
\[
\frac{1}{1 - \alpha} \log \mathbb{E} \left[ \exp \left( (1 - \alpha) \imath_n(X^n) \right) \right] \leq H_n^c(P^n) \leq \frac{1}{1 - \alpha} \log \mathbb{E} \left[ \exp \left( (1 - \alpha) \imath_n(X^n) \right) \right] + \frac{1}{1 - \alpha} \log 2
\tag{37}
\]
for every \( n \geq n_0 \). Now, Theorem 1 follows from (37) and Lemma 2 with \( s = 1 - \alpha \), completing the proof of Theorem 1.  

**Remark 2.** In the right-hand side of (22) stated in Theorem 1, while the first-order term \(+ n H(X)\) and the second-order term \(-\sqrt{n V(X)} \Phi^{-1}(\varepsilon)\) are independent of \( \alpha \), the third-order term \+(log n)/(2(1-\alpha))\) depends on \( \alpha \). As will be seen in (139) of Section III-B, these dependencies characterize the difference between the fundamental limits of fixed-to-fixed length (block) source coding [17] and Campbell’s source coding problems [2] in the almost lossless regime.

**Remark 3.** The left-hand side of (28) stated in Lemma 2 is asymptotically equal to the cumulant generating function of \((Z_1 + \cdots + Z_n)_e\) with the normalization factor \(1/s\), and we now consider the expectation of \((Z_1 + \cdots + Z_n)_e\). A minor extension of Kostina, Polyanskiy, and Verdú’s result [31, Lemma 1] shows that

\[
\mathbb{E} \left[ \left( \sum_{t=1}^{n} Z_t \right) / \varepsilon \right] = (1 - \varepsilon) E_n - \sqrt{V_n} f_G(\varepsilon) + O(1) \quad (n \to \infty),
\]

where the function \( f_G : [0, 1] \to [0, 1/\sqrt{2\pi}] \) is defined by

\[
f_G(s) := \begin{cases} \varphi(\Phi^{-1}(s)) & \text{if } 0 < s < 1, \\ 0 & \text{if } s = 0 \text{ or } s = 1. \end{cases}
\]

Refer to [16, Appendix C] for a proof of (38).

**Remark 4.** For a general source \( X = (X^n) = (X_1^{(n)}, \ldots, X_n^{(n)})_{n=1}^\infty \) satisfying the strong converse property, Koga [10, Theorem 3] showed the following asymptotic expansion:

\[
H^\varepsilon_n(X^n) = n H(X) + o(n) \quad (n \to \infty),
\]

where the spectral entropy rate \( H(X) \) is defined as the limit in probability of the sequence

\[
\left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \right\}_{n=1}^\infty.
\]

provided that \( X \) satisfies the strong converse property. If \( X^n \) consists of \( n \) i.i.d. copies of \( X \), then (40) can be specialized to

\[
H^\varepsilon_n(X^n) = n H(X) + o(n) \quad (n \to \infty),
\]

which is a more general result than the following asymptotic result

\[
\lim_{\varepsilon \to 0^+} \lim_{n \to \infty} \frac{1}{n} H^\varepsilon_n(X^n) = H(X)
\]

shown by Renner and Wolf [8, Lemma 3]. These prior results are consistent with the related results in Theorem 1.

Note that \( H^\varepsilon_n(X) \) can be a negative number. In fact, it is easy to see that

\[
\lim_{\varepsilon \to 1^-} H^\varepsilon_n(X) = -\infty
\]

for every discrete r.v. \( X \) and every \( 0 < \alpha < 1 \). Theorem 1 or (42) implies that \( H^\varepsilon_n(X^n) \) is positive for sufficiently large \( n \), provided that \( H(X) > 0 \).

While the Rényi entropy satisfies the additivity property for independent r.v.’s, i.e.,

\[
H_\alpha(X^n) = n H_\alpha(X)
\]

for i.i.d. r.v.’s \( X^n = (X_1, \ldots, X_n) \), the \( \varepsilon \)-smooth Rényi entropy does not satisfy the additivity property in general. Thus, we see from Theorem 1 and (45) that

\[
\lim_{n \to \infty} \frac{H^\varepsilon_n(X^n)}{H_\alpha(X^n)} = \frac{H(X)}{H_\alpha(X)} \leq 1
\]

for every \( 0 < \alpha < 1 \) and \( 0 < \varepsilon < 1 \), provided that \( 0 < H_\alpha(X) < \infty \), where note that \( H(X) \leq H_\alpha(X) \) with equality if and only if \( V(X) = 0 \). Namely, the \( \varepsilon \)-smoothing reduces the Rényi entropy in the first-order term.
B. Smooth Conditional Rényi Entropy—Kuzuoka’s Proposal

Recently, Kuzuoka [11] introduced the smooth conditional Rényi entropy to characterize fundamental limits of several information-theoretic problems subject to constraints on average error probabilities, where the averaging is taken with respect to the common side-information \( Y \) in this study. Let \( \mathcal{Y} \) be a countable alphabet. Consider a \( \mathcal{Y} \)-valued r.v. \( Y \) playing the role of side-information of \( X \). Denote by \( P_{X,Y} := P \circ (X,Y)^{-1} \) (resp. \( P_Y := P \circ Y^{-1} \)) the joint (resp. marginal) probability distribution induced by \((X,Y)\) (resp. \( Y \)). Assume without loss of generality that \( P_Y(y) > 0 \) for every \( y \in \mathcal{Y} \). Then, the conditional probability distribution \( P_{X|Y} \) of \( X \) given \( Y \) is defined as

\[
P_{X|Y}(x) = P_{X,Y}(x, y) / P_Y(y),
\]

for each \((x, y) \in X \times \mathcal{Y} \). Given \( \alpha \in (0, 1) \cup (1, \infty) \) and \( 0 \leq \epsilon < 1 \), Kuzuoka [11] defined the \( \epsilon \)-smooth conditional Rényi entropy of \( X \) given \( Y \) by

\[
H^\epsilon_\alpha(X \mid Y) := \frac{\alpha}{1 - \alpha} \log \inf_{Q \in \mathcal{B} \subset \mathcal{P}_{X \times \mathcal{Y}}} \left( \sum_{y \in \mathcal{Y}} \left( \sum_{x \in X} Q(x, y)^\alpha \right)^{1/\alpha} \right),
\]

where the infimum is taken over the collection \( \mathcal{B} \subset \mathcal{P}_{X \times \mathcal{Y}} \) of sub-probability distributions \( Q \) on \( X \times \mathcal{Y} \) given as

\[
\mathcal{B} = \left\{ Q \left| \sum_{(x, y) \in X \times \mathcal{Y}} Q(x, y) \geq 1 - \epsilon \right. \text{ and } 0 \leq Q(a, b) \leq P_{X,Y}(a, b) \text{ for all } (a, b) \in X \times \mathcal{Y} \right\}.
\]

Note that \( H^\epsilon_\alpha(X \mid Y) \) coincides with Arimoto’s conditional Rényi entropy \( H_\alpha(X \mid Y) \) [12] if \( \epsilon = 0 \). In other words, we see that

\[
H^\epsilon_\alpha(X \mid Y) \bigg|_{\epsilon=0} = H_\alpha(X \mid Y) := \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in \mathcal{Y}} \left( \sum_{x \in X} P_{X,Y}(x, y)^\alpha \right)^{1/\alpha} \right).
\]

Moreover, it is clear that \( H^\epsilon_\alpha(X \mid Y) \) coincides with \( H_\alpha^\epsilon(X) \) defined in (3) if \( \mathcal{Y} \) is a singleton.

**Remark 5.** In [11], Kuzuoka called \( H^\epsilon_\alpha(X \mid Y) \) the conditional smooth Rényi entropy. On the other hand, we see that it is defined by applying the smoothing operation on Arimoto’s conditional Rényi entropy \( H_\alpha(X \mid Y) \), where the smoothing operation is taken with respect to the joint distribution \( P_{X,Y} \) with the smoothness parameter \( \epsilon \). From this perspective, in this paper, we call \( H^\epsilon_\alpha(X \mid Y) \) the smooth conditional Rényi entropy.

Given a real-valued r.v. \( Z \), define the \( \alpha \)-expectation operator \( \mathbb{E}^{(\alpha)} \) as

\[
\mathbb{E}^{(\alpha)}[Z] := \frac{\alpha}{1 - \alpha} \log \mathbb{E} \left[ \exp \left( \frac{1 - \alpha}{\alpha} Z \right) \right]
\]

for each \( \alpha \in (0, 1) \cup (1, \infty) \), where \( \mathbb{E}[\cdot] \) denotes the expectation operator, and \( \exp(\cdot) := 2^\cdot \) denotes the exponential function of \( u \in \mathbb{R} \) with base-2. After some algebra, Arimoto’s conditional Rényi entropy \( H_\alpha(X \mid Y) \) can be written as

\[
H_\alpha(X \mid Y) = \mathbb{E}^{(\alpha)}[H_\alpha(P_{X,Y})] = \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in \mathcal{Y}} P_Y(y) \exp \left( \frac{1 - \alpha}{\alpha} H_\alpha(P_{X,Y=\cdot}) \right) \right).
\]

Inspired by (52), given a function \( \delta : \mathcal{Y} \to [0, 1] \) and a real number \( 0 < \alpha < 1 \), we define

\[
\mathcal{H}_\alpha^{(\cdot)}(X \mid Y) := \mathbb{E}^{(\alpha)}[H_\alpha^{(\cdot)}(P_{X,Y})] = \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in \mathcal{Y}} P_Y(y) \exp \left( \frac{1 - \alpha}{\alpha} H_\alpha^{(\cdot)}(P_{X,Y=\cdot}) \right) \right),
\]

where \( H_\alpha^{(\cdot)}(P_{X,Y=\cdot}) \) is given as

\[
H_\alpha^{(\cdot)}(P_{X,Y=\cdot}) = \begin{cases} 
\frac{1}{1 - \alpha} \log \left( \inf_{Q \in \mathcal{B}^{(\cdot)}} \left( \sum_{x \in X} Q(x)^\alpha \right) \right) & \text{if } 0 \leq \delta(y) < 1, \\
0 & \text{if } \delta(y) = 1
\end{cases}
\]

for each \( y \in \mathcal{Y} \), and \( \mathcal{B}^{(\cdot)} \) is defined as in (4) for each \( y \in \mathcal{Y} \). In particular, if \( \delta(y) = \epsilon \) for every \( y \in \mathcal{Y} \), then we write

\[
\mathcal{H}_\alpha^\epsilon(X \mid Y) = \mathcal{H}_\alpha^{(\cdot)}(X \mid Y)
\]

\[\text{In (52), assume that } \exp(-\infty) = 0.\]
for the sake of brevity. Kuzuoka [11, Theorem 1] derived a formula of $H^F_n(X \mid Y)$ as a generalization of Lemma 1: Kuzuoka’s formula can be restated as follows:

**Lemma 3** (Kuzuoka [11, Theorem 1]). For any $0 < \alpha < 1$ and $0 \leq \varepsilon < 1$, it holds that

$$H^F_n(X \mid Y) = \inf_{\delta(\cdot) \in \mathcal{E}_\varepsilon(X)} \mathcal{H}^{n(\cdot)}_{n}\big( X \mid Y \big),$$

where the infimum is taken over the collection $\mathcal{E}_\varepsilon(X)$ of functions $\delta : \mathcal{Y} \to [0, 1]$ satisfying $\mathbb{E}[\delta(Y)] = \varepsilon$.

Let $\{(X_i, Y_i)\}_{i=1}^\infty$ be i.i.d. copies of $(X, Y)$. Defining two quantities

$$H(X \mid Y) := \mathbb{E} \left[ \log \frac{1}{P_{X|Y}(X \mid Y)} \right] = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} P_{X,Y}(x,y) \log \frac{1}{P_{X|Y}(x \mid y)},$$

$$U(X \mid Y) := \mathbb{E} \left[ \log \frac{1}{P_{X|Y}(X \mid Y)} - H(X \mid Y) \right] = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} P_{X,Y}(x,y) \left( \log \frac{1}{P_{X|Y}(x \mid y)} - \sum_{b \in \mathcal{Y}} \sum_{a \in \mathcal{X}} P_{X,Y}(a, b) \log \frac{1}{P_{X|Y}(a \mid b)} \right)^2,$$

we state asymptotic expansions of $H^F_n(X^n \mid Y^n)$ as $n \to \infty$ for fixed real parameters $0 < \alpha < 1$ and $0 < \varepsilon < 1$ as follows:

**Theorem 2.** Fix two real numbers $0 < \alpha < 1$ and $0 < \varepsilon < 1$. If $U(X \mid Y) = 0$, then

$$H^F_n(X^n \mid Y^n) = n H(X \mid Y) + O(1) \quad (\text{as } n \to \infty).$$

On the other hand, if $0 < U(X \mid Y) < \infty$, then

$$H^F_n(X^n \mid Y^n) = n H(X \mid Y) + O(\sqrt{n}) \quad (\text{as } n \to \infty).$$

**Proof of Theorem 2:** Let $\delta : \mathcal{Y} \to [0, 1]$ be a map. For each $y \in \mathcal{Y}$ satisfying $0 \leq \delta(y) < 1$, denote by $\mathcal{A}^{\delta(y)}_{X^n}(y) \subseteq \mathcal{X}$ the proper subset defined as in $\mathcal{A}^{\delta(y)}_{X|Y}(y)$ stated in (6) and (7) so that the parameter $\varepsilon$ and the probability distribution $P_{X|Y}(\cdot \mid y)$ are replaced by $\delta(y)$ and $P_{X|Y}(\cdot \mid y)$, respectively. If $\delta(y) = 1$, then suppose that $\mathcal{A}^{\delta(y)}_{X^n|Y^n}(y) = \emptyset$. It follows from Lemmas 1 and 3 that

$$H^F_n(X \mid Y) = \frac{1}{1 - \alpha} \log \left( \inf_{\delta(\cdot) \in \mathcal{E}_\varepsilon(X)} \sum_{y \in \mathcal{Y}} P_Y(y) \left( \sum_{x \in \mathcal{A}^{\delta(y)}_{X^n|Y^n}(y)} P_{X|Y}(x | y)^\alpha + \left( 1 - \delta(y) - P_{X|Y}(\mathcal{A}^{\delta(y)}_{X|Y^n}(y) | y) \right)^a \right)^{1/\alpha} \right),$$

for every $0 < \alpha < 1$.

Suppose that $U(X \mid Y) = 0$. We firstly aim to now prove the asymptotic expansion in (59). If $H(X \mid Y) = 0$, it is immediate from the definition that $H^F_n(X^n \mid Y^n)$ is constant for $n$, and it remains to consider the case where $H(X \mid Y) > 0$. In this case, the upper bound part of (59) can also be directly proven by the definition of $H^F_n(X^n \mid Y^n)$ stated in (48). On the other hand, the lower bound part of (59) can be proven by employing Lemma 3 and the reverse Markov inequality (cf. [19, Lemma 5.6.1]).

See Appendix B for proofs of these asymptotic bounds.

Now suppose that $0 < U(X \mid Y) < \infty$, and we secondly prove the asymptotic expansion in (60). In this case, the upper bound part\(^3\) of (60) can be proven by applying Chebyshev’s inequality. On the other hand, the lower bound part of (60) can be proven by employing Lemma 3, Chebyshev’s inequality, and the reverse Markov inequality. See Appendix C for proofs of these asymptotic bounds. This completes the proof of Theorem 2.

As will be shown in Section III, together with certain one-shot coding theorems formulated by the smooth conditional Rényi entropy $H^F_n(X^n \mid Y^n)$, Theorem 2 characterizes the exact first-order result and the order of the residual term (the scaling of $\sqrt{n}$) for various information-theoretic problems [2]–[6], [11] under the average error criterion.

**Remark 6.** Kuzuoka [11, Theorem 2] showed an asymptotic expansion of the $\varepsilon$-smooth conditional Rényi entropy $H^F_n(X^n \mid Y^n)$, Theorem 2 characterizes the exact first-order result and the order of the residual term (the scaling of $\sqrt{n}$) for various information-theoretic problems [2]–[6], [11] under the average error criterion.

\[ H^F_n(X^n \mid Y^n) = n H(X \mid Y) + o(n) \quad (\text{as } n \to \infty), \]

and this first-order term is consistent with the results in Theorem 2.

\(^3\)Given an asymptotic expansion $f(n) = g(n) + O(h(n))$ as $n \to \infty$, its upper bound part means that there exist two positive constants $c_0$ and $n_0$ such that $f(n) \leq g(n) + c_0 \cdot h(n)$ for all $n \geq n_0$. Similarly, its lower bound part means that there exist two positive constants $c_1$ and $n_1$ such that $f(n) \geq g(n) - c_1 \cdot h(n)$ for all $n \geq n_1$. 
C. A Novel Definition: Conditional Smooth Rényi Entropy

Whereas Kuzuoka [11] proposed $H^\varepsilon_\alpha(X \mid Y)$ to handle the average error criterion in several information-theoretic problems, we now introduce the conditional $\varepsilon$-smooth Rényi entropy $\tilde{H}^\varepsilon_\alpha(X \mid Y)$ to handle the maximum error criterion, where the maximum (more precisely, the supremum) is taken with respect to all realizations $y$ of the side-information $Y$ in this study. For each $\alpha \in (0,1) \cup (1,\infty)$ and $0 \leq \varepsilon < 1$, define

$$\tilde{H}^\varepsilon_\alpha(X \mid Y) := \frac{\alpha}{1-\alpha} \log \left( \sum_{y \in \mathcal{Y}} P_Y(y) \inf_{Q \in \mathcal{B}^\varepsilon(X \mid Y \mid y)} \left( \sum_{x \in \mathcal{X}} Q(x)^{\alpha} \right)^{1/\alpha} \right), \quad (63)$$

where for each $y \in \mathcal{Y}$, the infimum is taken over the collection $\mathcal{B}^\varepsilon(X \mid Y \mid y)$ of sub-probability distributions $Q$ on $\mathcal{X}$; see (4) for the definition of $\mathcal{B}^\varepsilon(\cdot)$. Similar to Kuzuoka’s proposal $H^\varepsilon_\alpha(X \mid Y)$ defined in (48), note that $\tilde{H}^\varepsilon_\alpha(X \mid Y)$ coincides with Arimoto’s conditional Rényi entropy $H_\alpha(X \mid Y)$ if $\varepsilon = 0$, and $\tilde{H}^\varepsilon_\alpha(X \mid Y)$ coincides with $H^\varepsilon_\alpha(X)$ defined in (3) if $X$ and $Y$ are independent. In contrast to Lemma 3, it can be verified that

$$\tilde{H}^\varepsilon_\alpha(X \mid Y) = \tilde{H}^\varepsilon_\alpha(X \mid Y) = \frac{\alpha}{1-\alpha} \log \left( \sum_{y \in \mathcal{Y}} P_Y(y) \exp \left( \frac{1-\alpha}{\alpha} H^\varepsilon_\alpha(P_X \mid Y \mid y) \right) \right). \quad (64)$$

**Remark 7.** In contrast to the smooth conditional Rényi entropy $H^\varepsilon_\alpha(X \mid Y)$, we call $\tilde{H}^\varepsilon_\alpha(X \mid Y)$ the conditional smooth Rényi entropy in this paper. This terminology comes from the observation that $\tilde{H}^\varepsilon_\alpha(X \mid Y)$ is defined by conditioning the smooth Rényi entropy with the smoothness parameter $\varepsilon$; see (52) and (64).

Now, we shall provide asymptotic expansions of $\tilde{H}^\varepsilon_\alpha(X^n \mid Y^n)$ as $n \to \infty$. Define

$$H(P_{X^n \mid Y^n}) := \mathbb{E} \log \left[ \frac{1}{P_{X^n \mid Y^n}(X \mid Y)} \right], \quad (65)$$

$$V(X \mid Y) := \mathbb{E} \left( \log \left[ \frac{1}{P_{X \mid Y}(X \mid Y)} - H(P_{X \mid Y}) \right] \right)^2 \quad = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} P_{X,Y}(x,y) \log \left[ \frac{1}{P_{X \mid Y}(x \mid y)} - \sum_{a \in \mathcal{X}} P_{X \mid Y}(a \mid y) \log \frac{1}{P_{X \mid Y}(a \mid y)} \right]^2, \quad (66)$$

$$H^{(\alpha)}(X \mid Y) := \mathbb{E}^{[\alpha]}[H(P_{X \mid Y})] \quad = \frac{\alpha}{1-\alpha} \log \left( \sum_{y \in \mathcal{Y}} P_Y(y) \exp \left( \frac{1-\alpha}{\alpha} \sum_{x \in \mathcal{X}} P_{X \mid Y}(x \mid y) \log \frac{1}{P_{X \mid Y}(x \mid y)} \right) \right). \quad (67)$$

**Theorem 3.** Fix two real numbers $0 < \alpha < 1$ and $0 < \varepsilon < 1$. If $V(X \mid Y) = 0$, then

$$\tilde{H}^\varepsilon_\alpha(X^n \mid Y^n) = n H^{(\alpha)}(X \mid Y) + O(1) \quad \text{as } n \to \infty. \quad (68)$$

If $V(X \mid Y) > 0$ and $\sup_{y \in \mathcal{Y}} V(P_{X \mid Y \mid y}) < \infty$, then

$$\tilde{H}^\varepsilon_\alpha(X^n \mid Y^n) = n H^{(\alpha)}(X \mid Y) + O(\sqrt{n}) \quad \text{as } n \to \infty. \quad (69)$$

**Proof of Theorem 3:** It follows from Lemma 1 and (64) that

$$\tilde{H}^\varepsilon_\alpha(X \mid Y) = \frac{\alpha}{1-\alpha} \log \left( \sum_{y \in \mathcal{Y}} P_Y(y) \left( \sum_{x \in \mathcal{X}} P_{X \mid Y}(x \mid y)^{\alpha} + \left( 1 - \varepsilon - P_{X \mid Y}(\mathcal{A}_{X \mid Y \mid y}) \right)^{\alpha} \right)^{1/\alpha} \right) \quad (70)$$

for every $0 < \alpha < 1$ and $0 < \varepsilon < 1$, where the proper subset $\mathcal{A}_{X \mid Y \mid y}$ of $\mathcal{X}$ is given by (6) and (7) with $\delta(y) = \varepsilon$ for each $y \in \mathcal{Y}$.

Suppose that $V(X \mid Y) = 0$, and we firstly verify the asymptotic expansion in (68). It is clear that

$$P_{X \mid Y}(x \mid y) = \begin{cases} \exp \left( -H(P_{X \mid Y \mid y}) \right) & \text{if } P_{X \mid Y}(x \mid y) > 0, \\ 0 & \text{if } P_{X \mid Y}(x \mid y) = 0. \end{cases} \quad (71)$$

In this case, the upper bound part of (68) can be directly proven by the definition of $\tilde{H}^\varepsilon_\alpha(X \mid Y)$ stated in (63). On the other hand, we prove the lower bound part of (68) by employing (70) and the one-sided Chebyshev inequality, and by dividing into two cases: either $U(X \mid Y) = 0$ or $U(X \mid Y) > 0$. See Appendix D for proofs of these asymptotic bounds.
Next, suppose that $V(X \mid Y) > 0$ and $\sup_{y \in Y} V(P_{X \mid Y = y}) < \infty$. We then verify the asymptotic expansion in (69). The upper bound part of (69) can be proven by employing Chebyshev’s inequality. On the other hand, we prove the lower bound part of (69) by employing (70) and the one-sided Chebyshev’s inequality, and by dividing into two cases: either $U(X \mid Y) = V(X \mid Y)$ or $U(X \mid Y) > V(X \mid Y)$. See Appendix E-A for proofs of these asymptotic bounds. This completes the proof of Theorem 3.

As will be shown in Section III, together with certain one-shot coding theorems formulated by the smooth conditional Rényi entropy $\hat{H}_\alpha^n(X^n \mid Y^n)$, Theorem 3 characterizes the exact first-order result and the order of the residual term (the scaling of $\sqrt{n}$) for various information-theoretic problems [2]–[6], [11] under the maximum error criterion.

The following proposition delineates the difference between the first-order terms stated in Theorems 2 and 3.

**Proposition 1.** For any $0 < \alpha < 1$, it holds that

$$H(X \mid Y) \leq H^{(\alpha)}(X \mid Y) \leq H_\alpha(X \mid Y). \quad (72)$$

More precisely, we observe the following equality conditions:

- the left-hand inequality of (72) holds with equality if and only if $U(X \mid Y) = V(X \mid Y)$, and
- the right-hand inequality of (72) holds with equality if and only if $V(X \mid Y) = 0$.

**Proof of Proposition 1:** The left-hand inequality of (72) follows from Jensen’s inequality, and it follows from the equality condition of Jensen’s inequality that it holds with equality if and only if $H(P_{X \mid Y = y})$ is constant for every $y \in \mathcal{Y}$. On the other hand, it follows by the law of total variance that

$$U(X \mid Y) = V(X \mid Y) + \sum_{y \in \mathcal{Y}} P_Y(y) \left( H(X \mid Y) - H(P_{X \mid Y = y}) \right)^2, \quad (73)$$

which implies that $U(X \mid Y) = V(X \mid Y)$ if and only if $H(P_{X \mid Y = y})$ is constant for every $y \in \mathcal{Y}$. This is indeed the equality condition of the left-hand inequality of (72).

On the other hand, it is known that the Rényi entropy $\alpha \mapsto H_\alpha(X)$ is nonincreasing in $\alpha \geq 0$. More precisely, we readily see that $\alpha \mapsto H_\alpha(X)$ is strictly decreasing in $\alpha \geq 0$ if and only if $V(X) > 0$. Therefore, for each $y \in \mathcal{Y}$ and each $0 < \alpha < 1$, we observe that

$$H(P_{X \mid Y = y}) \leq H_\alpha(P_{X \mid Y = y}) \quad (74)$$

with equality if and only if $V(P_{X \mid Y = y}) = 0$. Applying (74) straightforwardly to the definition of $H^{(\alpha)}(X \mid Y)$ stated in (67), we obtain the right-hand inequality of (72) and the condition for equality. This completes the proof of Proposition 1.

**Example 1** (binary symmetric source). Let $X = \mathcal{Y} = \{0,1\}$. Given $0 < \delta < 1$, consider a pair $(X,Y)$ satisfying

$$P_{X,Y}(x,y) = \begin{cases} (1-\delta)/2 & \text{if } x = y, \\ \delta/2 & \text{if } x \neq y. \end{cases} \quad (75)$$

It is easy to see that

$$0 < U(X \mid Y) = V(X \mid Y) < \infty; \quad (76)$$

thus, it follows from Proposition 1 that

$$0 < H(X \mid Y) = H^{(\alpha)}(X \mid Y) < H_\alpha(X \mid Y). \quad (77)$$

**Example 2** (binary erasure source). Let $X = \{0,1\}$ and $\mathcal{Y} = \{0,1,?\}$. Given $0 < \delta < 1$, consider a pair $(X,Y)$ satisfying

$$P_{X,Y}(x,y) = \begin{cases} (1-\delta)/2 & \text{if } x = y, \\ \delta/2 & \text{if } y = ?, \\ 0 & \text{if } x \neq y \text{ and } y \neq ?. \end{cases} \quad (78)$$

It is easy to see that

$$0 = V(X \mid Y) < U(X \mid Y) < \infty; \quad (79)$$

thus, it follows from Proposition 1 that

$$0 < H(X \mid Y) < H^{(\alpha)}(X \mid Y) = H_\alpha(X \mid Y). \quad (80)$$

**Example 3** (binary symmetric erasure source). Let $X = \{0,1\}$ and $\mathcal{Y} = \{0,1,?\}$. Given two real numbers $0 < \delta_e < 1$ and $0 < \delta_c < 1$, consider a pair $(X,Y)$ satisfying

$$P_{X,Y}(x,y) = \begin{cases} (1-\delta_e - \delta_c)/2 & \text{if } x = y, \\ \delta_e/2 & \text{if } y = ?, \\ \delta_c/2 & \text{if } x \neq y \text{ and } y \neq ?. \end{cases} \quad (81)$$
It is easy to see that
\[ 0 < V(X \mid Y) < U(X \mid Y) < \infty; \]
thus, it follows from Proposition 1 that
\[ 0 < H(X \mid Y) < H^{(\alpha)}(X \mid Y) < H_\alpha(X \mid Y). \]

III. APPLICATIONS

This section provides applications of the results in Section II to Campbell’s source coding problem [2], [11], the guessing problem [4], [5], [11], and the task partition problem [6], all allowing errors. In this section, suppose that the order $\alpha$ is given as
\[ \alpha = \frac{1}{1 + \rho}, \]
for a given positive parameter $\rho$. Namely, note that $0 < \alpha < 1$.

A. Unified Approach—Converse and Achievability Bounds

In this subsection, we establish unified converse and achievability bounds that are applicable to the three above mentioned information-theoretic problems [2]–[6], [11]. Let $\rho > 0$ and $0 \leq \epsilon < 1$ be fixed. Given two deterministic maps $\epsilon : X \times Y \to [0, 1]$ and $\kappa : X \times Y \to (0, \infty)$, construct a stochastic map $K : X \times Y \to [0, 1]$ given as
\[ K(x, y) = \begin{cases} \kappa(x, y) & \text{with probability } 1 - \epsilon(x, y), \\ 0 & \text{with probability } \epsilon(x, y) \end{cases} \]
for each $(x, y) \in X \times Y$. In this study, all fundamental limits can be formulated by the $\rho$-th moment of $K(X, Y)$ with properly chosen functions $\epsilon : X \times Y \to [0, 1]$ and $\kappa : X \times Y \to (0, \infty)$. Therefore, we shall give certain lower and upper bounds on the $\rho$-th moment of $K(X, Y)$ under certain constraints.

We first provide unified converse bounds under the average and maximum error formalisms. Define
\[ R(\epsilon, \kappa) := \sup_{y \in Y} \frac{1}{\kappa(x, y)} \sum_{x \in X} \mathbb{1}_{\epsilon(x, y) < 1} \]
which represents a certain redundancy term in the converse bounds. It is worth mentioning that the sum in the right-hand side of (86) is taken over all $x \in X$ satisfying the constraint that $\epsilon(x, y) < 1$, and this constraint aids to establish valid converse bounds in the guessing and the task encoding problems over the countably infinite alphabet $X$. The following two lemmas are generalizations of Courtade and Verdù’s unified converse bound [3, Lemma 2] and Kumar, Sunny, Thakre, and Kumar’s unified converse bound [15, Theorem 18] from error-free settings (i.e., $\epsilon = 0$) to almost lossless settings in which the error probability is at most $\epsilon > 0$.

**Lemma 4** (unified converse bound—average error criterion). If
\[ \mathbb{E}[(\epsilon(X, Y))] \leq \epsilon, \]
then it holds that
\[ \frac{1}{\rho} \log \mathbb{E}[K(X, Y)^\rho] \geq H_\alpha^\epsilon(X \mid Y) - \log R(\epsilon, \kappa). \]

**Proof of Lemma 4:** See Appendix F.

**Lemma 5** (unified converse bound—maximum error criterion). If
\[ \sup_{y \in Y} \mathbb{E}[(\epsilon(X, Y) \mid Y = y)] \leq \epsilon, \]
then it holds that
\[ \frac{1}{\rho} \log \mathbb{E}[K(X, Y)^\rho] \geq H_\alpha^\epsilon(X \mid Y) - \log R(\epsilon, \kappa). \]

**Proof of Lemma 5:** See Appendix G.

Next, we provide a unified achievability bound that is applicable to both maximum and average error formalisms. Let $\delta : Y \to [0, 1]$ be a deterministic map. Recall that for each $y \in Y$, the proper subset $\mathcal{A}^\delta_{X \mid Y}(y) \subseteq X$ is defined to satisfy

\[ \mathbb{P}[X = x \mid Y = y] \geq \delta(y) \]
and (7), where \( \varepsilon \) and \( P_X(\cdot) \) are replaced by \( \delta(y) \) and \( P_{X|Y}(\cdot \mid y) \), respectively. For each \( y \in \mathcal{Y} \), choose an element \( x^*(y) \in X \setminus \mathcal{A}_{X|Y}^{\delta(y)}(y) \) so that
\[
x^*(y) \in \arg \max_{x \in X \setminus \mathcal{A}_{X|Y}^{\delta(y)}(y)} P_{X|Y}(x \mid y).
\]
and choose the number \( M(y) \) as
\[
M(y) = 1 - \delta(y) - P_{X|Y}(\mathcal{A}_{X|Y}^{\delta(y)}(y) \mid y).
\]
Moreover, define the conditional probability distribution \( Q_{X|Y}^{(\alpha, \delta(\cdot))}(x \mid y) \) as
\[
Q_{X|Y}^{(\alpha, \delta(\cdot))}(x \mid y) := \begin{cases} 
P_{X|Y}(x \mid y)^\alpha & \text{if } x \in \mathcal{A}_{X|Y}^{\delta(y)}(y), \\
\frac{M(y)^\alpha}{\sum_{a \in \mathcal{A}_{X|Y}^{\delta(y)}(y)} P_{X|Y}(a \mid y)^\alpha + M(y)^\alpha} & \text{if } x = x^*(y), \\
0 & \text{otherwise}
\end{cases}
\]
for each \((x, y) \in X \times \mathcal{Y}\). Here, note that \( Q_{X|Y}^{(\alpha, \delta(\cdot))} \) depends only on \( P_{X|Y}, \alpha \), and \( \delta(\cdot) \).

**Lemma 6** (unified achievability bound). Suppose that there exists a positive constant \( c \) such that
\[
k(x, y) Q_{X|Y}^{(\alpha, \delta(\cdot))}(x \mid y) \leq c
\]
for every \((x, y) \in X \times \mathcal{Y}\). Then, there exists a deterministic map \( \varepsilon : X \times \mathcal{Y} \to [0, 1] \) satisfying
\[
\mathbb{E}[\varepsilon(X, Y) \mid Y = y] = \delta(y) \quad \text{(for } y \in \mathcal{Y}),
\]
\[
\frac{1}{\rho} \log \mathbb{E}[K(X, Y)^\rho] \leq \bar{H}_\alpha(\cdot \mid Y) + \log c,
\]
where \( \bar{H}_\alpha(\cdot \mid Y) \) is defined in (53).

**Proof of Lemma 6:** See Appendix H.

To apply Lemma 6 for the average and maximum error formalisms to various information-theoretic problems, we choose an appropriate map \( \delta : \mathcal{Y} \to [0, 1] \) by referring to the identities of (56) in Lemma 3 and (64), respectively. More precisely, for the average error formalism, we find a map \( \delta^*(\cdot) \in \mathcal{E}_0 \) so that
\[
H_\alpha^\varepsilon(X \mid Y) \geq \bar{H}_\alpha^{\delta^*(\cdot)}(X \mid Y) - \zeta
\]
for an arbitrarily small \( \zeta > 0 \). On the other hand, for the maximum error formalism, we choose a constant function \( \delta : \mathcal{Y} \to [0, 1] \) as \( \delta(y) = \varepsilon \) for every \( y \in \mathcal{Y} \). In Campbell’s source coding problems \([2, 11]\) stated in the next subsection, our achievability bounds are proven by constructing Shannon codes with the conditional distribution \( Q_{X|Y}^{(\alpha, \delta(\cdot))} \) defined in (93), where the 1-bit redundancy terms of the Shannon codes can be obtained by choosing the constant \( c = 2^\zeta \).

**B. Campbell’s Source Coding Problem**

Given a correlated source \((X, Y)\), we consider compressing the source \( X \) into a variable-length binary string when the side-information \( Y \) is available at both encoder and decoder. Denote by
\[
\{0, 1\}^* := \{\emptyset\} \cup \bigcup_{n=1}^{\infty} \{0, 1\}^n
\]
the set of finite-length binary strings containing the empty string \( \emptyset \). Let \( F : X \times \mathcal{Y} \to \{0, 1\}^* \) and \( G : \{0, 1\}^* \times \mathcal{Y} \to X \) be two random maps playing the roles of a stochastic encoder and a stochastic decoder, respectively. We call this pair \((F, G)\) a variable-length stochastic code. For each \( y \in \mathcal{Y} \), we say that the codeword set
\[
C_y(X, Y, F) := \{\mathbf{b} \in \{0, 1\}^* \mid |P(F(X, Y) = \mathbf{b} \mid Y = y| > 0\}
\]
is prefix-free if for every distinct \( \mathbf{b}_1, \mathbf{b}_2 \in C_y(X, Y, F) \), a codeword \( \mathbf{b}_1 \) is not a prefix of another codeword \( \mathbf{b}_2 \).

Given a \( \{0, 1\}^* \)-valued r.v. \( B \), consider the cumulant generating function of codeword lengths \([2, 3]\) defined as
\[
\Lambda(B \| \rho) := \log \mathbb{E}[2^{\rho(B)}]
\]
for a positive parameter $\rho$, where $\ell : \{0,1\}^* \to \mathbb{N} \cup \{0\}$ stands for the length function of a binary string, i.e.,
\begin{align}
\ell(\emptyset) &= 0, \\
\ell(0) &= \ell(1) = 1, \\
\ell(00) &= \ell(01) = \ell(10) = \ell(11) = 2, \\
\ell(000) &= \ell(001) = \ell(010) = \ell(011) = \ell(100) = \ell(101) = \ell(110) = \ell(111) = 3,
\end{align}
and so on. Then, we are interested to characterize fundamental limits defined as the infimum of positive real numbers $L$ such that there exists a variable-length stochastic code $(F, G)$ satisfying
\[
\Lambda(F(X, Y) \parallel \rho) \leq \rho L
\]
under a certain constraint on error probabilities. We now introduce two error formalisms as follows:

**Definition 1** (average error criterion). A $(\rho, L, \varepsilon)_{\text{avg}}$-code for a correlated source $(X, Y)$ is a variable-length stochastic code $(F, G)$ such that (105) holds, the codeword set $C_y(X, Y, F)$ is prefix-free for every $y \in \mathcal{Y}$, and
\[
P\{X \neq g(F(X, Y), Y)\} \leq \varepsilon.
\]

**Definition 2** (maximum error criterion). A $(\rho, L, \varepsilon)_{\text{max}}$-code for a correlated source $(X, Y)$ is a variable-length stochastic code $(F, G)$ such that (105) holds, the codeword set $C_y(X, Y, F)$ is prefix-free for every $y \in \mathcal{Y}$, and
\[
\sup_{y \in \mathcal{Y}} P\{X \neq g(F(X, Y), Y) | Y = y\} \leq \varepsilon.
\]

Given two real numbers $\rho > 0$ and $0 \leq \varepsilon < 1$, define the following two fundamental limits:
\begin{align}
\Lambda_{\text{avg}}^*(X, Y \parallel \rho, \varepsilon) &:= \inf\{L > 0 \mid \text{there exists a } (\rho, L, \varepsilon)_{\text{avg}}\text{-code for } (X, Y)\}, \\
\Lambda_{\text{max}}^*(X, Y \parallel \rho, \varepsilon) &:= \inf\{L > 0 \mid \text{there exists a } (\rho, L, \varepsilon)_{\text{max}}\text{-code for } (X, Y)\}.
\end{align}

We now state the following one-shot coding theorems.

**Theorem 4** (average error criterion). For every $\rho > 0$ and $0 \leq \varepsilon < 1$, it holds that
\[
H_0^e(X | Y) \leq \Lambda_{\text{avg}}^*(X, Y \parallel \rho, \varepsilon) = H_0^e(X | Y) + 1 + \frac{1}{\rho} \log \left(\frac{1}{1 - \varepsilon}\right).
\]

**Proof of Theorem 4:** It is clear that for any stochastic code $(F, G)$ and a correlated source $(X, Y)$, there exists a deterministic decoder $g : \{0,1\}^* \times \mathcal{Y} \to X$ satisfying
\[
P\{X \neq g(F(X, Y), Y)\} \leq P\{X \neq g(F(X, Y), Y)\}.
\]
Thus, it suffices to consider deterministic decoders $g : \{0,1\}^* \times \mathcal{Y} \to X$. In this paper, a variable-length stochastic code $(F, g)$ is called a variable-length semi-stochastic code if $g$ is deterministic. Instead of $\Lambda(B \parallel \rho)$ defined in (100), we now consider a cutoff version of the cumulant generating function of codeword lengths as follows:
\[
\tilde{\Lambda}(X, Y, F, g \parallel \rho) := \log \mathbb{E}\left[2^{\rho f(F(X,Y))} 1_{\{X=g(F(X,Y),Y)\}}\right].
\]
Namely, instead of (105), we are interested in the infimum of positive real numbers $L$ such that there exists a variable-length semi-stochastic code $(F, g)$ satisfying
\[
\tilde{\Lambda}(X, Y, F, g \parallel \rho) \leq \rho L
\]
under the average error criterion.

**Definition 3.** Let $\rho > 0$, $L \geq 0$, and $0 \leq \varepsilon < 1$ be real numbers. Given a source $X$, a $(\rho, L, \varepsilon)_{\text{avg}}$-weak-code for the source $X$ is a variable-length semi-stochastic code $(F, g)$ such that (113) holds, the codeword set $C_y(X, Y, F)$ is prefix-free for every $y \in \mathcal{Y}$, and
\[
P\{X \neq g(F(X, Y), Y)\} \leq \varepsilon.
\]
Similar to $\Lambda_{\text{avg}}^*(X, Y \parallel \rho, \varepsilon)$ defined in (108), define
\[
\tilde{\Lambda}_{\text{avg}}^*(X, Y \parallel \rho, \varepsilon) := \inf\{L > 0 \mid \text{there exists a } (\rho, L, \varepsilon)_{\text{avg}}\text{-weak-code for } (X, Y)\}.
\]
Instead of $\Lambda_{\text{avg}}^*(X, Y \parallel \rho, \varepsilon)$, the following lemma establishes one-shot bounds on $\tilde{\Lambda}_{\text{avg}}^*(X, Y \parallel \rho, \varepsilon)$.

**Lemma 7.** For any $\rho > 0$ and $0 \leq \varepsilon < 1$, it holds that
\[
H_0^e(X | Y) \leq \tilde{\Lambda}_{\text{avg}}^*(X, Y \parallel \rho, \varepsilon) = H_0^e(X | Y) + 1.
\]
Proof of Lemma 7: The converse and achievability bounds can be proven via the unified approaches as stated in Lemmas 4 and 6, respectively; see Appendix I for details.

The following lemma provides inequalities between $\Lambda_{\text{avg}}^*(X, Y \mid \rho, \varepsilon)$ and $\tilde{\Lambda}_{\text{avg}}^*(X, Y \mid \rho, \varepsilon)$.

Lemma 8. For any $\rho > 0$ and $0 < \varepsilon < 1$, it holds that

$$\tilde{\Lambda}_{\text{avg}}^*(X, Y \mid \rho, \varepsilon) \leq \Lambda_{\text{avg}}^*(X, Y \mid \rho, \varepsilon) \leq \tilde{\Lambda}_{\text{avg}}^*(X, Y \mid \rho, \varepsilon) + \frac{1}{\rho} \log \left( \frac{1}{1 - \varepsilon} \right).$$

(117)

Proof of Lemma 8: See Appendix J.
The proof of Theorem 4 is immediately completed by combining Lemmas 7 and 8.

Theorem 5 (maximum error criterion). For every $\rho > 0$ and $0 \leq \varepsilon < 1$, it holds that

$$\tilde{H}_{\alpha}^\rho(X \mid Y) \leq \Lambda_{\text{max}}^*(X, Y \mid \rho, \varepsilon) < \tilde{H}_{\alpha}^\rho(X \mid Y) + 1 + \frac{1}{\rho} \log \left( \frac{1}{1 - \varepsilon} \right).$$

(118)

Proof of Theorem 5: Similar to the previous subsection, we introduce a weaker version of semi-stochastic codes as follows:

Definition 4. Let $\rho > 0$, $L \geq 0$, and $0 \leq \varepsilon < 1$ be real numbers. Given a source $X$, a $(\rho, L, \varepsilon)_{\text{max}}$-weak-code for the source $X$ is a variable-length semi-stochastic code $(F, g)$ such that (113) holds, the codeword set $C_y(X, Y, F)$ is prefix-free for every $y \in Y$, and

$$\sup_{y \in Y} \mathbb{P}\{X \neq g(F(X, Y), Y) \mid Y = y\} \leq \varepsilon.$$  

(119)

Similar to $\Lambda_{\text{max}}^*(X, Y \mid \rho, \varepsilon)$ defined in (109), define

$$\tilde{\Lambda}_{\text{max}}^*(X, Y \mid \rho, \varepsilon) := \inf\{L > 0 \mid \text{there exists a } (\rho, L, \varepsilon)_{\text{max}}-\text{weak-code for the correlated source } (X, Y)\}.$$  

(120)

Instead of $\Lambda_{\text{max}}^*(X, Y \mid \rho, \varepsilon)$, the following lemma establishes one-shot bounds on $\tilde{\Lambda}_{\text{max}}^*(X, Y \mid \rho, \varepsilon)$.

Lemma 9. For any $\rho > 0$ and $0 \leq \varepsilon < 1$, it holds that

$$\tilde{H}_{\alpha}^\rho(X \mid Y) \leq \tilde{\Lambda}_{\text{max}}^*(X, Y \mid \rho, \varepsilon) < \tilde{H}_{\alpha}^\rho(X \mid Y) + 1.$$  

(121)

Proof of Lemma 9: See Appendix K.
The following lemma provides inequalities between $\Lambda_{\text{max}}^*(X, Y \mid \rho, \varepsilon)$ and $\tilde{\Lambda}_{\text{max}}^*(X, Y \mid \rho, \varepsilon)$.

Lemma 10. For any $\rho > 0$ and $0 < \varepsilon < 1$, it holds that

$$\tilde{\Lambda}_{\text{max}}^*(X, Y \mid \rho, \varepsilon) \leq \Lambda_{\text{max}}^*(X, Y \mid \rho, \varepsilon) \leq \tilde{\Lambda}_{\text{max}}^*(X, Y \mid \rho, \varepsilon) + \frac{1}{\rho} \log \left( \frac{1}{1 - \varepsilon} \right).$$  

(122)

Proof of Lemma 10: See Appendix L.
The proof of Theorem 5 is immediately completed by combining Lemmas 9 and 10.

Remark 8. The converse bound of Theorem 4 is the same as Kuzuoka’s converse bound [11, Theorem 6], and the achievability bound differs slightly compared to [11, Theorem 7].

Now, we can obtain the following asymptotic results.

Corollary 1 (average error criterion). Let $\rho > 0$ and $0 < \varepsilon < 1$ be fixed. If $U(X \mid Y) = 0$, then

$$\Lambda_{\text{avg}}^*(X^n, Y^n \mid \rho, \varepsilon) = n H(X \mid Y) + O(1) \quad \text{(as } n \to \infty).$$  

(123)

On the other hand, if $0 < U(X \mid Y) < \infty$, then

$$\Lambda_{\text{avg}}^*(X^n, Y^n \mid \rho, \varepsilon) = n H(X \mid Y) + O(\sqrt{n}) \quad \text{(as } n \to \infty).$$  

(124)

Proof of Corollary 1: Corollary 1 follows from Theorems 2 and 4.

Corollary 2 (maximum error criterion). Let $\rho > 0$ and $0 < \varepsilon < 1$ be fixed. If $V(X \mid Y) = 0$, then

$$\Lambda_{\text{max}}^*(X^n, Y^n \mid \rho, \varepsilon) = n H^{(\alpha)}(X \mid Y) + O(1) \quad \text{(as } n \to \infty).$$  

(125)

On the other hand, if $V(X \mid Y) > 0$ and $\sup_{y \in Y} V(P_{X \mid Y = y}) < \infty$, then

$$\Lambda_{\text{max}}^*(X^n, Y^n \mid \rho, \varepsilon) = n H^{(\alpha)}(X \mid Y) + O(\sqrt{n}) \quad \text{(as } n \to \infty).$$  

(126)

Proof of Corollary 2: Corollary 2 follows from Theorems 3 and 5.
Now, consider compressing the source $X$ into a variable-length binary string in the absence of the side-information $Y$. Formally, when $Y$ is almost surely constant, we denote by

$$\Lambda^*(X \parallel \rho, \varepsilon) := \Lambda^*_{\text{avg}}(X, Y \parallel \rho, \varepsilon) = \Lambda^*_{\text{max}}(X, Y \parallel \rho, \varepsilon)$$  \hspace{1cm} (127)

our considered fundamental limit here.

**Remark 9.** If $\varepsilon = 0$ and $|\mathcal{Y}| = 1$, then Theorems 4 and 5 coincide with Campbell’s one-shot bounds [2, Equation (18)]:

$$H_d(X) \leq \Lambda^*(X \parallel \rho, 0) < H_d(X) + 1. \hspace{1cm} (128)$$

Now, we can get the following third-order asymptotic result.

**Corollary 3** (in the absence of side-information $Y$). Let $\rho > 0$ and $0 < \varepsilon < 1$ be fixed. If $V(X) = 0$, then

$$\Lambda^*(X^n \parallel \rho, \varepsilon) = n H(X) + O(1) \quad \text{(as } n \to \infty).$$  \hspace{1cm} (129)

On the other hand, if $0 < V(X) < \infty$, then

$$\Lambda^*(X^n \parallel \rho, \varepsilon) = n H(X) + O(\sqrt{n}) \quad \text{(as } n \to \infty).$$  \hspace{1cm} (130)

In particular, if $V(X) > 0$ and $T(X) < \infty$, then

$$\Lambda^*(X^n \parallel \rho, \varepsilon) = n H(X) - \sqrt{nV(X)\Phi^{-1}(\varepsilon)} - \frac{1 + \rho}{2\rho} \log n + O(1) \quad \text{(as } n \to \infty).$$  \hspace{1cm} (131)

**Proof of Corollary 3:** Corollary 3 follows from Theorems 1 and 4. □

Finally, consider compressing the source $X$ into a fixed-length binary string in the absence of the side-information $Y$. Namely, on the stochastic encoders $F : X \to \{0, 1\}^*$, we now impose the constraint that each codeword has the same length, i.e., the codeword length $\ell(F(X))$ is almost surely constant. This is the well-known fixed-to-fixed length (block) source coding problem.

**Definition 5** (block source coding). A $(\rho, L, \varepsilon)_{\text{FF}}$-code for a source $X$ is a stochastic code $(F, G)$ such that

$$\Lambda(F(X) \parallel \rho) \leq \rho L,$$  \hspace{1cm} (132)

$$\mathbb{P}\{X \neq G(F(X))\} \leq \varepsilon,$$  \hspace{1cm} (133)

and $\ell(F(X))$ is almost surely constant.

Given $\rho > 0$ and $0 \leq \varepsilon < 1$, consider the following fundamental limit:

$$\Lambda^*(X \parallel \rho, \varepsilon) := \inf\{L > 0 \mid \text{there exists a } (\rho, L, \varepsilon)_{\text{FF}}\text{-code for the source } X\}. \hspace{1cm} (134)$$

It is clear that

$$\Lambda^*(X \parallel \rho, \varepsilon) = \ell(F(X)) \quad \text{(a.s.),}$$  \hspace{1cm} (135)

provided that $\ell(F(X))$ is almost surely constant. Hence, the fundamental limit $\Lambda^*_{\text{FF}}(X \parallel \rho, \varepsilon)$ is independent of the parameter $\rho > 0$, and it can be written as

$$\Lambda^*_{\text{FF}}(X \parallel \rho, \varepsilon) = \lfloor \log(1 + |\mathcal{A}_{X_n}^c|) \rfloor,$$  \hspace{1cm} (136)

where the proper subset $\mathcal{A}_{X_n}^c \subseteq X$ is defined in (6) and (7). Therefore, it follows by Strassen’s seminal result [17] that

$$\Lambda^*_{\text{FF}}(X^n \parallel \rho, \varepsilon) = n H(X) - \sqrt{nV(X)\Phi^{-1}(\varepsilon)} - \frac{1}{2} \log n + O(1),$$  \hspace{1cm} (137)

provided that $V(X) > 0$ and $T(X) < \infty$.

We shall compare the two fundamental limits $\Lambda^*(X \parallel \rho, \varepsilon)$ and $\Lambda^*_{\text{FF}}(X \parallel \rho, \varepsilon)$. Since every fixed-to-fixed length source code is prefix-free, it is clear that a $(\rho, L, \varepsilon)_{\text{FF}}$-code for a source $X$ is a $(\rho, L, \varepsilon)$-code for the source $X$. Namely, we readily see that

$$\Lambda^*(X \parallel \rho, \varepsilon) \leq \Lambda^*_{\text{FF}}(X \parallel \rho, \varepsilon).$$  \hspace{1cm} (138)

Indeed, it follows from Corollary 3 and (138) that

$$\Lambda^*_{\text{FF}}(X^n \parallel \rho, \varepsilon) - \Lambda^*(X^n \parallel \rho, \varepsilon) = \frac{1}{2\rho} \log n + O(1) \quad \text{(as } n \to \infty),$$  \hspace{1cm} (139)

provided that $V(X) > 0$ and $T(X) < \infty$. In other words, the first- and second-order terms (i.e., the $n$ and $\sqrt{n}$ scales, respectively) of $\Lambda^*(X^n \parallel \rho, \varepsilon)$ and $\Lambda^*_{\text{FF}}(X^n \parallel \rho, \varepsilon)$ are the same, and the third-order term (i.e., the $\log n$ scale) of $\Lambda^*(X^n \parallel \rho, \varepsilon)$ is strictly smaller than that of $\Lambda^*_{\text{FF}}(X^n \parallel \rho, \varepsilon)$. Roughly speaking, the benefit of variable-length codewords appears only in the third-order term in Campbell’s fixed-to-variable length source coding problem [2].
C. Guessing Problem

We introduce Kuzuoka’s guessing problem [11, Section III]. A guessing strategy with a giving-up policy is a pair \((g, \pi)\) of deterministic maps \(g : \mathcal{X} \times \mathcal{Y} \to \mathbb{N}\) and \(\pi : \mathbb{N} \times \mathcal{Y} \to [0, 1]\) in which \(g(\cdot, y) : \mathcal{X} \to \mathbb{N}\) is bijective for each \(y \in \mathcal{Y}\). This pair \((g, \pi)\) induces the following strategy: Assume that the guesser knows the side-information \(Y = y\). For each guessing epoch, the guesser can stochastically give-up guessing based on the giving-up policy \(\pi\). Formally, at the \(k\)-th stage \((k \geq 1)\), he declares an error with probability \(\pi(k, y)\), or he asks the question “Is \(X = x_k\)?” with probability \(1 - \pi(k, y)\), where the candidate \(x_k\) is chosen by the guessing function \(g\) as \(g(x_k, y) = k\). The guesser repeats these epochs until he succeeds in guessing \(X\) or he declares an error. Construct a stochastic map \(\bar{G} : \mathcal{X} \times \mathcal{Y} \to \mathbb{N} \cup \{0\}\) so that

\[
\bar{G}(x, y) = \begin{cases} 
  g(x, y) \text{ with probability } \prod_{k=1}^{g(x,y)} (1 - \pi(k, y)) \\
   0 \text{ with probability } 1 - \prod_{k=1}^{g(x,y)} (1 - \pi(k, y)) 
\end{cases}
\]

for each \((x, y) \in \mathcal{X} \times \mathcal{Y}\). We call this stochastic map \(\bar{G} : \mathcal{X} \times \mathcal{Y} \to \mathbb{N} \cup \{0\}\) the giving-up guessing function induced by the guessing strategy \((g, \pi)\). Then, it is clear that the guesser declares some error if and only if \(\bar{G}(X, Y) = 0\). We now aim to minimize the guessing \(\rho\)-th moment \(\mathbb{E}[\bar{G}(X, Y)^\rho]\) for a fixed \(\rho > 0\) subject to certain error constraints. In other words, we are interested to characterize fundamental limits defined as the infimum of positive real numbers \(M\) such that there exists a guessing strategy \((g, \pi)\) satisfying

\[
\log \mathbb{E}[\bar{G}(X, Y)^\rho] \leq \rho M
\]

under a certain constraint on error probabilities. We now introduce two error formalisms as follows:

**Definition 6** (average error criterion). A \((\rho, M, \varepsilon)_{\text{avg}}\)-strategy for a correlated source \((X, Y)\) is a guessing strategy \((g, \pi)\) such that (141) holds and

\[
\mathbb{P}\{\bar{G}(X, Y) = 0\} \leq \varepsilon.
\]

**Definition 7** (maximum error criterion). A \((\rho, M, \varepsilon)_{\text{max}}\)-strategy for a correlated source \((X, Y)\) is a guessing strategy \((g, \pi)\) such that (141) holds and

\[
\sup_{y \in \mathcal{Y}} \mathbb{P}\{\bar{G}(X, Y) = 0 \mid Y = y\} \leq \varepsilon.
\]

Given two real numbers \(\rho > 0\) and \(0 \leq \varepsilon < 1\), define the following two fundamental limits

\[
G_{\text{avg}}(X, Y \| \rho, \varepsilon) := \inf\{M > 0 \mid \text{there exists a \((\rho, M, \varepsilon)_{\text{avg}}\)-strategy for } (X, Y)\}, \quad G_{\text{max}}^*(X, Y \| \rho, \varepsilon) := \inf\{M > 0 \mid \text{there exists a \((\rho, M, \varepsilon)_{\text{max}}\)-strategy for } (X, Y)\}.
\]

Now, we state the following one-shot bounds.

**Theorem 6** (average error criterion). For every \(\rho > 0\) and \(0 < \varepsilon < 1\), it holds that

\[
H_\alpha^e(X \mid Y) - \log \left(1 + \frac{H(X \mid Y)}{\varepsilon}\right) \leq G_{\text{avg}}(X, Y \| \rho, \varepsilon) \leq H_\alpha^e(X \mid Y).
\]

**Proof of Theorem 6:** For each \(y \in \mathcal{Y}\), denote by \(\varsigma_y : \mathbb{N} \to \mathcal{X}\) a bijection satisfying

\[
P_{X|Y}(\varsigma_y(1) \mid y) \geq P_{X|Y}(\varsigma_y(2) \mid y) \geq P_{X|Y}(\varsigma_y(3) \mid y) \geq P_{X|Y}(\varsigma_y(4) \mid y) \geq P_{X|Y}(\varsigma_y(5) \mid y) \geq \cdots.
\]

Define two parameters \(J\) and \(\xi\) by

\[
J := \sup_j \left\{ j \geq 0 \left| \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{k=1}^j P_{X|Y}(\varsigma_y(k) \mid y) < 1 - \varepsilon \right. \right\},
\]

\[
\xi := 1 - \varepsilon - \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{k=1}^J P_{X|Y}(\varsigma_y(k) \mid y),
\]

respectively. The following lemma characterizes an optimal guessing strategy under the average error criterion.

---

*Originally, Kuzuoka [11, Section III] introduced a positive error cost when the guesser declares an error. This setting was argued to be more practical; see also [16, Section IV]. On the other hand, he derived his one-shot bounds and general formula in the absence of the error cost. Our study also focuses on the guessing problem without the error cost.*
Lemma 11 (optimal guessing strategy—average error criterion). Consider a guessing strategy \((g^*, \pi^*_{\text{avg}})\) given by

\[
g^*(x, y) = \xi^{-1}_y(x), \tag{150}
\]

\[
\pi^*_{\text{avg}}(k, y) = \begin{cases} 
0 & \text{if } 1 \leq k \leq J, \\
1 - \frac{\xi}{P_{X,Y}(\xi_y(k), y)} & \text{if } k = J + 1, \\
1 & \text{if } J + 2 \leq k < \infty.
\end{cases} \tag{151}
\]

Denote by \(\bar{G}^*_\text{avg} : X \times Y \rightarrow \mathbb{N} \cup \{0\}\) the giving-up guessing function induced by \((g^*, \pi^*_{\text{avg}})\). For any \(\rho > 0\), it holds that

\[
\frac{1}{\rho} \log \mathbb{E}[\bar{G}^*_\text{avg}(X, Y)\rho] = G^*_\text{avg}(X, Y \parallel \rho, \varepsilon),
\]

\[
P(\bar{G}^*_\text{avg}(X, Y) = 0) = \varepsilon. \tag{153}
\]

Proof of Lemma 11: See Appendix M. ■

The left-hand inequality of (146) can be proven by combining Lemma 11 and the unified converse bound stated in Lemma 4. On the other hand, the right-hand inequality of (146) can be proven by the unified achievability bound stated in Lemma 6. See Appendix N for proofs of these lower and upper bounds. ■

Theorem 7 (maximum error criterion). For every \(\rho > 0\) and \(0 < \varepsilon < 1\), it holds that

\[
\bar{H}^*_\alpha(X \mid Y) - \log \left(1 + \frac{1}{\varepsilon} \sup_{y \in Y} H(P_{X,Y}^y)\right) \leq G^*_\text{max}(X, Y \parallel \rho, \varepsilon) \leq \bar{H}^*_\alpha(X \mid Y). \tag{154}
\]

Proof of Theorem 7: For each \(y \in Y\), define two parameters \(J(y)\) and \(\xi(y)\) by

\[
J(y) := \sup \left\{ j \geq 0 \mid \sum_{k=1}^j P_{X,Y}(\xi_y(k) \mid y) < 1 - \varepsilon \right\},
\]

\[
\xi(y) := 1 - \varepsilon - \sum_{k=1}^{J(y)} P_{X,Y}(\xi_y(k) \mid y),
\]

respectively. In contrast to Lemma 11, the following lemma characterizes an optimal guessing strategy under the maximum error criterion.

Lemma 12 (optimal guessing strategy—maximum error criterion). Consider a guessing strategy \((g^*, \pi^*_\text{max})\) given by (150) and

\[
\pi^*_{\text{avg}}(k, y) = \begin{cases} 
0 & \text{if } 1 \leq k \leq J(y), \\
1 - \frac{\xi(y)}{P_{X,Y}(\xi_y(k), y)} & \text{if } k = J(y) + 1, \\
1 & \text{if } J(y) + 2 \leq k < \infty.
\end{cases} \tag{157}
\]

Denote by \(\bar{G}^*_\text{max} : X \times Y \rightarrow \mathbb{N} \cup \{0\}\) the giving-up guessing function induced by \((g^*, \pi^*_\text{max})\). For any \(\rho > 0\), it holds that

\[
\frac{1}{\rho} \log \mathbb{E}[\bar{G}^*_\text{max}(X, Y)\rho] = G^*_\text{max}(X, Y \parallel \rho, \varepsilon), \tag{158}
\]

\[
P(\bar{G}^*_\text{max}(X, Y) = 0 \mid Y = y) = \varepsilon \quad \text{(for all } y \in Y\text{).} \tag{159}
\]

Proof of Lemma 12: See Appendix O. ■

Now, the left-hand inequality of (154) can be proven by combining Lemma 12 and the unified converse bound stated in Lemma 5. On the other hand, the right-hand inequality of (154) can be proven by the unified achievability bound stated in Lemma 6. See Appendix P for proofs of these lower and upper bounds. ■

Remark 10. Kuzuoka provided one-shot bounds [11, Theorems 3 and 4] on the same guessing problem when \(X\) takes values in a finite alphabet. His converse bound [11, Theorem 3] is proven by taking the sum in the right-hand side of (86) over all \(x \in X\) even if \(\varepsilon(x, y) < 1\), and this works only if \(X\) is finite due to the divergence of the harmonic series (see also [5, Section II-A]). In other words, his one-shot bounds [11, Theorems 3 and 4] can be written as

\[
H^*_\alpha(X \mid Y) - \log(1 + |\mathcal{A}|) \leq G^*_\text{avg}(X, Y \parallel \rho, \varepsilon) \leq \bar{H}^*_\alpha(X \mid Y), \tag{160}
\]

provided that \(X\) is supported on a finite subalphabet \(\mathcal{A} \subset X\). On the other hand, our one-shot bounds stated in Theorems 6 and 7 are also applicable to sources \(X\) with countably infinite alphabets \(X\). This holds because our converse bounds are proven by considering the optimal guessing strategies stated in Lemmas 11 and 12 and by restricting the sum in the right-hand side of (86) over all \(x \in X\) satisfying \(\varepsilon(x, y) < 1\).
Using these one-shot bounds, we obtain the following asymptotic expansions of the fundamental limits.

**Corollary 4** (average error criterion). Let $\rho > 0$ and $0 < \varepsilon < 1$ be fixed. If $U(X \mid Y) = 0$, then
\begin{equation}
G_{\text{avg}}^*(X^n, Y^n \parallel \rho, \varepsilon) = n H(X \mid Y) + O(1) \quad (n \to \infty).
\end{equation}

On the other hand, if $U(X \mid Y) > 0$, then
\begin{equation}
G_{\text{avg}}^*(X^n, Y^n \parallel \rho, \varepsilon) = n H(X \mid Y) + O(\sqrt{n}) \quad (n \to \infty).
\end{equation}

**Proof of Corollary 4:** Corollary 4 follows from Theorems 2 and 6.

**Corollary 5** (maximum error criterion). Let $\rho > 0$ and $0 < \varepsilon < 1$ be fixed. If $V(X \mid Y) = 0$, then
\begin{equation}
G_{\text{max}}^*(X^n, Y^n \parallel \rho, \varepsilon) = n H(X \mid Y) + O(1) \quad (n \to \infty).
\end{equation}

On the other hand, if $V(X \mid Y) > 0$ and $\sup_{y \in Y} V(P_X | Y = y) < \infty$, then
\begin{equation}
G_{\text{max}}^*(X^n, Y^n \parallel \rho, \varepsilon) = n H^o(X \mid Y) + O(\sqrt{n}) \quad (n \to \infty).
\end{equation}

**Proof of Corollary 5:** Corollary 5 follows from Theorems 3 and 7.

**Remark 11.** Suppose here that $X$ is supported on some finite sub-alphabet $\mathcal{A} \subseteq X$. Since there is no difference between the average and maximum error criteria in the error-free setting (i.e., $\varepsilon = 0$), we now define
\begin{equation}
G_{\text{error-free}}^*(X, Y \parallel \rho) := G_{\text{avg}}^*(X, Y \parallel \rho, 0) = G_{\text{max}}^*(X, Y \parallel \rho, 0).
\end{equation}

Then, it follows by Arıkan’s seminal result [5] that
\begin{equation}
G_{\text{error-free}}^*(X^n, Y^n \parallel \rho) = n H_\sigma(X \mid Y) + O(1) \quad (n \to \infty).
\end{equation}

Therefore, the differences among the first-order terms of the asymptotic expansions of the three fundamental limits $G_{\text{avg}}^*(X, Y \parallel \rho, \varepsilon)$, $G_{\text{max}}^*(X, Y \parallel \rho, \varepsilon)$, and $G_{\text{error-free}}^*(X^n, Y^n \parallel \rho)$ are characterized by the relations amongst the different conditional entropies as delineated by Proposition 1.

Now, consider guessing problems for the source $X$ in the absence of the side-information $Y$. Formally, when $Y$ is almost surely constant, we denote by
\begin{equation}
G^*(X \parallel \rho, \varepsilon) := G_{\text{avg}}^*(X, Y \parallel \rho, \varepsilon) = G_{\text{max}}^*(X, Y \parallel \rho, \varepsilon).
\end{equation}

our considered fundamental limit here. Then, we can get the following second-order asymptotic result.

**Corollary 6** (in the absence of side-information $Y$). Let $\rho > 0$ and $0 < \varepsilon < 1$ be fixed. If $V(X) = 0$, then
\begin{equation}
G^*(X^n \parallel \rho, \varepsilon) = n H(X) + O(1) \quad (n \to \infty).
\end{equation}

On the other hand, if $0 < V(X) < \infty$, then
\begin{equation}
G^*(X^n \parallel \rho, \varepsilon) = n H(X) + O(\sqrt{n}) \quad (n \to \infty).
\end{equation}

In particular, if $V(X) > 0$ and $T(X) < \infty$, then
\begin{equation}
G^*(X^n \parallel \rho, \varepsilon) = n H(X) - \sqrt{n V(X) \Phi^{-1}(\varepsilon)} + O(\log n) \quad (n \to \infty).
\end{equation}

**Proof of Corollary 6:** Corollary 6 follows from Theorems 1 and 6.

**D. Encoding Tasks**

Suppose that $X$ is a countably infinite set of tasks we wish to execute. Given a correlated source $(X, Y)$ and a positive integer $M$, let us consider assigning a randomly occurred task $X$ into $M$ messages with the help of some side-information $Y$ of $X$. Bunte and Lapidoth [6] proposed this problem and assumed that none of the tasks are ignored. In this case, we can think of such an assignment as a finite partition of $X$ in which the cardinality of the partition does not exceed the desired threshold $M$.

In this study, under certain error constraints, we allow the possibility of ignoring some tasks. Given a deterministic map $f : X \times Y \to \{0, 1, 2, \ldots, M\}$ called an assignment function, consider the following assignment rule: If $f(X, Y) = m$ for some $1 \leq m \leq M$, then a task $X$ is assigned to a message $m$. On the other hand, a task $X$ is ignored if and only if $f(X, Y) = 0$. Define the deterministic map $\mathcal{L} : \{0, 1, 2, \ldots, M\} \times Y \to 2^X$ by
\begin{equation}
\mathcal{L}(m, y) := \begin{cases} 
0 & \text{if } m = 0, \\
\{x \in X \mid f(x, y) = m\} & \text{if } m = 1, \ldots, M
\end{cases}
\end{equation}
for each $1 \leq m \leq M$ and $y \in \mathcal{Y}$. Then, the family $\{\mathcal{L}(m, y)\}_{m=0}^{M}$ forms a sub-partition$^3$ of $\mathcal{X}$ for each $y \in \mathcal{Y}$. When a task $X$ occurs, all tasks in $\mathcal{L}(f(X, Y), Y)$ are executed. Namely, if $f(X, Y) = 0$, then no task in $X$ is executed. Furthermore, we allow the probability of not executing any task even if $f(X, Y) \neq 0$, but this occurs with a certain fixed probability. More precisely, given a stochastic map $\mathbb{E} : 2^X \times \mathcal{Y} \rightarrow 2^X$ satisfying

$$\mathbb{P}\{\mathbb{E}(\mathcal{A}, y) = \emptyset\} + \mathbb{P}\{\mathbb{E}(\mathcal{A}, y) = \mathcal{A}\} = 1$$

(172)

for each $\mathcal{A} \subset \mathcal{X}$ and $y \in \mathcal{Y}$, define the stochastic map $\mathcal{L} : \{0, 1, 2, \ldots, M\} \times \mathcal{Y} \rightarrow 2^X$ by

$$\mathcal{L}(f(x, y), y) := \mathbb{E}(\mathcal{L}(f(x, y), y), y)$$

(173)

for each $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Then, an error occurs if and only if $X \notin \mathcal{L}(f(X, Y), Y)$. For this stochastic sub-partition $\mathcal{L}$ induced by the pair $(f, \mathbb{E})$, we aim to minimize the task sub-partitioning $\rho$-th moment $\mathbb{E}[|\mathcal{L}(f(X, Y), Y)|^\rho]$ for a fixed $\rho > 0$ subject to certain error constraints. In other words, we are interested to characterize fundamental limits defined as the infimum of positive real numbers $N$ such that there exist an assignment function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1, 2, \ldots, M\}$ and a stochastic map $\mathbb{E} : 2^X \times \mathcal{Y} \rightarrow 2^X$ satisfying

$$\log \mathbb{E}[|\mathcal{L}(f(X, Y), Y)|^\rho] \leq \rho N$$

(174)

under a certain constraint on error probabilities. We now introduce two error formalisms as follows:

**Definition 8** (average error criterion). A $(\rho, M, N, \varepsilon)_{\text{avg}}$-assignment for a correlated source $(X, Y)$ is a pair $(f, \mathbb{E})$ consisting of an assignment function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1, 2, \ldots, M\}$ and a stochastic map $\mathbb{E} : 2^X \times \mathcal{Y} \rightarrow 2^X$ satisfying (172) and (174) hold, and

$$\mathbb{P}\{X \notin \mathcal{L}(f(X, Y), Y)\} \leq \varepsilon$$

(175)

**Definition 9** (maximum error criterion). A $(\rho, M, N, \varepsilon)_{\text{max}}$-assignment for a correlated source $(X, Y)$ is a pair $(f, \mathbb{E})$ consisting of an assignment function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1, 2, \ldots, M\}$ and a stochastic map $\mathbb{E} : 2^X \times \mathcal{Y} \rightarrow 2^X$ satisfying (172) and (174) hold, and

$$\sup_{y \in \mathcal{Y}} \mathbb{P}\{X \notin \mathcal{L}(f(X, Y), Y) \mid Y = y\} \leq \varepsilon$$

(176)

Given a positive integer $M$ and two real numbers $\rho > 0$ and $0 < \varepsilon < 1$, consider the following two fundamental limits:

$$L^*_{\text{avg}}(X, Y, M \parallel \rho, \varepsilon) := \inf\{N > 0 \mid \text{there exists a } (\rho, M, N, \varepsilon)_{\text{avg}}\text{-assignment for } (X, Y)\},$$

(177)

$$L^*_{\text{max}}(X, Y, M \parallel \rho, \varepsilon) := \inf\{N > 0 \mid \text{there exists a } (\rho, M, N, \varepsilon)_{\text{max}}\text{-assignment for } (X, Y)\}. $$

(178)

We now state the following one-shot bounds.

**Theorem 8** (average error criterion). Let $\rho > 0$, $0 < \varepsilon < 1$, and $M \geq 1$ be fixed. It holds that

$$L^*_{\text{avg}}(X, Y, M \parallel \rho, \varepsilon) \geq H^*_{\alpha}(X \mid Y) - \log M.$$  

(179)

Moreover, if $M > 2 + H(X \mid Y)/\varepsilon$, then

$$L^*_{\text{avg}}(X, Y, M \parallel \rho, \varepsilon) \leq H^*_{\alpha}(X \mid Y) - \log \left(\frac{\varepsilon(M - 2) - H(X \mid Y)}{4\varepsilon + 4H(X \mid Y)}\right) + \frac{1}{\rho} \log 2,$$

(180)

where $|u|_{\alpha} := \max\{0, u\}$ for $u \in \mathbb{R}$.

**Proof of Theorem 8:** The converse bound stated in (179) is proven in Appendix Q via the unified converse bound stated in Lemma 4. In the following, we shall prove the achievability bound stated in (180).

Recall that the numbers $J$ and $\xi$ are defined in (148) and (149), respectively. In addition, define the number

$$\nu := \frac{\xi}{\sum_{b \in \mathcal{Y}} P_{X, Y}(\xi_b(J + 1), b)}$$

(181)

where the bijection $\varsigma_y : \mathbb{N} \rightarrow \mathcal{X}$ is defined to satisfy (147) for each $y \in \mathcal{Y}$. We now introduce yet another definition of the $\varepsilon$-smooth conditional Rényi entropy as

$$\tilde{H}^*_{\alpha}(X \mid Y) := \frac{\alpha}{1 - \alpha} \log \left(\sum_{y \in \mathcal{Y}} \left(\sum_{k=1}^{J} P_{X, Y}(\varsigma_y(k), y)^{\alpha} + \nu^\alpha P_{X, Y}(\varsigma_y(J + 1), y)^{\alpha}\right)^{1/\alpha}\right).$$

(182)

Instead of establishing (180), we establish the following one-shot achievability bound, which serves as an intermediate result in proving (180).

\footnote{A sub-partition of a set $\mathcal{S}$ is a subset of a partition of the set $\mathcal{S}$.}
Lemma 13. Suppose that the integer $M$ is large enough so that
\[ M > 2 + \frac{H(X \mid Y)}{\varepsilon}. \] (183)

Then, it holds that
\[ \mathbf{L}_{\text{avg}}^*(X, Y, M \parallel \rho, \varepsilon) \leq \left| \tilde{H}^e_\alpha(X \mid Y) - \log \left( \frac{\varepsilon(M - 2) - H(X \mid Y)}{4\varepsilon} \right) \right|_+ + \frac{1}{\rho} \log 2. \] (184)

Proof of Lemma 13: See Appendix R.

The following lemma provides inequalities between $H^e_\alpha(X \mid Y)$ and $\tilde{H}^e_\alpha(X \mid Y)$.

Lemma 14. For any $0 < \alpha < 1$ and $0 \leq \varepsilon < 1$, it holds that
\[ H^e_\alpha(X \mid Y) \leq \tilde{H}^e_\alpha(X \mid Y) \leq H^e_\alpha(X \mid Y) + \log \left( 1 + \frac{H(X \mid Y)}{\varepsilon} \right). \] (185)

Proof of Lemma 14: See Appendix S.

Remark 12. By Lemma 14, we observe that the asymptotic expansions of $H^e_\alpha(X \mid Y)$ and $\tilde{H}^e_\alpha(X^n \mid Y^n)$ are the same up to the reminder term +O(log $n$), provided that $H(X \mid Y) < \infty$.

Combining Lemmas 13 and 14, we have the achievability bound stated in (180). This completes the proof of Theorem 8.

Theorem 9 (maximum error criterion). Let $\rho > 0$, $0 < \varepsilon < 1$, and $M \geq 1$ be fixed. It holds that
\[ \mathbf{L}_{\text{max}}^*(X, Y, M \parallel \rho, \varepsilon) \geq \tilde{H}^e_\alpha(X \mid Y) - \log M. \] (186)

Moreover, if $M > 2 + \sup_{y \in Y} H(P_X[y]) / \varepsilon$, then
\[ \mathbf{L}_{\text{max}}^*(X, Y, M \parallel \rho, \varepsilon) \leq \left| \tilde{H}^e_\alpha(X \mid Y) - \log \left( \frac{\varepsilon(M - 2) - \sup_{y \in Y} H(P_X[y])}{4\varepsilon} \right) \right|_+ + \frac{1}{\rho} \log 2. \] (187)

Proof of Theorem 9: The converse and achievability bounds stated in (186) and (187), respectively, are proven in Appendices T-A and T-B, respectively. These utilize the unified approaches stated in Lemmas 5 and 6, respectively. This completes the proof of Theorem 9.

Remark 13. Suppose that there exists a $y \in Y$ such that the support set $\{ x \in X \mid P(X \mid y) > 0 \}$ is infinite. Since every finite partition of an infinite set contains an infinite subset, it is clear that
\[ \mathbf{L}_{\text{avg}}^*(X, Y, M \parallel \rho, 0) = \mathbf{L}_{\text{avg}}^*(X, Y, M \parallel \rho, 0) = \infty \] (188)
in the zero-error setting (i.e., $\varepsilon = 0$). However, if $0 < \varepsilon < 1$, then Theorems 8 and 9 state that these fundamental limits can be finite, and the task encoding problem can be considered over a countably infinite alphabet $X$ when we ignore some tasks.

Consider a sequence $\{ M_n \}_{n=1}^\infty$ of positive integers satisfying
\[ \lim_{n \to \infty} \frac{M_n}{n} = \infty. \] (189)

Define
\[ \tau_n := \log M_n \] (190)
for each $n \geq 1$. Using the above one-shot bounds, we obtain the following asymptotic results.

Corollary 7 (average error criterion). Let $\rho > 0$ and $0 < \varepsilon < 1$ be fixed. If $0 < U(X \mid Y) < \infty$, then
\[ \mathbf{L}_{\text{avg}}^*(X^n, Y^n, M_n \parallel \rho, \varepsilon) = \left| n H(X \mid Y) - \tau_n \right|_+ + O(\sqrt{n}) \quad \text{as } n \to \infty. \] (191)

Proof of Corollary 7: Corollary 7 follows from Theorems 2 and 8.

Corollary 8 (maximum error criterion). Let $\rho > 0$ and $0 < \varepsilon < 1$ be fixed. If $V(X \mid Y) > 0$ and $\sup_{y \in Y} V(P_X[y]) < \infty$, then
\[ \mathbf{L}_{\text{max}}^*(X^n, Y^n, M_n \parallel \rho, \varepsilon) = \left| n H^{(\alpha)}(X \mid Y) - \tau_n \right|_+ + O(\sqrt{n}) \quad \text{as } n \to \infty. \] (192)

Proof of Corollary 8: Corollary 8 follows from Theorems 3 and 9.

Remark 14. Suppose here that $X$ is supported on some finite sub-alphabet $\mathcal{A} \subseteq X$. Since there is no difference between the average and maximum error criteria in the error-free setting (i.e., $\varepsilon = 0$), we now define
\[ \mathbf{L}_{\text{error-free}}^*(X, Y \parallel \rho) := \mathbf{L}_{\text{avg}}^*(X, Y \parallel \rho, 0) = \mathbf{L}_{\text{max}}^*(X, Y \parallel \rho, 0). \] (193)
To this fundamental limit, Bunte and Lapidoth [6] proved that

\[ L_{\text{error-free}}^*(X^n, Y^n \parallel \rho) = \left| n H_\alpha(X \mid Y) - \tau_n \right|_+ + O(1) \quad \text{(as } n \to \infty) \]  

Therefore, the differences among the first-order terms of asymptotic expansions of the three fundamental limits \( L_{\text{avg}}^*(X, Y \parallel \rho, \varepsilon) \), \( L_{\text{max}}^*(X, Y \parallel \rho, \varepsilon) \), and \( L_{\text{error-free}}^*(X^n, Y^n \parallel \rho) \) can be characterized by Proposition 1.

Now, consider task encoding problems for the source \( X \) in the absence of the side-information \( Y \). Formally, when \( Y \) is almost surely constant, we denote by

\[ L^*(X, M \parallel \rho, \varepsilon) := L_{\text{avg}}^*(X, Y, M \parallel \rho, \varepsilon) = L_{\text{max}}^*(X, Y, M \parallel \rho, \varepsilon) \]  

our considered fundamental limit here. Then, we obtain the following third-order asymptotic result.

**Corollary 9** (in the absence of side-information \( Y \)). Let \( \rho > 0 \) and \( 0 < \varepsilon < 1 \) be fixed. If \( V(X) = 0 \), then

\[ L^*(X^n, M_n \parallel \rho, \varepsilon) = \left| n H(X) - \tau_n \right|_+ + O(1) \quad \text{(as } n \to \infty) \]  

On the other hand, if \( 0 < V(X) < \infty \), then

\[ L^*(X^n, M_n \parallel \rho, \varepsilon) = \left| n H(X) - \tau_n \right|_+ + O(\sqrt{n}) \quad \text{(as } n \to \infty) \]  

In particular, if \( V(X) > 0 \) and \( T(X) < \infty \), then

\[ L^*(X^n, M_n \parallel \rho, \varepsilon) = \left| n H(X) - \tau_n - \sqrt{n V(X)} \Phi^{-1}(\varepsilon) - \frac{1 + \rho}{2 \rho} \log n \right|_+ + O(1) \quad \text{(as } n \to \infty) \]  

**Proof of Corollary 9:** Corollary 9 follows from Theorems 1 and 9.

### IV. Concluding Remarks

We characterized asymptotic expansions of the unconditional and two versions of the smooth conditional Rényi entropies, and derived fundamental limits of several information-theoretic problems [2]–[6] as applications of the asymptotic expansions. Specifically, we compared the third-order asymptotic analyses for the classical fixed-to-fixed source coding and Campbell’s source coding problems allowing errors, and showed in (139) that the difference between these two asymptotic expansions are manifested in their third-order terms. In contrast to traditional results [3], [5], [6], [11], [15] requiring the assumption of finite alphabets, due to the fact that we allow errors in the various problems we study, our results on guessing and task encoding problems are applicable to sources \( \mathcal{X} \) defined over countably infinite alphabets \( \mathcal{X} \).

In [16], the present authors considered the following limiting case of the cumulant generating function of codeword lengths:

\[ \lim_{\rho \to 0^+} \frac{1}{\rho} \log \mathbb{E}[2^{n\ell(F(X,Y))}] = \mathbb{E}[\ell(F(X,Y))], \]  

without prefix-free constraints. We [16] then showed that the optimal first-order coding rates (i.e., the \( n \) scale) are the same under both average and maximum error criteria, and the optimal second-order coding rates (i.e., the \( \sqrt{n} \) scale) differ under these two error formalisms. This difference is characterized by the law of total variance (see (73)). On the other hand, in this study, Corollaries 1 and 2 state that the optimal first-order coding rates differ under these two error formalisms, and this difference can also be characterized by the law of total variance (see Proposition 1).

In Theorems 4–9, we provided one-shot coding theorems in various information-theoretic problems, and these are formulated by two conditional versions of smooth Rényi entropies \( H_\alpha^n(X^n \mid Y^n) \) and \( \tilde{H}_\alpha^n(X^n \mid Y^n) \). Hence, further asymptotic analyses of \( H_\alpha^n(X^n \mid Y^n) \) and \( \tilde{H}_\alpha^n(X^n \mid Y^n) \) would yield further asymptotic results on the problems. While the exact second- and third-order terms of the unconditional version of the smooth Rényi entropy \( H_\alpha^n(X^n) \) were derived in Theorem 1, we showed in Theorems 2 and 3 the exact first-order terms of \( H_\alpha^n(X^n \mid Y^n) \) and \( \tilde{H}_\alpha^n(X^n \mid Y^n) \), respectively, and that both remainder terms scale as +O(\( \sqrt{n} \)) due to Chebyshev’s inequality. Namely, finding the coefficients of the second- and third-order terms of \( H_\alpha^n(X^n \mid Y^n) \) and \( \tilde{H}_\alpha^n(X^n \mid Y^n) \) remain open problems. As explained in Remark 6, Kuzuoka [11, Theorem 2] provided the first-order term of \( H_\alpha^n(X^n \mid Y^n) \) when \( (X^n, Y^n) \) is a mixture of i.i.d. sources, and his result can be straightforwardly extended to the source \( (X^n, Y^n) \) satisfying the AEP, e.g., a mixture of stationary and ergodic sources (see [11, Remark 2]). General formulas of the two conditional versions \( H_\alpha^n(X^n \mid Y^n) \) and \( \tilde{H}_\alpha^n(X^n \mid Y^n) \) for a general source \( (X, Y) = \{(X^n, Y^n)\}_{n=1}^\infty \) remains open problems as well. Finally, in this study, we only considered the smooth Rényi entropies in the case where \( 0 < \alpha < 1 \). Asymptotic expansions and operational interpretations of the smooth Rényi entropies with the order \( 1 < \alpha < \infty \) are of interest in future works.
APPENDIX A
PROOF OF LEMMA 2

It follows by the definition of $\langle \rangle_e$ stated in (23) that

$$
\mathbb{E} \left[ \exp \left( \sum_{i=1}^{n} Z_i \right) \right] = \mathbb{E} \left[ \exp \left( \sum_{i=1}^{n} Z_i \right) 1_{\{\sum_{i=1}^{n} Z_i < \delta \eta_n \}} \right] + \alpha_n e^{\delta \eta_n} \mathbb{P} \left\{ \sum_{i=1}^{n} Z_i = \eta_n \right\},
$$

where two real parameters $\eta_n \in \mathbb{R}$ and $0 \leq \alpha_n < 1$ are chosen so that

$$
\mathbb{P} \left\{ \sum_{i=1}^{n} Z_i > \eta_n \right\} + \alpha_n \mathbb{P} \left\{ \sum_{i=1}^{n} Z_i = \eta_n \right\} = \varepsilon. \tag{201}
$$

Since we have assumed that there exist two positive constants $c_1$ and $c_2$ satisfying $V_n > n c_1$ and $T_n < c_2 V_n$ for sufficiently large $n \geq n_0$, it can be verified by the Berry–Esseen theorem (see, e.g., [32, Theorem 2 in Chapter XVI.5]) and Taylor’s theorem for the map $\Phi^{-1} : (0,1) \to \mathbb{R}$ that there exists a positive constant $c_3$ depending only on $0 < \varepsilon < 1$ such that

$$
s E_n + s \sqrt{V_n} \Phi^{-1}(1-\varepsilon) - c_2 c_3 s \leq \eta_n \leq s E_n + s \sqrt{V_n} \Phi^{-1}(1-\varepsilon) + c_2 c_3 s \tag{202}
$$

for all $n \geq n_0$. Now, choose an integer $n_1 \geq n_0$ so that

$$
c_2 c_3 + s \leq s \sqrt{n c_1} - s \log n \tag{203}
$$

for all $n \geq n_1$. In addition, fix a real number $\gamma$ so that

$$
\gamma \geq s \left( 12 c_2 + \frac{1}{\sqrt{c_1}} \right) \sqrt{2\pi e^{(\Phi^{-1}(1-\varepsilon)+1)^2/2}}. \tag{204}
$$

Then, we observe that

$$
\mathbb{E} \left[ \exp \left( \sum_{i=1}^{n} Z_i \right) 1_{\{\sum_{i=1}^{n} Z_i < \delta \eta_n \}} \right] \geq \sum_{k=1}^{\infty} 2^{\eta_n \gamma} \mathbb{P} \left\{ \sum_{i=1}^{n} Z_i < \eta_n - (k-1) \gamma \right\} \geq \frac{1}{\sqrt{2\pi}} \int_{[\eta_n - s E_n - (k-1) \gamma] / (s \sqrt{V_n})}^{[\eta_n - s E_n] / (s \sqrt{V_n})} e^{-t^2/2} dt - \frac{12 T_n}{V_n^{3/2}}.
$$

for sufficiently large $n \geq n_1$, where

- (a) follows from the Berry–Esseen theorem (see, e.g., [32, Theorem 2 in Chapter XVI.5]),
• (b) follows from (202), (203), and the fact that \( t \mapsto e^{-t^2/2} \) is quasiconcave in \( t \in \mathbb{R} \),
• (c) follows from the hypothesis that \( T_n < c_2 V_n \) for sufficiently large \( n \geq n_1 \),
• (d) follows from the hypothesis that \( V_n < n/c_1 \) for sufficiently large \( n \geq n_1 \), and
• (e) follows from the choice of \( \gamma \) stated in (204).

Therefore, we have

\[
\frac{1}{s} \log \mathbb{E} \left[ \exp \left( \sum_{i=1}^{n} Z_i \right) \right] \geq (a) \frac{1}{s} \log \mathbb{E} \left[ \exp \left( \sum_{i=1}^{n} Z_i \right) 1_{\left\{ s \sum_{i=1}^{n} Z_i \leq \eta_n \right\}} \right] \geq (b) \frac{1}{s} \log \left( \frac{2^{\eta_n} (1 - n^{-\gamma})}{2 \sqrt{n}} \right) = \frac{1}{s} \left( \eta_n - \gamma + \log(1 - n^{-\gamma}) - \frac{1}{2} \log n \right) \geq (c) E_n + \sqrt{V_n} \Phi^{-1}(1 - \varepsilon) - \frac{1}{2s} \log n - c_2 c_3 + \frac{\gamma}{s} + \frac{1}{s} \log(1 - n^{-\gamma}) \tag{206}
\]

for sufficiently large \( n \geq n_1 \), where

• (a) follows from (200),
• (b) follows from (205), and
• (c) follows from (202).

On the other hand, it can be verified by the same way as [14, Lemma 47] that

\[
\mathbb{E} \left[ \exp \left( t \sum_{i=1}^{n} Z_i \right) 1_{\left\{ t \sum_{i=1}^{n} Z_i \leq \eta_n \right\}} \right] \leq 2^{\eta_n + 1} \left( \frac{1}{s \sqrt{2\pi}} + 12 c_2 \right) \frac{1}{\sqrt{n c_1}} \tag{207}
\]

for sufficiently large \( n \geq n_0 \). Thus, we obtain

\[
\frac{1}{s} \log \mathbb{E} \left[ \exp \left( \sum_{i=1}^{n} Z_i \right) \right] \leq (a) \frac{1}{s} \log \mathbb{E} \left[ \exp \left( \sum_{i=1}^{n} Z_i \right) 1_{\left\{ s \sum_{i=1}^{n} Z_i \leq \eta_n \right\}} \right] \leq (b) \frac{1}{s} \log \left( \frac{2^{\eta_n + 1} \left( \frac{1}{s \sqrt{2\pi}} + 12 c_2 \right)}{\sqrt{n c_1}} \right) = \frac{1}{s} \left( \eta_n + 1 + \log \left( \frac{1}{s \sqrt{2\pi}} + 12 c_2 \right) - \frac{1}{2} \log n - \frac{1}{2} \log c_1 \right) \geq (c) E_n + \sqrt{V_n} \Phi^{-1}(1 - \varepsilon) - \frac{1}{2s} \log n + c_2 c_3 + \frac{1}{s} + \frac{1}{s} \log \left( \frac{1}{s \sqrt{2\pi}} + 12 c_2 \right) - \frac{1}{2s} \log c_1 \tag{208}
\]

for sufficiently large \( n \geq n_0 \), where

• (a) follows from (200),
• (b) follows from (207), and
• (c) follows from (202).

The proof of Lemma 2 is now completed by combining (206) and (208).

\[\blacksquare\]

APPENDIX B

PROOF OF THEOREM 2—ZERO VARIANCE \( U(X \mid Y) = 0 \)

A. Proof of (59) When \( H(X \mid Y) = 0 \)

Since we have assumed that \( U(X \mid Y) = 0 \), it is clear that

\[
P_{X\mid Y}(x \mid y) = \begin{cases} \exp \left( - H(X \mid Y) \right) & \text{if } P_{X\mid Y}(x \mid y) > 0, \\ 0 & \text{if } P_{X\mid Y}(x \mid y) = 0 \end{cases} \tag{209}
\]

for every \((x, y) \in X \times \mathcal{Y}\. Consider the case in which \( H(X \mid Y) = 0 \). Then, it follows from (6), (7), and (209) that

\[
\mathcal{R}_{X\mid Y}^{\alpha(y)}(y) = \emptyset \tag{210}
\]

for every \( y \in \mathcal{Y}\. Hence, it follows from (61) and (210) that

\[
H_{\alpha}^c(X^n \mid Y^n) = \frac{\alpha}{1 - \alpha} \log \left( \inf_{\delta_n(\cdot)} \sum_{y \in \mathcal{Y}^n} P_{Y^n}(y) (1 - \delta_n(y)) \right) = \frac{\alpha}{1 - \alpha} \log(1 - \varepsilon) \tag{211}
\]
for every \( n \geq 1 \), where the infimum is taken over the mappings \( \delta_n : \mathcal{Y}^n \to [0, 1] \) satisfying
\[
\sum_{y \in \mathcal{Y}^n} P_{Y^n}(y) \delta_n(y) = \varepsilon. \tag{212}
\]
Therefore, the asymptotic expansion in (59) holds if \( H(X \mid Y) = 0 \), completing the proof.

\[ \]

B. Proof of Upper Bound Part of (59) When \( H(X \mid Y) > 0 \)

For each \( n \geq 1 \) and \( y \in \mathcal{Y}^n \), define the sub-probability distribution \( \tilde{Q}_{X^n | Y^n}(\cdot \mid y) \) on \( X^n \) as
\[
\tilde{Q}_{X^n | Y^n}(x \mid y) = (1 - \varepsilon) P_{X^n | Y^n}(x \mid y).
\]
Moreover, define the joint sub-probability distribution \( \tilde{Q}_{X^n, Y^n} \) on \( X^n \times \mathcal{Y}^n \) as
\[
\tilde{Q}_{X^n, Y^n}(x, y) = \sum_{y' \in \mathcal{Y}^n} P_{Y^n}(y) \tilde{Q}_{X^n | Y^n}(x \mid y).
\]

Then, we observe that
\[
H^\alpha_\varepsilon(X^n \mid Y^n) = \frac{\alpha}{1 - \alpha} \log \left( \frac{\tilde{Q}_{Y^n}(y)}{\tilde{Q}_{X^n | Y^n}(y)} \right) \]
\[
= \frac{\alpha}{1 - \alpha} \log \left( \frac{\sum_{y' \in \mathcal{Y}^n} P_{Y^n}(y') \tilde{Q}_{X^n | Y^n}(x, y')^{1/\alpha}}{\sum_{y' \in \mathcal{Y}^n} P_{Y^n}(y) \tilde{Q}_{X^n | Y^n}(x, y')^{1/\alpha}} \right) \]
\[
= \frac{\alpha}{1 - \alpha} \log \left( \frac{\sum_{y' \in \mathcal{Y}^n} P_{Y^n}(y') \tilde{Q}_{X^n | Y^n}(x, y')^{1/\alpha}}{\sum_{y' \in \mathcal{Y}^n} P_{Y^n}(y) \tilde{Q}_{X^n | Y^n}(x, y')^{1/\alpha}} \right) + \frac{\alpha}{1 - \alpha} \log(1 - \varepsilon)
\]
\[ \]
where
- (a) follows from the fact that \( \tilde{Q}_{X^n, Y^n} \) belongs to \( \mathcal{B}^\varepsilon_{X^n \times Y^n}(P_{X^n, Y^n}) \),
- (b) follows from (214),
- (c) follows from (213), and
- (d) follows from (209).

This completes the proof of the upper bound part of (59).

C. Proof of Lower Bound Part of (59) When \( H(X \mid Y) > 0 \)

Consider a mapping \( \delta_n : \mathcal{Y}^n \to [0, 1] \) satisfying (212). For the sake of brevity, we write
\[
\mathcal{A}_n(y) = \mathcal{A}^{\delta_n(y)}_{X^n | Y^n}(y) \tag{216}
\]
for each \( y \in \mathcal{Y}^n \); see (6) and (7) for the definition of the right-hand side of (216) by replacing \( \varepsilon \) and \( P_X(\cdot) \) by \( \delta_n(y) \) and \( P_{X^n | Y^n}(\cdot \mid y) \), respectively. Moreover, define
\[
\tilde{\mathcal{A}}_n(y) = \mathcal{A}_n(y) \cup \{ x^*(y) \}, \tag{217}
\]
where \( x^*(y) \in X^n \) is chosen so that
\[
x^*(y) \in \arg \max_{x \in X^n \setminus \mathcal{A}_n(y)} P_{X^n | Y^n}(x \mid y). \tag{218}
\]

Furthermore, define the subset \( \mathcal{U}_n \) of \( \mathcal{Y}^n \) by
\[
\mathcal{U}_n := \left\{ y \mid P_{X^n | Y^n}(\tilde{\mathcal{A}}_n(y) \mid y) \geq \frac{1 - \varepsilon}{2} \right\}. \tag{219}
\]
Note that both $\mathcal{A}_n(\cdot)$ and $\mathcal{U}_n$ depend on $\delta_n(\cdot)$. A direct calculation shows

$$1 - \varepsilon \equiv \sum_{y \in \mathcal{Y}_n} P_{Y^n}(y) (1 - \delta_n(y))$$

$$\leq \sum_{y \in \mathcal{Y}_n} P_{Y^n}(y) P_{X^n|Y^n}(\mathcal{A}_n(y) \mid y)$$

$$\leq \sum_{y \in \mathcal{U}_n} P_{Y^n}(y) + \frac{1 - \varepsilon}{2} \sum_{y \in \mathcal{Y}_n \setminus \mathcal{U}_n} P_{Y^n}(y)$$

$$= P_{Y^n}(\mathcal{U}_n) \left(1 + \frac{1 - \varepsilon}{2}\right) + \frac{1 - \varepsilon}{2},$$

where

- (a) follows from (212),
- (b) follows from the right-hand inequality of (7), and
- (c) follows by the definition of $\mathcal{U}_n$ stated in (219).

Therefore, we get

$$P_{Y^n}(\mathcal{U}_n) \geq \frac{1 - \varepsilon}{1 + \varepsilon}. \quad (220)$$

In addition, we see that

$$\frac{1 - \varepsilon}{2} \leq P_{X^n|Y^n}(\mathcal{A}_n(y) \mid y)$$

$$\equiv |\mathcal{A}_n(y)| \exp \left(-n H(X \mid Y)\right) \quad (222)$$

for each $y \in \mathcal{U}_n$, where

- (a) follows by the definition of $\mathcal{U}_n$ stated in (219), and
- (b) follows from (209).

Furthermore, we have

$$\sum_{x \in \mathcal{A}_n(y)} P_{X^n|Y^n}(x \mid y)^\alpha \equiv |\mathcal{A}_n(y)| \exp \left(-\alpha n H(X \mid Y)\right)$$

$$\geq \frac{1 - \varepsilon}{2} \exp \left((1 - \alpha) n H(X \mid Y)\right) \quad (223)$$

for every $y \in \mathcal{U}_n$, where

- (a) follows from (209), and
- (b) follows from (222).

Now, noting that $H(X \mid Y) > 0$, choose an integer $n_0$ by

$$n_0 = n_0(\alpha, \varepsilon, H(X \mid Y)) := \left[\frac{1}{(1 - \alpha) H(X \mid Y) \log \frac{4}{1 - \varepsilon}}\right], \quad (224)$$

where $[u] := \min\{z \in \mathbb{Z} \mid z \geq u\}$ stands for the ceiling function. Then, we observe that

$$H_{\varepsilon}^\alpha(X^n \mid Y^n) \equiv \frac{\alpha}{1 - \alpha} \log \left(\inf_{\delta_n(\cdot)} \sum_{y \in \mathcal{Y}_n} P_{Y^n}(y) \left(\sum_{x \in \mathcal{A}_n(y)} P_{X^n|Y^n}(x \mid y)^\alpha + (1 - \delta_n(y) - P_{X^n|Y^n}(\mathcal{A}_n(y) \mid y))^\alpha\right)\right)^{1/\alpha}$$

$$\geq \frac{\alpha}{1 - \alpha} \log \left(\inf_{\delta_n(\cdot)} \sum_{y \in \mathcal{U}_n} P_{Y^n}(y) \left(\sum_{x \in \mathcal{A}_n(y)} P_{X^n|Y^n}(x \mid y)^\alpha - 1\right)\right)^{1/\alpha}$$

$$\geq \frac{\alpha}{1 - \alpha} \log \left(\inf_{\delta_n(\cdot)} \sum_{y \in \mathcal{U}_n} P_{Y^n}(y) \left(\frac{1 - \varepsilon}{2} \exp \left((1 - \alpha) n H(X \mid Y)\right) - 1\right)^{1/\alpha}\right)$$

$$\geq \frac{\alpha}{1 - \alpha} \log \left(\inf_{\delta_n(\cdot)} \sum_{y \in \mathcal{U}_n} P_{Y^n}(y) \left(\frac{1 - \varepsilon}{4} \exp \left(\frac{1 - \alpha}{\alpha} n H(X \mid Y)\right)\right)^{1/\alpha}\right)$$

Inequality (221) is indeed a reverse Markov inequality (cf. [19, Lemma 5.6.1]).
\[ = n H(X \mid Y) + \frac{1}{1 - \alpha} \log \frac{1 - \varepsilon}{4} + \frac{\alpha}{1 - \alpha} \log \left( \inf_{\delta_0} P_{Y^n}(\mathcal{U}_n) \right) \]

\[
\geq n H(X \mid Y) + \frac{1}{1 - \alpha} \log \frac{1 - \varepsilon}{4} + \frac{\alpha}{1 - \alpha} \log \frac{1 - \varepsilon}{1 + \varepsilon} \tag{225}
\]

for sufficiently large \( n \geq n_0 \), where

- (a) follows from (61),
- (b) follows from the right-hand inequality of (7), i.e.,
\[
\left( 1 - \delta_n(y) - P_{X^n|Y^n}(\mathcal{A}_n(y) \mid y) \right) \geq P_{X^n|Y^n}(x^*(y) \mid y)^\alpha - 1, \tag{226}
\]

- (c) follows from (223),
- (d) follows by the choice of \( n_0 \) stated in (224), and
- (e) follows from (221).

This completes the proof of the lower bound part of (59). \( \blacksquare \)

**APPENDIX C**

**Proof of Theorem 2—Positive Variance \( 0 < U(X \mid Y) < \infty \)**

**A. Proof of Upper Bound Part of (60)**

For each positive integer \( n \), define a subset \( \mathcal{T}_e^{(n)} \) of \( \mathcal{X}^n \times \mathcal{Y}^n \) as

\[
\mathcal{T}_e^{(n)} := \{ (x, y) \mid \log \frac{1}{P_{X^n|Y^n}(x \mid y)} - n H(X \mid Y) \leq \sqrt{\frac{n U(X \mid Y)}{\varepsilon}} \}. \tag{227}
\]

In addition, define

\[
\mathcal{T}_e^{(n)}(y) := \{ x \mid (x, y) \in \mathcal{T}_e^{(n)} \}. \tag{228}
\]

Note that for every \( (x, y) \in \mathcal{T}_e^{(n)} \), it holds that

\[
\exp \left( -n H(X \mid Y) - \sqrt{\frac{n U(X \mid Y)}{\varepsilon}} \right) \leq P_{X^n|Y^n}(x \mid y) \leq \exp \left( -n H(X \mid Y) + \sqrt{\frac{n U(X \mid Y)}{\varepsilon}} \right). \tag{229}
\]

Since \( 0 < U(X \mid Y) < \infty \), it follows from Chebyshev’s inequality that the \( P_{X^n,Y^n} \)-probability of \( \mathcal{T}_e^{(n)} \) is bounded from below as

\[
P_{X^n,Y^n}(\mathcal{T}_e^{(n)}) = 1 - \mathbb{P}\{ (X^n, Y^n) \not\in \mathcal{T}_e^{(n)} \} \geq 1 - \varepsilon. \tag{230}
\]

On the other hand, we see that

\[
1 \geq \sum_{x \in \mathcal{X}^n : (x, y) \in \mathcal{T}_e^{(n)}} P_{X^n|Y^n}(x \mid y) \geq \sum_{x \in \mathcal{X}^n : (x, y) \in \mathcal{T}_e^{(n)}} \exp \left[ -n H(X \mid Y) - \sqrt{\frac{n U(X \mid Y)}{\varepsilon}} \right] = |\mathcal{T}_e^{(n)}(y)| \exp \left( -n H(X \mid Y) - \sqrt{\frac{n U(X \mid Y)}{\varepsilon}} \right) \tag{231}
\]
for every $y \in \mathcal{Y}^n$, where the second inequality follows from the left-hand inequality of (229). Hence, we observe that

\[
H^\alpha_n(X^n \mid Y^n) \geq (a) \frac{\alpha}{1 - \alpha} \log \left( \inf_{Q \in \mathcal{G}_{X^n} \times \mathcal{Y}^n} \sum_{y \in \mathcal{Y}^n} Q(x, y) \left( \sum_{x \in X^n} Q(x, y)^{1/\alpha} \right)^{1/\alpha} \right)
\]

\[
\leq (b) \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in \mathcal{Y}^n} P_{Y^n}(y) \left( \sum_{x \in X^n} P_{X^n,Y^n}(x, y)^{1/\alpha} \right)^{1/\alpha} \right)
\]

\[
= (c) \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in \mathcal{Y}^n} P_{Y^n}(y) \left( \sum_{x \in X^n} \exp \left( -\alpha \left( n H(X \mid Y) - \frac{\sqrt{n U(X \mid Y)}}{\varepsilon} \right) \right) \right)^{1/\alpha} \right)
\]

\[
\leq (d) \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in \mathcal{Y}^n} P_{Y^n}(y) \exp \left( \frac{1 - \alpha}{\alpha} n H(X \mid Y) + \frac{1 + \alpha}{\alpha} \frac{\sqrt{n U(X \mid Y)}}{\varepsilon} \right) \right)
\]

\[
= \frac{\alpha}{1 - \alpha} n H(X \mid Y) + \frac{1 + \alpha}{1 - \alpha} \frac{\sqrt{n U(X \mid Y)}}{\varepsilon}
\]

(232)

where

- (a) follows by the definition of $H^\alpha_n(X^n \mid Y^n)$ stated in (48),
- (b) follows from (230) and the definition of $\mathcal{B}_{X^n,Y^n}(P_{X^n,Y^n})$ stated in (49),
- (c) follows from the right-hand inequality of (229), and
- (d) follows from (231).

This completes the proof of the upper bound part of (60).

**B. Proof of Lower Bound Part of (60)**

Consider a mapping $\delta_n : \mathcal{Y}^n \rightarrow [0, 1]$ satisfying (212). Recall that $A_n(y)$ and $\tilde{A}_n(y)$ are given as (216) and (217), respectively, for each $y \in \mathcal{Y}^n$. Fix $\gamma \in (0, 1 - \varepsilon)$ arbitrarily, and consider the subset $\mathcal{T}_\gamma^{(n)}$ of $X^n \times \mathcal{Y}^n$ defined as in (227) by replacing $\varepsilon$ by $\gamma$. Moreover, define the subset $\mathcal{V}_n$ of $\mathcal{Y}^n$ by

\[
\mathcal{V}_n := \left\{ y \in \mathcal{Y}^n \left| P_{X^n,Y^n}(\tilde{A}_n(y) \cap \mathcal{T}_\gamma^{(n)}(y) \mid y) \geq \frac{1 - \varepsilon - \gamma}{2} \right. \right\}
\]

(233)

where note that $\mathcal{V}_n$ depends on $\varepsilon_n(\cdot)$. A simple calculation yields

\[
1 - \varepsilon - \gamma = \sum_{y \in \mathcal{Y}^n} P_{Y^n}(y) (1 - \varepsilon_n(y)) - \gamma
\]

\[
\leq \sum_{y \in \mathcal{Y}^n} P_{Y^n}(y) P_{X^n,Y^n}(\tilde{A}_n(y) \mid y) - \gamma
\]

\[
\leq \sum_{y \in \mathcal{Y}^n} P_{Y^n}(y) P_{X^n,Y^n}(\tilde{A}_n(y) \mid y) + \sum_{y \in \mathcal{Y}^n} P_{Y^n}(y) P_{X^n,Y^n}(\mathcal{T}_\gamma^{(n)}(y) \mid y) - 1
\]

\[
= \sum_{y \in \mathcal{Y}^n} P_{Y^n}(y) \left( P_{X^n,Y^n}(\tilde{A}_n(y) \mid y) + P_{X^n,Y^n}(\mathcal{T}_\gamma^{(n)}(y) \mid y) - 1 \right)
\]

\[
\leq \sum_{y \in \mathcal{Y}^n} P_{Y^n}(y) P_{X^n,Y^n}(\tilde{A}_n(y) \cap \mathcal{T}_\gamma^{(n)}(y) \mid y)
\]

(234)
\[
\begin{align*}
&\sum_{y \in \mathcal{V}_n} P_{\mathcal{V}_n}(y) \leq \frac{1 - \varepsilon - \gamma}{2} \sum_{y \in \mathcal{Y} \setminus \mathcal{V}_n} P_{\mathcal{V}_n}(y) \\
&= P_{\mathcal{V}_n}(\mathcal{V}_n) \left( \frac{1 + \varepsilon + \gamma}{2} \right) + \frac{1 - \varepsilon - \gamma}{2},
\end{align*}
\] (234)

where
- (a) follows from (212),
- (b) follows from the right-hand inequality of (7),
- (c) follows from (230),
- (d) follows from the fact that \( P(E_1 \cap E_2) \geq P(E_1) + P(E_2) - 1 \) for two events \( E_1 \) and \( E_2 \), and
- (e) follows by the definition of \( \mathcal{V}_n \) stated in (233).

Thus, similar to (221), we obtain
\[
P_{\mathcal{V}_n}(\mathcal{V}_n) \geq \frac{1 - \varepsilon - \gamma}{1 + \varepsilon + \gamma}.
\] (235)

Moreover, we see that
\[
\begin{align*}
\frac{1 - \varepsilon - \gamma}{2} &\leq \sum_{x \in \tilde{A}_n(y) \cap \mathcal{T}_n(y)} P_{X^n \mid Y_n}(x \mid y) \\
&\leq \sum_{x \in \tilde{A}_n(y) \cap \mathcal{T}_n(y)} \exp \left( -n H(X \mid Y) + \sqrt{\frac{n U(X \mid Y)}{\gamma}} \right) \\
&= |\tilde{A}_n(y) \cap \mathcal{T}_n(y)| \exp \left( -n H(X \mid Y) + \sqrt{\frac{n U(X \mid Y)}{\gamma}} \right)
\end{align*}
\] (236)

for every \( y \in \mathcal{V}_n \), where
- (a) follows by the definition of \( \mathcal{V}_n \) stated in (233), and
- (b) follows from the right-hand inequality of (229).

Furthermore, we get
\[
\begin{align*}
\sum_{x \in \tilde{A}_n(y) \cap \mathcal{T}_n(y)} P_{X^n \mid Y_n}(x \mid y) &\geq \sum_{x \in \tilde{A}_n(y) \cap \mathcal{T}_n(y)} \exp \left( -\alpha n H(X \mid Y) + \sqrt{\frac{n U(X \mid Y)}{\gamma}} \right) \\
&= |\tilde{A}_n(y) \cap \mathcal{T}_n(y)| \exp \left( -\alpha n H(X \mid Y) + \sqrt{\frac{n U(X \mid Y)}{\gamma}} \right) \\
&\geq \frac{1 - \varepsilon - \gamma}{2} \exp \left( (1 - \alpha) n H(X \mid Y) - (1 + \alpha) \sqrt{\frac{n U(X \mid Y)}{\gamma}} \right)
\end{align*}
\] (237)

for every \( y \in \mathcal{V}_n \), where
- (a) follows from the left-hand inequality of (229), and
- (b) follows from (236).

Now, since \( 0 < U(X \mid Y) < \infty \) implies that \( 0 < H(X \mid Y) < \infty \), one can choose an integer \( n_1 = n_1(\varepsilon, \delta, \alpha, H(X \mid Y), U(X \mid Y)) \) so that
\[
(1 - \alpha) n H(X \mid Y) - (1 + \alpha) \sqrt{\frac{n U(X \mid Y)}{\gamma}} \geq \log \frac{4}{1 - \varepsilon - \gamma}
\] (238)
for every $n \geq n_1$. Then, we observe that

\[
H^c_{\alpha}(X^n | Y^n) &\overset{(a)}{=} \frac{\alpha}{1-\alpha} \log \left( \inf_{\delta_n(\cdot)} \sum_{y \in Y_n} P_{Y^n}(y) \left( \sum_{x \in \mathcal{X}_n(y)} P_{X^n | Y^n}(x | y) + \left(1 - \delta_n(y) - P_{X^n | Y^n}(\mathcal{X}_n(y) | y)\right)^{1/\alpha} \right) \right) \\
\geq & \frac{\alpha}{1-\alpha} \log \left( \inf_{\delta_n(\cdot)} \sum_{y \in Y_n} P_{Y^n}(y) \left( \sum_{x \in \mathcal{X}_n(y)} P_{X^n | Y^n}(x | y) + \left(1 - \delta_n(y) - P_{X^n | Y^n}(\mathcal{X}_n(y) | y)\right)^{1/\alpha} \right) \right) \\
\overset{(b)}{=} & \frac{\alpha}{1-\alpha} \log \left( \inf_{\delta_n(\cdot)} \sum_{y \in Y_n} P_{Y^n}(y) \left( \sum_{x \in \mathcal{X}_n(y)} P_{X^n | Y^n}(x | y)^{1/\alpha} \right) \right) \\
\overset{(c)}{=} & \frac{\alpha}{1-\alpha} \log \left( \inf_{\delta_n(\cdot)} \sum_{y \in Y_n} P_{Y^n}(y) \left( \sum_{x \in \mathcal{X}_n(y)} P_{X^n | Y^n}(x | y)^{1/\alpha} - 1 \right) \right) \\
\overset{(d)}{=} & \frac{\alpha}{1-\alpha} \log \left( \inf_{\delta_n(\cdot)} \sum_{y \in Y_n} P_{Y^n}(y) \left( \frac{1 - \epsilon - \gamma}{2} \exp \left( (1 - \alpha) n H(X | Y) - (1 + \alpha) \sqrt{\frac{n U(X | Y)}{\gamma}} \right) - 1 \right) \right) \\
\overset{(e)}{=} & \frac{\alpha}{1-\alpha} \log \left( \inf_{\delta_n(\cdot)} \sum_{y \in Y_n} P_{Y^n}(y) \left( \frac{1 - \epsilon - \gamma}{4} \exp \left( (1 - \alpha) n H(X | Y) - (1 + \alpha) \sqrt{\frac{n U(X | Y)}{\gamma}} \right) \right) \right) \\
= & n H(X | Y) - \frac{1 + \alpha}{1 - \alpha} \sqrt{\frac{n U(X | Y)}{\gamma}} + \frac{1}{1 - \alpha} \log \left( \frac{1 - \epsilon - \gamma}{4} \exp \left( (1 - \alpha) n H(X | Y) - (1 + \alpha) \sqrt{\frac{n U(X | Y)}{\gamma}} \right) \right) + \frac{\alpha}{1 - \alpha} \log \left( \inf_{\delta_n(\cdot)} P_{Y^n}(V_n) \right) \\
\geq & n H(X | Y) - \frac{1 + \alpha}{1 - \alpha} \sqrt{\frac{n U(X | Y)}{\gamma}} + \frac{1}{1 - \alpha} \log \left( \frac{1 - \epsilon - \gamma}{4} \exp \left( (1 - \alpha) n H(X | Y) - (1 + \alpha) \sqrt{\frac{n U(X | Y)}{\gamma}} \right) \right) + \frac{\alpha}{1 - \alpha} \log \left( \inf_{\delta_n(\cdot)} P_{Y^n}(V_n) \right) \\
\text{for sufficiently large } n \geq n_1, \text{ where} \\
\bullet \ (a) \text{ follows from (61)}, \\
\bullet \ (b) \text{ follows from (226)}, \\
\bullet \ (c) \text{ follows from (237)}, \\
\bullet \ (d) \text{ follows by the choice of } n_1 \text{ stated in (238)}, \text{ and} \\
\bullet \ (e) \text{ follows from (235)}. \\
\text{This completes the proof of the lower bound part of (60).} \]

\section*{APPENDIX D}

\textbf{Proof of Theorem 3—Zero Variance $V(X | Y) = 0$}

\textbf{A. Proof of Upper Bound Part of (68)}

Consider the conditional sub-probability distribution $\tilde{Q}_{X^n | Y^n}$ given as

\[
\tilde{Q}_{X^n | Y^n}(x | y) = (1 - \epsilon) P_{X^n | Y^n}(x | y)
\]

for each $(x, y) \in X^n \times Y^n$. Then, we observe that

\[
\tilde{H}^c_{\alpha}(X^n | Y^n) \overset{(a)}{=} \frac{\alpha}{1-\alpha} \log \left( \sum_{y \in Y} P_{Y^n}(y) \right) \left( \sum_{x \in X^n} \tilde{Q}(x)^{1/\alpha} \right) \\
\overset{(b)}{=} \frac{\alpha}{1-\alpha} \log \left( \sum_{y \in Y} P_{Y^n}(y) \left( \sum_{x \in X^n} \tilde{Q}_{X^n | Y^n}(x | y)^{1/\alpha} \right) \right) \\
\overset{(c)}{=} \frac{\alpha}{1-\alpha} \log \left( \sum_{y \in Y} P_{Y^n}(y) \left( \sum_{x \in X^n} P_{X^n | Y^n}(x | y)^{1/\alpha} \right) \right) + \frac{\alpha}{1 - \alpha} \log(1 - \epsilon) \\
\overset{(d)}{=} \frac{\alpha}{1-\alpha} \log \left( \sum_{y \in Y} P_{Y^n}(y) \exp \left( \frac{1 - \alpha}{\alpha} \sum_{i=1}^{n} H(P_{X_i | Y_i = y_i}) \right) \right) + \frac{\alpha}{1 - \alpha} \log(1 - \epsilon)
\]
where

- (a) follows from the fact that \( \tilde{Q}_{X^n|Y^n}(\cdot | y) \in \mathcal{B}_{X^n|Y^n}(P_{X^n|Y^n}=y) \) for each \( y \in Y^n \),
- (b) follows by the definition of \( \tilde{Q}_{X^n|Y^n} \) stated in (240),
- (c) follows from (71),
- (d) follows from the fact that \((X_1, Y_1), \ldots, (X_n, Y_n)\) are \( n \) i.i.d. copies of \((X, Y)\), and
- (e) follows by the definition of \( H^{(\alpha)}(X | Y) \) stated in (67).

This completes the proof of the upper bound part of (68).

\[ \blacksquare \]

\section*{B. Proof of Lower Bound Part of (68)}

Firstly, suppose that \( U(X | Y) = 0 \). Then, it is clear that

\[ P_{X^n|Y^n}(x | y) = \begin{cases} \exp \left( -n H(X | Y) \right) & \text{if } P_{X^n|Y^n}(x | y) > 0, \\ 0 & \text{if } P_{X^n|Y^n}(x | y) = 0 \end{cases} \]  

for every \((x, y) \in X^n \times Y^n\). For the sake of brevity, denote by

\[ \mathcal{B}_n(y) := \mathcal{R}_{X^n|Y^n}(y) \]  

(243)

for each \( y \in Y^n \); see (6) and (7) for the definition of the right-hand side of (243) by replacing \( P_X(\cdot) \) by \( P_{X^n|Y^n}(\cdot | y) \). Moreover, define

\[ \tilde{\mathcal{B}}_n(y) := \mathcal{B}_n(y) \cup \{ \tilde{x}(y) \}, \]  

(244)

where \( \tilde{x}(y) \in X^n \) is chosen so that

\[ \tilde{x}(y) \in \arg \max_{x \in X^n \setminus \mathcal{B}_n(y)} P_{X^n|Y^n}(x | y). \]  

(245)

for each \( y \in Y^n \). We get

\[ 1 - \varepsilon \overset{(a)}{=} \sum_{x \in \mathcal{B}_n(y)} P_{X^n|Y^n}(x | y) \]

\[ \overset{(b)}{=} \sum_{x \in \mathcal{B}_n(y)} \exp \left( -n H(X | Y) \right) \]

\[ = |\mathcal{B}_n(y)| \exp \left( -n H(X | Y) \right) \]  

(246)

for every \( y \in Y^n \), where

- (a) follows from the right-hand inequality of (7), and
- (b) follows from (6) and (242).

Moreover, we see that

\[ \sum_{x \in \mathcal{B}_n(y)} P_{X^n|Y^n}(x | y) \overset{(a)}{=} \sum_{x \in \mathcal{B}_n(y)} \exp \left( -\alpha n H(X | Y) \right) \]

\[ = |\tilde{\mathcal{B}}_n(y)| \exp \left( -\alpha n H(X | Y) \right) \]

\[ \overset{(b)}{\geq} (1 - \varepsilon) \exp \left( (1 - \alpha) n H(X | Y) \right) \]  

(a.s.)  

(247)

for every \( y \in Y^n \), where

- (a) follows from (6) and (242), and
- (b) follows from (246).

Since \( H(X | Y) > 0 \), one can choose an integer \( n_0 \) as

\[ n_0 := \left\lceil \frac{1}{(1 - \alpha) H(X | Y)} \log \frac{2}{1 - \varepsilon} \right\rceil. \]  

(248)
Then, we observe that
\[
\hat{H}_n^{(\alpha)}(X^n | Y^n) \overset{(a)}{=} \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in Y^n} P_{Y^n}(y) \left( \sum_{x \in \mathcal{B}_n(y)} P_{X^n | Y^n}(x \mid y)^\alpha + \left( 1 - \varepsilon - P_{X^n | Y^n}(\mathcal{B}_n(y) \mid y) \right)^\alpha \right) \right)^{1/\alpha} \\
\overset{(b)}{\geq} \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in Y^n} P_{Y^n}(y) \left( \sum_{x \in \mathcal{B}_n(y)} P_{X^n | Y^n}(x \mid y)^\alpha - 1 \right) \right)^{1/\alpha} \\
\overset{(c)}{\geq} \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in Y^n} P_{Y^n}(y) \left( 1 - \varepsilon \right) \exp \left( \left( 1 - \frac{\alpha}{2} \right) n H(X \mid Y) \right) \right)^{1/\alpha} \\
\overset{(d)}{\geq} \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in Y^n} P_{Y^n}(y) \left( \frac{1 - \varepsilon}{2} \right) \exp \left( \frac{1 - \alpha}{1 - \alpha} n H(X \mid Y) \right) \right) \\
= n H(X \mid Y) + \frac{1}{1 - \alpha} \log \frac{1 - \varepsilon}{2} \\
\overset{(e)}{=} n H^{(\alpha)}(X \mid Y) + \frac{1}{1 - \alpha} \log \frac{1 - \varepsilon}{2} 
\]  
(249)
for sufficiently large \( n \geq n_0 \), where
- (a) follows from (70),
- (b) follows from the right-hand inequality of (7), i.e.,
\[
\left( 1 - \varepsilon - P_{X^n | Y^n}(\mathcal{B}_n(y) \mid y) \right)^\alpha \geq P_{X^n | Y^n}(\mathcal{B}(y) \mid y) - 1.
\]  
(250)
- (c) follows from (247),
- (d) follows from the choice of \( n_0 \) stated in (248), and
- (e) follows from Proposition 1 and the hypothesis that \( U(X \mid Y) = V(X \mid Y) = 0 \).

This completes the proof of the lower bound part of (68) in the case where \( U(X \mid Y) = 0 \).

Secondly, suppose that \( U(X \mid Y) > 0 \). Defining a subset \( \mathcal{E}_n \) of \( Y^n \) by
\[
\mathcal{E}_n := \left\{ y \mid H(P_{X^n | Y^n = y}) \geq n H(X \mid Y) - \sqrt{n(U(X \mid Y) - V(X \mid Y))} \right\},
\]  
(251)
we see that
\[
P_{Y^n}(\mathcal{E}_n) \overset{(a)}{=} 1 - \frac{\sum_{y^n} P_{Y^n}(y) \left( H(P_{X^n | Y^n = y}) - H(X^n | Y^n) \right)^2}{\sum_{y^n} P_{Y^n}(y) \left( H(P_{X^n | Y^n = y}) - H(X^n | Y^n) \right)^2 + (U(X^n | Y^n) - V(X^n | Y^n))} \\
\overset{(b)}{=} \frac{1}{2},
\]  
(252)
where
- (a) follows from Cantelli’s inequality (or the one-sided Chebyshev inequality), and
- (b) follows from the law of total variance.

Similar to (246), we see that
\[
1 - \varepsilon \overset{(a)}{=} \sum_{x \in \mathcal{B}_n(x)} P_{X^n | Y^n}(x \mid y) \\
\overset{(b)}{=} \sum_{x \in \mathcal{B}_n(x)} \exp \left( -H(P_{X^n | Y^n = y}) \right) \\
= |\mathcal{B}_n(x)| \exp \left( -H(P_{X^n | Y^n = y}) \right) 
\]  
(253)
for every \( y \in Y^n \), where
- (a) follows from the right-hand inequality of (7), and
- (b) follows from (6) and (71).
Moreover, we observe that

\[
\sum_{x \in \mathcal{B}_n(y)} P_{X^n|Y^n}(x \mid y) \geq \sum_{x \in \mathcal{B}_n(y)} \exp \left( -\alpha H(P_{X^n|Y^n}) \right)
\]

(a) follows from (6) and (71), and

(b) follows from (253).

Since \( U(X \mid Y) > 0 \) implies that \( H(X \mid Y) > 0 \), one can choose an integer \( n_0 \) so that

\[
n H(X \mid Y) - \sqrt{n(U(X \mid Y) - V(X \mid Y))} \geq \frac{1}{1-\alpha} \log \frac{2}{1-\varepsilon}
\]

(255)

for every \( n \geq n_0 \). Then, we observe that

\[
\tilde{H}_\alpha^n(X^n \mid Y^n) \geq \frac{\alpha}{1-\alpha} \log \left( \sum_{y \in \mathcal{Y}^n} P_{Y^n}(y) \left( \sum_{x \in \mathcal{B}_n(y)} P_{X^n|Y^n}(x \mid y)^\alpha - 1 \right)^{1/\alpha} \right)
\]

(b) follows from (254).
\[
\begin{aligned}
&\left(\phi\right) \geq \frac{\alpha}{1-\alpha} \log \left( \sum_{y \in Y} P_Y(y) \exp \left( \frac{1 - \alpha}{\alpha} H(P_{X|Y=y}) \right) \right)^n - \frac{1}{2} \exp \left( \frac{1 - \alpha}{\alpha} n H(X | Y) \right) + \frac{1}{1 - \alpha} \log \frac{1 - \epsilon}{2} \\
&\left(\gamma\right) \geq \frac{\alpha}{1-\alpha} \log \left( \frac{1}{2} \sum_{y \in Y} P_Y(y) \exp \left( \frac{1 - \alpha}{\alpha} H(P_{X|Y=y}) \right) \right)^n + \frac{1}{1 - \alpha} \log \frac{1 - \epsilon}{2} \\
&= n H^{(\alpha)}(X | Y) + \frac{\alpha}{1-\alpha} \log \frac{1}{2} + \frac{1}{1 - \alpha} \log \frac{1 - \epsilon}{2}
\end{aligned}
\] (256)

for sufficiently large \( n \geq n_0 \), where

- (a) follows as in Steps (a) and (b) of (249),
- (b) follows from (254),
- (c) follows by the definition of \( E_n \) stated in (251) and the choice of \( n_0 \) stated in (255),
- (d) follows from the fact that \( X_1, Y_1, \ldots, X_n, Y_n \) are i.i.d. copies of \( (X, Y) \),
- (e) follows by the definition of \( E_n \) stated in (251),
- (f) follows from (252), and
- (g) follows from Jensen’s inequality.

This completes the proof of the lower bound part of (68) in the case where \( U(X | Y) > 0 \). \( \blacksquare \)

**APPENDIX E**

**Proof of Theorem 3—Positive Variance 0 < V(X | Y) < \infty**

**A. Proof of Upper Bound Part of (69)**

For each \( y \in Y^n \), consider a subset \( D^{(n)}_e(y) \) of \( X^n \) given by

\[
D^{(n)}_e(y) := \left\{ x \middle| \log \frac{1}{P_{X^n|Y^n}(x | y)} - H(P_{X^n|Y^n=y}) \leq \sqrt{\frac{V(P_{X^n|Y^n=y})}{\epsilon}} \right\}.
\] (257)

Note that

\[
\exp \left( -H(P_{X^n|Y^n=y}) - \sqrt{\frac{V(P_{X^n|Y^n=y})}{\epsilon}} \right) \leq P_{X^n|Y^n}(x | y) \leq \exp \left( -H(P_{X^n|Y^n=y}) + \sqrt{\frac{V(P_{X^n|Y^n=y})}{\epsilon}} \right)
\] (258)

whenever \( x \in D^{(n)}_e(y) \). By the left-hand inequality of (258), it can be verified by the same way as (231) that

\[
|D^{(n)}_e(y)| \leq \exp \left( H(P_{X^n|Y^n=y}) + \sqrt{\frac{V(P_{X^n|Y^n=y})}{\epsilon}} \right)
\] (259)

for every \( y \in Y^n \). It follows from Chebyshev’s inequality that

\[
P_{X^n|Y^n}(D^{(n)}_e(y) | y) \geq 1 - \epsilon
\] (260)

for every \( y \in Y^n \) in which \( V(P_{X^n|Y^n=y}) > 0 \). On the other hand, it follows from (71) that

\[
P_{X^n|Y^n}(D^{(n)}_e(y) | y) = 1
\] (261)
for every $y \in Y^n$ in which $V(P_{X^n|Y^n=y}) = 0$. Hence, we have

$$
\mathcal{H}_n^e(X^n | Y^n) = \left( a \right) \frac{\alpha}{1 - \alpha} \log \left( \frac{\prod_{y \in Y^n} P_{Y^n}(y)}{\inf_{P_X \in B_n^e(P_{X^n|Y^n=y})} \left( \prod_{x \in X^n} Q(x)^\alpha \right)} \right)^{1/\alpha}
$$

$$
\leq \left( b \right) \frac{\alpha}{1 - \alpha} \log \left( \frac{\prod_{y \in Y^n} P_{Y^n}(y)}{\prod_{x \in Y^n(y)} \left( \prod_{x \in X^n(y)} P_{X^n|Y^n}(x | y)^\alpha \right)} \right)^{1/\alpha}
$$

$$
\leq \left( c \right) \frac{\alpha}{1 - \alpha} \log \left( \frac{\prod_{y \in Y^n} P_{Y^n}(y)}{\prod_{x \in Y^n(y)} \exp \left( -\alpha \left( H(P_{X^n|Y^n=y}) - \sqrt{\frac{V(P_{X^n|Y^n=y})}{\epsilon}} \right) \right)} \right)^{1/\alpha}
$$

$$
\leq \left( d \right) \frac{\alpha}{1 - \alpha} \log \left( \frac{\prod_{y \in Y^n} P_{Y^n}(y)}{\prod_{x \in Y^n(y)} \exp \left( -\frac{1 - \alpha}{\alpha} H(P_{X^n|Y^n=y}) - \frac{1}{\alpha} \sqrt{\frac{V(P_{X^n|Y^n=y})}{\epsilon}} \right)} \right)^{1/\alpha}
$$

$$
\leq \left( e \right) \frac{n\alpha}{1 - \alpha} \log \left( \frac{\prod_{y \in Y} P_{Y}(y)}{\prod_{x \in Y} \exp \left( -\frac{1 - \alpha}{\alpha} H(P_{X|Y=y}) - \frac{1}{\alpha} \sqrt{\frac{n \sup_{y \in Y} V(P_{X|Y=y})}{\epsilon}} \right)} \right)^{1/\alpha}
$$

$$
= n H^{(\alpha)}(X | Y) + \frac{1 + \alpha}{1 - \alpha} \sqrt{n \sup_{y \in Y} V(P_{X|Y=y})} \epsilon (262)
$$

where

- (a) follows by the definition of $\mathcal{H}_n^e(X^n | Y^n)$ stated in (63),
- (b) follows from (260) and (261),
- (c) follows from the right-hand inequality of (258),
- (d) follows from (259),
- (e) follows from the hypothesis that $\sup_{y \in Y} V(P_{X|Y=y}) < \infty$, and
- (f) follows from the fact that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. copies of $(X, Y)$.

This completes the proof of the upper bound part of (69).

\[ \blacksquare \]

**B. Proof of Lower Bound Part of (69)**

Recall that $B_n(y)$ and $\tilde{B}_n(y)$ are defined as (243) and (244), respectively, for each $y \in Y^n$. Fix $\gamma \in (0, 1 - \epsilon)$ arbitrarily, and consider the subset $D_{\gamma}^{(n)}(y)$ of $X^n$ defined as in (257) by replacing $\epsilon$ by $\gamma$. A simple calculation yields

$$
1 - \epsilon - \gamma \leq P_{X^n|Y^n}(\tilde{B}_n(y) | y) - \gamma
$$

$$
\leq P_{X^n|Y^n}(\tilde{B}_n(y) | y) + P_{X^n|Y^n}(D_{\gamma}^{(n)}(y) | y) - 1
$$

$$
\leq P_{X^n|Y^n}(\tilde{B}_n(y) \cap D_{\gamma}^{(n)}(y) | y^n)
$$

$$
\leq |\tilde{B}_n(y) \cap D_{\gamma}^{(n)}(y)| \exp \left( -H(P_{X^n|Y^n=y}) - \frac{\sqrt{V(P_{X^n|Y^n=y})}}{\gamma} \right) \epsilon (263)
$$

for every $y \in Y^n$, where

- (a) follows from the right-hand inequality of (7),
- (b) follows from (260),
- (c) follows from the fact that $\mu(\mathcal{A}) + \mu(\mathcal{B}) \leq \mu(\mathcal{A} \cap \mathcal{B}) + 1$ for every probability measure $\mu$, and
- (f) follows from the right-hand inequality of (258).
In addition, we have

\[
\sum_{x \in \hat{\mathcal{B}}_n(y) \cap \mathcal{D}_n(y)} P_{X^n|Y^n}(x | y)^\alpha \geq \sum_{x \in \hat{\mathcal{B}}_n(y) \cap \mathcal{D}_n(y)} \exp \left( -\alpha H(P_{X^n|Y^n}) - \alpha \sqrt{\frac{V(P_{X^n|Y^n})}{\gamma}} \right)
\]

\[
= |\hat{\mathcal{B}}_n(y) \cap \mathcal{D}_n(y)| \exp \left( -\alpha H(P_{X^n|Y^n}) - \alpha \sqrt{\frac{V(P_{X^n|Y^n})}{\gamma}} \right)
\]

\[
\geq (1 - \varepsilon - \gamma) \exp \left( (1 - \alpha) H(P_{X^n|Y^n}) - (1 + \alpha) \sqrt{\frac{V(P_{X^n|Y^n})}{\gamma}} \right)
\]

\[
\geq (1 - \varepsilon - \gamma) \exp \left( (1 - \alpha) H(P_{X^n|Y^n}) - (1 + \alpha) \frac{n \sup_{y \in \mathcal{Y}} V(P_{X^n|Y^n})}{\gamma} \right)
\]

(264)

for every \( y \in \mathcal{Y}^n \), where

- (a) follows from the left-hand inequality of (258), and
- (b) follows from (263).

Firstly, suppose that \( U(X | Y) = V(X | Y) \). Since \( V(X | Y) > 0 \) implies that \( H(X | Y) > 0 \), and since we have assumed that \( \sup_{y \in \mathcal{Y}} V(P_{X|Y=y}) < \infty \), one can find a positive integer \( n_2 = n_2(\varepsilon, \delta, \alpha, H(X | Y), \sup_{y \in \mathcal{Y}} V(P_{X|Y=y})) \) such that

\[
n(1 - \alpha) H(X | Y) - (1 + \alpha) \frac{n \sup_{y \in \mathcal{Y}} V(P_{X|Y=y})}{\delta} \geq \log \frac{2}{1 - \varepsilon - \delta}
\]

(265)

for every \( n \geq n_2 \). We observe that

\[
\hat{H}_a^n(X^n | Y^n) \geq \left( \frac{\alpha}{1 - \alpha} \right) \log \left( \sum_{y \in \mathcal{Y}^n} P_{Y^n}(y) \sum_{x \in \hat{\mathcal{B}}_n(y) \cap \mathcal{D}_n(y)} P_{X^n|Y^n}(x | y)^\alpha - 1 \right)^{1/\alpha}
\]

\[
\geq \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in \mathcal{Y}^n} P_{Y^n}(y) \sum_{x \in \hat{\mathcal{B}}_n(y) \cap \mathcal{D}_n(y)} P_{X^n|Y^n}(x | y)^\alpha - 1 \right)^{1/\alpha}
\]

\[
\geq \frac{\alpha}{1 - \alpha} \log \left( (1 - \varepsilon - \gamma) \exp \left( (1 - \alpha) H(P_{X^n|Y^n}) - (1 + \alpha) \frac{n \sup_{y \in \mathcal{Y}} V(P_{X|Y=y})}{\gamma} \right) - 1 \right)^{1/\alpha}
\]

\[
= \frac{\alpha}{1 - \alpha} \log \left( (1 - \varepsilon - \gamma) \exp \left( (1 - \alpha) H(X | Y) - (1 + \alpha) \frac{n \sup_{y \in \mathcal{Y}} V(P_{X|Y=y})}{\gamma} \right) - 1 \right)^{1/\alpha}
\]

\[
= \frac{\alpha}{1 - \alpha} \log \left( (1 - \varepsilon - \gamma) \exp \left( (1 - \alpha) H(X | Y) - (1 + \alpha) \frac{n \sup_{y \in \mathcal{Y}} V(P_{X|Y=y})}{\gamma} \right) - 1 \right)^{1/\alpha}
\]

\[
= \frac{\alpha}{1 - \alpha} \log \left( \frac{1 - \varepsilon - \gamma}{2} \exp \left( \frac{1 - \alpha}{\alpha} n H(X | Y) - 1 + \frac{1 - \alpha}{\alpha} \frac{n \sup_{y \in \mathcal{Y}} V(P_{X|Y=y})}{\gamma} \right) \right)^{1/\alpha}
\]

\[
= n H(X | Y) - \frac{1 + \alpha}{1 - \alpha} \frac{n \sup_{y \in \mathcal{Y}} V(P_{X|Y=y})}{\gamma} + \frac{1}{1 - \alpha} \log \frac{1 - \varepsilon - \gamma}{2}
\]

\[
= n H^{(a)}(X | Y) - \frac{1 + \alpha}{1 - \alpha} \frac{n \sup_{y \in \mathcal{Y}} V(P_{X|Y=y})}{\gamma} + \frac{1}{1 - \alpha} \log \frac{1 - \varepsilon - \gamma}{2}
\]

(266)

for sufficiently large \( n \geq n_2 \), where

- (a) follows as in Steps (a) and (b) of (249),
- (b) follows from (264),
- (c) follows by the law of total variance and the hypothesis that \( U(X | Y) = V(X | Y) \),
- (d) follows by the choice of \( n_2 \); see (265), and
- (e) follows from Proposition 1.
This completes the proof of the lower bound part of (69) in the case where $U(X \mid Y) = V(X \mid Y)$.

Secondly, suppose that $U(X \mid Y) > V(X \mid Y)$. Recall that the subset $E_n$ of $Y^n$ is defined in (251). Since $V(X \mid Y) > 0$ implies that $H(X \mid Y) > 0$, and since we have assumed that $\sup_{y \in Y} V(P_X \mid y < \infty$, one can find a positive integer $n_3 = n_3(\epsilon, \delta, H(X \mid Y), U(X \mid Y), V(X \mid Y), \sup_{y \in Y} V(P_X \mid y))$ such that

$$
n (1 - \alpha) H(X \mid Y) - \sqrt{\alpha} \left( \frac{\langle U(X \mid Y) - V(X \mid Y) \rangle}{\delta} \right) \geq \log \frac{2}{1 - \epsilon - \delta} \tag{267}
$$

for every $n \geq n_3$. We then observe that

$$
\begin{align*}
\tilde{H}_n(X^n \mid Y^n) & \geq \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in Y^n} P_{Y^n}(y) \left( \sum_{x \in B_n(y)} P_{X^n \mid y}(x \mid y)^\alpha - 1 \right)^{1/\alpha} \right) \\
& \geq \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in E_n} P_{Y^n}(y) \left( \sum_{x \in B_n(y)} P_{X^n \mid y}(x \mid y)^\alpha - 1 \right)^{1/\alpha} \right) \\
& \geq \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in E_n} P_{Y^n}(y) \left( 1 - \epsilon - \gamma \right) \exp \left( (1 - \alpha) H(P_{X^n \mid y^n}) - (1 + \alpha) \sqrt{\frac{n \sup_{y \in Y} V(P_X \mid y)}{\gamma}} \right) \\
& \quad - \frac{1 - \epsilon - \delta}{2} \exp \left( n (1 - \alpha) H(X \mid Y) - (1 - \alpha) \sqrt{\frac{n \sup_{y \in Y} V(P_X \mid y)}{\gamma}} \right) \right)^{1/\alpha} \right) \\
& \geq \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in E_n} P_{Y^n}(y) \left( \frac{1 - \epsilon - \gamma}{2} \right) \exp \left( \frac{1 - \alpha}{\alpha} H(P_{X^n \mid y^n}) - \frac{1 + \alpha}{\alpha} \sqrt{\frac{n \sup_{y \in Y} V(P_X \mid y)}{\gamma}} \right) \right) \\
& = \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in E_n} P_{Y^n}(y) \left( \frac{1 - \alpha}{\alpha} H(P_{X^n \mid y^n}) \right) \right) - \frac{1 + \alpha}{1 - \alpha} \sqrt{\frac{n \sup_{y \in Y} V(P_X \mid y)}{\gamma}} + \frac{1}{1 - \alpha} \log \frac{1 - \epsilon - \gamma}{2} \\
& = \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in Y^n} P_{Y^n}(y) \left( \frac{1 - \alpha}{\alpha} H(P_{X^n \mid y^n}) \right) \right) - \frac{1 + \alpha}{1 - \alpha} \sqrt{\frac{n \sup_{y \in Y} V(P_X \mid y)}{\gamma}} + \frac{1}{1 - \alpha} \log \frac{1 - \epsilon - \gamma}{2} \\
& \geq \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in Y} P_{Y}(y) \left( \frac{1 - \alpha}{\alpha} H(P_{X \mid y}) \right) \right) - \frac{1 + \alpha}{1 - \alpha} \sqrt{\frac{n \sup_{y \in Y} V(P_X \mid y)}{\gamma}} + \frac{1}{1 - \alpha} \log \frac{1 - \epsilon - \gamma}{2} \\
& = \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in Y} P_{Y}(y) \left( \frac{1 - \alpha}{\alpha} H(P_{X \mid y}) \right) \right) - \frac{1 + \alpha}{1 - \alpha} \sqrt{\frac{n \sup_{y \in Y} V(P_X \mid y)}{\gamma}} + \frac{1}{1 - \alpha} \log \frac{1 - \epsilon - \gamma}{2} \\
& \geq \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in Y} P_{Y}(y) \left( \frac{1 - \alpha}{\alpha} H(P_{X \mid y}) \right) \right) - \frac{1 + \alpha}{1 - \alpha} \sqrt{\frac{n \sup_{y \in Y} V(P_X \mid y)}{\gamma}} + \frac{1}{1 - \alpha} \log \frac{1 - \epsilon - \gamma}{2} \\
& = \frac{\alpha}{1 - \alpha} \log \left( \sum_{y \in Y} P_{Y}(y) \left( \frac{1 - \alpha}{\alpha} H(P_{X \mid y}) \right) \right) - \frac{1 + \alpha}{1 - \alpha} \sqrt{\frac{n \sup_{y \in Y} V(P_X \mid y)}{\gamma}} + \frac{1}{1 - \alpha} \log \frac{1 - \epsilon - \gamma}{2}
\end{align*}
\]
This completes the proof of the lower bound part of (69) in the case where $U(X \mid Y) > V(X \mid Y)$.

\[ \text{APPENDIX F} \]

\text{PROOF OF LEmMA 4} 

For each $(x, y) \in \mathcal{X} \times \mathcal{Y}$, define

\[
a(x, y) := \begin{cases} \kappa(x, y)^{-\rho/(1 + \rho)} & \text{if } 0 \leq \epsilon(x, y) < 1, \\ 0 & \text{if } \epsilon(x, y) = 1, \end{cases}
\]

(269)

\[
b(x, y) := \kappa(x, y)^{\rho/(1 + \rho)} \left( 1 - \epsilon(x, y) \right) P_{X \mid Y}(x \mid y)^{1/(1 + \rho)}. \tag{270}
\]

Applying Hölder’s inequality (see, e.g., [33, Problem 4.15]),

\[
\sum_{x \in \mathcal{X}} a(x, y) b(x, y) \leq \left( \sum_{x \in \mathcal{X}} a(x, y)^{1/\lambda} \right)^{\lambda} \left( \sum_{x \in \mathcal{X}} b(x, y)^{1/(1 - \lambda)} \right)^{1 - \lambda}
\]

(271)

with $\lambda = \rho/(1 + \rho)$, we get

\[
\left( \sum_{x \in \mathcal{X}} \left( 1 - \epsilon(x, y) \right) P_{X \mid Y}(x \mid y)^{1/(1 + \rho)} \right)^{1 + \rho} \leq \left( \sum_{x \in \mathcal{X}; 0 \leq \epsilon(x, y) < 1} \frac{1}{\kappa(x, y)} \right)^{\rho} \left( \sum_{x \in \mathcal{X}} \left( 1 - \epsilon(x, y) \right) P_{X \mid Y}(x \mid y) \kappa(x, y)^{\rho} \right)^{1/(1 + \rho)}
\]

(272)

for every $y \in \mathcal{Y}$. Now, choose the function $\delta : \mathcal{Y} \to [0, 1]$ so that

\[
\delta(y) = \mathbb{E}[\epsilon(X, Y) \mid Y = y] = \sum_{x \in \mathcal{X}} \epsilon(x, y) P_{X \mid Y}(x \mid y).
\]

(273)
Then, we have
\[
\mathbb{E}[K(X, Y)\rho \mid Y = y] = \sum_{x \in \mathcal{X}} (1 - \varepsilon(x, y)) P_{X|Y}(x \mid y) \kappa(x, y)^\rho
\]

\[
\overset{(a)}{\geq} \left( \sum_{x \in \mathcal{X}} \left( (1 - \varepsilon(x, y)) P_{X|Y}(x \mid y) \right)^{1/(1 + \rho)} \right)^{1 + \rho} \left( \sum_{x \in \mathcal{X}, \kappa(x, y) < 1} \frac{1}{\kappa(x, y)} \right)^{-\rho}
\]

\[
\overset{(b)}{=} \inf_{\delta : \mathcal{Y} \rightarrow [0, 1]} \left( \sum_{x \in \mathcal{X}} \left( (1 - \bar{\varepsilon}(x)) P_{X|Y}(x \mid y) \right)^{1/(1 + \rho)} \right)^{1 + \rho} \left( \sum_{x \in \mathcal{X}, \kappa(x, y) < 1} \frac{1}{\kappa(x, y)} \right)^{-\rho}
\]

\[
\overset{(c)}{=} \exp \left( \rho H_{1/(1 + \rho)}^\delta(y)(P_{X|Y=y}) - \rho \log \sum_{x \in \mathcal{X}, \kappa(x, y) < 1} \frac{1}{\kappa(x, y)} \right)
\]

for every \( y \in \mathcal{Y} \), where
- (a) follows from (272),
- (b) follows by the choice of \( \delta : \mathcal{Y} \rightarrow [0, 1] \) stated in (273), and
- (c) follows by the definition of \( H_{1/(1 + \rho)}^\delta(y)(P_{X|Y=y}) \) stated in (3).

Therefore, we obtain
\[
\mathbb{E}[K(X, Y)\rho] = \sum_{y \in \mathcal{Y}} P_Y(y) \mathbb{E}[K(X, Y)\rho \mid Y = y]
\]

\[
\overset{(a)}{\geq} \sum_{y \in \mathcal{Y}} P_Y(y) \exp \left( \rho H_{1/(1 + \rho)}^\delta(y)(P_{X|Y=y}) - \rho \log \sum_{x \in \mathcal{X}, \kappa(x, y) < 1} \frac{1}{\kappa(x, y)} \right)
\]

\[
\overset{(b)}{=} \left( \sum_{y \in \mathcal{Y}} P_Y(y) \exp \left( \rho H_{1/(1 + \rho)}^\delta(y)(P_{X|Y=y}) \right) \right) \exp \left( - \rho \log R(\varepsilon, \kappa) \right)
\]

\[
\overset{(c)}{=} \exp \left( \rho \tilde{H}_{1/(1 + \rho)}(X \mid Y) \right) \exp \left( - \rho \log R(\varepsilon, \kappa) \right)
\]

\[
\overset{(d)}{\geq} \exp \left( \rho \inf_{\delta : \mathcal{Y} \rightarrow [0, 1]} \tilde{H}_{1/(1 + \rho)}(X \mid Y) \right) \exp \left( - \rho \log R(\varepsilon, \kappa) \right)
\]

\[
\overset{(e)}{=} \tilde{H}_{1/(1 + \rho)}(X \mid Y) \exp \left( - \rho \log R(\varepsilon, \kappa) \right)
\]

\[
\overset{(f)}{=} \exp \left( \rho \tilde{H}_{1/(1 + \rho)}(X \mid Y) \right) \exp \left( - \rho \log R(\varepsilon, \kappa) \right)
\]

where
- (a) follows from (274),
- (b) follows by the definition of \( R(\varepsilon, \kappa) \) stated in (86),
- (c) follows by the definition of \( \tilde{H}_{\varepsilon}^\delta(X \mid Y) \) stated in (53),
- (d) follows from (87) and (273),
- (e) follows from the fact that the unsmooth conditional Rényi entropy \( \varepsilon \mapsto \tilde{H}_{\varepsilon}^\delta \) is nonincreasing in \( \varepsilon \in [0, 1] \), and
- (f) follows from Lemma 3.

This completes the proof of Lemma 4.

### Appendix G

#### Proof of Lemma 5

A direct calculation shows
\[
\mathbb{E}[K(X, Y)\rho] \overset{(a)}{\geq} \exp \left( \rho \tilde{H}_{1/(1 + \rho)}(X \mid Y) \right) \exp \left( - \rho \log R(\varepsilon, \kappa) \right)
\]

\[
\overset{(b)}{\geq} \exp \left( \rho \inf_{\delta : \mathcal{Y} \rightarrow [0, 1]} \tilde{H}_{1/(1 + \rho)}(X \mid Y) \right) \exp \left( - \rho \log R(\varepsilon, \kappa) \right)
\]
\[\varepsilon \exp \left( \rho \tilde{H}^{(i)}_{1/(1+p)}(X \mid Y) \right) \exp \left( -\rho \log R(\varepsilon, \kappa) \right) \]

where

- (a) follows as in Steps (a)–(c) of (275),
- (b) follows from (89) and (273),
- (c) follows from the fact that the unsmooth conditional Rényi entropy \( \varepsilon \mapsto H^c_{\alpha} \) is nonincreasing in \( \varepsilon \in [0,1] \), and
- (d) follows from (64).

This completes the proof of Lemma 5.

**Appendix H**

**Proof of Lemma 6**

For each \( y \in \mathcal{Y} \), choose two real parameters \( \eta(y) \geq 1 \) and \( 0 \leq \beta(y) < 1 \) so that

\[
\Pr\{ \kappa(X, Y) > \eta(Y) \mid Y = y \} \leq \beta(y) \Pr\{ \kappa(X, Y) = \eta(Y) \mid Y = y \} = \delta(y). \tag{277}
\]

Construct a deterministic map \( \varepsilon : X \times \mathcal{Y} \to [0,1] \) as

\[
\varepsilon(x, y) = \begin{cases} 
1 & \text{if } \kappa(x, y) < \eta(y), \\
\beta(y) & \text{if } \kappa(x, y) = \eta(y), \\
0 & \text{if } \kappa(x, y) > \eta(y).
\end{cases}
\tag{278}
\]

for each \( (x, y) \in X \times \mathcal{Y} \). Then, it is clear that

\[
\mathbb{E}[\varepsilon(X, Y) \mid Y = y] = \delta(y) \tag{279}
\]

for each \( y \in \mathcal{Y} \). Moreover, a direct calculation shows

\[
\begin{align*}
\mathbb{E}[K(X, Y)^p] & \geq \left( \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{x : x \in X, \kappa(x, y) < \eta(y)} P_{X|Y}(x \mid y) \kappa(x, y)^p + \beta(y) \eta(y)^p \right) \\
& \leq c^p \sum_{y \in \mathcal{Y}} P_Y(y) \left( \sum_{x : x \in X, \kappa(x, y) < \eta(y)} P_{X|Y}(x \mid y)^{1/(1+p)} + M(y)^{1/(1+p)} \right)^{1+p} \\
& \geq \left( \sum_{y \in \mathcal{Y}} P_Y(y) \exp \left( \rho \tilde{H}^{(i)}_{1/(1+p)}(P_{X|Y = y}) \right) \right)^{1+p} \\
& \geq \left( \sum_{y \in \mathcal{Y}} P_Y(y) \exp \left( \rho \tilde{H}^{(i)}_{1/(1+p)}(X \mid Y) \right) \right)^{1+p}.
\end{align*}
\tag{280}
\]

where

- (a) follows from (85) and (278),
- (b) follows from (94) and the definition of \( Q^{1/(1+p), \delta(i)}_{X|Y} \) stated in (93),
- (c) follows from Lemma 1, and
- (d) follows from the definition of \( \tilde{H}^{(i)}_{1/(1+p)}(X \mid Y) \) stated in (53).

This completes the proof of Lemma 6.

**Appendix I**

**Proof of Lemma 7**

Firstly, we shall verify the converse bound of Lemma 7, i.e., the left-hand inequality of (116). Consider a variable-length semi-stochastic code \( (F, g) \) such that \( C_g(X, Y, F) \) is prefix-free for every \( y \in \mathcal{Y} \) and

\[
\Pr\{ X \neq g(F(X, Y), Y) \} \leq \varepsilon. \tag{281}
\]

Construct another stochastic encoder \( F_0 : X \times \mathcal{Y} \to \{0,1\}^* \) as follows:

\[
F_0(x, y) = \begin{cases} 
\emptyset & \text{if } x \neq g(F(x, y), y), \\
F(x, y) & \text{if } x = g(F(x, y), y). 
\end{cases}
\tag{282}
\]
for each $x \in X$. It is clear that
\[ X \neq g(F_0(X,Y),Y) \implies X \neq g(F(X,Y),Y). \] (283)
Consider a collection $\{B(x,y)\}_{(x,y) \in X \times Y}$ of subsets of $\{0,1\}^*$ given as
\[
B(x,y) = \begin{cases} \{\varnothing\} & \text{if } x = g(\varnothing, y), \\
\{b \in \{0,1\}^* \setminus \{\varnothing\} \mid P(F_0(x,y) = b) > 0\} & \text{if } x \neq g(\varnothing, y).
\end{cases}
\] (284)
Furthermore, for each $(x,y) \in X \times Y$, choose a binary string $b(x,y) \in \{0,1\}^*$ so that
\[
b(x,y) \in \arg\min_{b \in B(x,y)} \ell(b),
\] (285)
where suppose that $b(x,y) = \varnothing$ if $B(x,y) = \emptyset$. Note that this map $b(\cdot, \cdot) : X \times Y \rightarrow \{0,1\}^*$ is deterministic. Now, construct another stochastic encoder $F_1 : X \times Y \rightarrow \{0,1\}^*$ so that
\[
F_1(x,y) = \begin{cases} \varnothing & \text{if } x \neq g(F_0(x,y),y), \\
b(x,y) & \text{if } x = g(F_0(x,y),y).
\end{cases}
\] (286)
for each $(x,y) \in X \times Y$. Then, it follows from (283) and (286) that
\[
\exp \left[ \hat{A}(X,Y,F,g \| \rho) \right] = \mathbb{E}[2^{\ell(F(X,Y),Y)}1_{\{x=g(F(X,Y),Y)\}}] \geq \mathbb{E}[2^{\ell(b(X,Y))}1_{\{x=g(F(X,Y),Y)\}}].
\] (287)
Choosing $\epsilon : X \times Y \rightarrow [0,1]$ so that
\[
\epsilon(x,y) = \mathbb{P}\{X \neq g(F(X,Y),Y) \mid (X,Y) = (x,y)\}
\] (288)
for each $(x,y) \in X \times Y$, we observe that
\[
2^{\ell(b(x,y))}1_{\{x=g(F(X,Y),y)\}} = \begin{cases} 2^{\ell(b(x,y))} & \text{with probability } 1 - \epsilon(x,y), \\
0 & \text{with probability } \epsilon(x,y).
\end{cases}
\] (289)
Therefore, since (281) implies that (87) holds, it follows from Lemma 4 and (287) that
\[
\frac{\hat{A}(X,Y,F,g \| \rho)}{\rho} \geq H^E_{1/(1+\rho)}(X \mid Y) - \log R(\epsilon,2^{\ell(b(\cdot,\cdot))})
\] (290)
Finally, since $C_y(X,Y,F)$ is prefix-free for every $y \in Y$, it follows from the Kraft–McMillan inequality that
\[
\sum_{x \in X : \epsilon(x,y) < 1} 2^{-\ell(b(x,y))} \leq \sum_{b \in C_y(X,Y,F)} 2^{-\ell(b)} \leq 1
\] (291)
for every $y \in Y$, which implies that
\[
R(\epsilon,2^{\ell(b(\cdot,\cdot))}) \leq 1.
\] (292)
This completes the proof of the left-hand inequality of (116).\footnote{Here, both constructed stochastic encoders $F_0$ and $F_1$ do not satisfy the prefix-free constraint, and it does not affect the proof.}
Secondly, we shall verify the achievability bound of Lemma 7, i.e., the right-hand inequality of (116). Fix an arbitrary small positive number $\zeta$. It follows from Lemma 3 that one can find a deterministic map $\delta(\cdot) \in E_0(\epsilon)$ so that
\[
H^E_{1/(1+\rho)}(X \mid Y) \geq H^E_{1/(1+\rho)}(X \mid Y) - \zeta.
\] (293)
Recall that for each $y \in Y$, the proper subset $A^{\delta(y)}_{X|Y=y}$ of $X$ is defined to satisfy (6) and (7), and the element $x^*(y)$ of $X$ is chosen as (91). Denote by
\[
\tilde{A}^{\delta(y)}_{X|Y=y} \doteq A^{\delta(y)}_{X|Y=y} \cup \{x^*(y)\}
\] (294)
for each $y \in Y$. Let $Z$ be a subset of $X \times Y$ given as
\[
Z = \{(x,y) \in X \times Y \mid x \in \tilde{A}^{\delta(y)}_{X|Y=y}\}
\] (295)
Based on the conditional distribution $Q^{(1/(1+\rho),\delta(\cdot))}_{X|Y}$ defined in (93), consider the Shannon code $f_{sh} : Z \rightarrow \{0,1\}^*$ satisfying
\[
\ell(f_{sh}(x,y)) = \left[ \log \frac{1}{Q^{(1/(1+\rho),\delta(\cdot))}_{X|Y}(x \mid y)} \right].
\] (296)
for every \((x, y) \in \mathcal{Z}\), where note that \(x \mapsto f_{\text{Sh}}(x, y)\) is prefix-free for every fixed \(y\). It follows from (296) that
\[
2^{\ell(f_{\text{Sh}}(x, y))} Q_x^{(1/(1+r))} \delta(x \mid y) < 2
\] (297)
for every \((x, y) \in \mathcal{Z}\). Fix a pair \((a, b) \in \mathcal{Z}\) arbitrarily. Now, construct a stochastic encoder \(F : \mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1\}^*\) so that
\[
F(x, y) = \begin{cases} 
  f_{\text{Sh}}(a, b) & \text{if } (x, y) \in (\mathcal{X} \times \mathcal{Y}) \setminus \mathcal{Z}, \\
  B^* & \text{if } x = x^*(y), \\
  f_{\text{Sh}}(x, y) & \text{if } x \in \mathcal{A}^{\delta^*(y)}
\end{cases}
\] (298)
for each \((x, y) \in \mathcal{X} \times \mathcal{Y}\), where \(B^*_y\) is a r.v. given as
\[
B^*_y = \begin{cases} 
  f_{\text{Sh}}(a, b) & \text{with probability } 1 - \frac{1 - \delta(y) - P_{X \mid Y}(\mathcal{A}^{\delta^*(y)}_{X \mid Y = y} \mid y)}{P_{X \mid Y}(x^*(y) \mid y)}, \\
  f_{\text{Sh}}(x^*(y), y) & \text{with probability } \frac{1 - \delta(y) - P_{X \mid Y}(\mathcal{A}^{\delta^*(y)}_{X \mid Y = y} \mid y)}{P_{X \mid Y}(x^*(y) \mid y)}
\end{cases}
\] (299)
for each \(y \in \mathcal{Y}\). Since the Shannon code \(x \mapsto f_{\text{Sh}}(x, y)\) is prefix-free for fixed \(y\), it is clear that \(C_y(X, Y, F)\) is prefix-free for every \(y \in \mathcal{Y}\). On the other hand, construct a deterministic decoder \(g : \{0, 1\}^* \times \mathcal{Y} \rightarrow \mathcal{X}\) so that
\[
g(b, y) = \begin{cases} 
  x & \text{if } x \in \mathcal{A}^{\delta^*(y)}_{X \mid Y = y} \text{ and } b = f_{\text{Sh}}(x, y), \\
  x^*(y) & \text{otherwise}
\end{cases}
\] (300)
for each \((b, y) \in \{0, 1\}^* \times \mathcal{Y}\). We observe that
\[
\mathbb{P}\{X \neq g(F(X, Y), Y)\} = 1 - \sum_{y \in \mathcal{Y}} P_Y(y) \mathbb{P}\{X = g(F(X, Y), Y) \mid Y = y\}
\]
\[
\overset{(a)}{=} 1 - \sum_{y \in \mathcal{Y}} P_Y(y) \left( \mathbb{P}\{X \in \mathcal{A}^{\delta^*(y)}_{X \mid Y = y} \mid Y = y\} + \frac{1 - \delta(y) - P_{X \mid Y}(\mathcal{A}^{\delta^*(y)}_{X \mid Y = y} \mid y)}{P_{X \mid Y}(x^*(y) \mid y)} \mathbb{P}\{X = x^*(y) \mid Y = y\} \right)
\]
\[
= \sum_{y \in \mathcal{Y}} P_Y(y) \delta^*(y)
\]
\[
\overset{(b)}{=} \varepsilon,
\] (301)
where
- (a) follows by the construction of \((F, g)\) stated in (298) and (300),
- (b) follows from the fact that \(\delta^*(\cdot) \in \mathcal{E}_0(\varepsilon)\); see Lemma 3.

Therefore, we have
\[
\bar{\Lambda}_{\text{avg}}(X, Y \parallel \rho, \varepsilon) \overset{(a)}{=} \bar{\Lambda}(X, Y, F, g \parallel \rho)
\]
\[
= \frac{1}{\rho} \log \mathbb{E}[2^{\rho F(X, Y)} \mathbb{1}_{\{X = g(F(X, Y), Y)\}}]
\]
\[
\overset{(b)}{=} \bar{H}^\rho(\cdot)(X \mid Y) + 1
\]
\[
\overset{(c)}{=} H^\rho_{1/(1+r)}(X \mid Y) + 1 + \zeta,
\] (302)
where
- (a) follows from (301) and the fact that \(C_y(X, Y, F)\) is prefix-free for every \(y \in \mathcal{Y}\),
- (b) follows from Lemma 6 and (297) with \(\kappa(x, y) = 2^{\ell(f_{\text{Sh}}(x, y))}\), and
- (c) follows from (293).

Since \(\zeta > 0\) is arbitrary, this completes the proof of Lemma 7.

**APPENDIX J**

**PROOF OF LEMMA 8**

The left-hand inequality of (117) is clear by the definitions of \(\Lambda_{\text{avg}}(X, Y \parallel \rho, \varepsilon)\) and \(\bar{\Lambda}_{\text{avg}}(X, Y \parallel \rho, \varepsilon)\) stated in (108) and (115), respectively.
It remains to verify the right-hand inequality of (117). Let ζ be an arbitrary positive number. By the definition of \( \tilde{\Lambda}_\text{avg}^\ast(X, Y \mid \rho, \varepsilon) \) stated in (115), one can choose a variable-length semi-stochastic code \((F, g)\) such that \( C_g(X, Y, F) \) is prefix-free for every \( y \in \mathcal{Y} \) and

\[
\rho \tilde{\Lambda}_\text{avg}^\ast(X, Y \mid \rho, \varepsilon) \geq \tilde{\Lambda}(X, Y, F, g \mid \rho) - \zeta, \tag{303}
\]

\[
\mathbb{P}\{X \neq g(F(X, Y), Y)\} \leq \varepsilon. \tag{304}
\]

Now, construct another stochastic encoder \( F' : X \times \mathcal{Y} \rightarrow \{0, 1\}^* \) so that

\[
F'(x, y) = \begin{cases} \hat{b}(x, y) & \text{if } x \neq g(F(x, y), y), \\ F(x, y) & \text{if } x = g(F(x, y), y) \end{cases}
\]

for each \( x \in \mathcal{X} \), where \( \hat{b}(x, y) \) is chosen so that

\[
\hat{b}(x, y) \in \arg \min_{b \in C_g(X, Y, F)} \ell(b). \tag{305}
\]

Since \( C_g(X, Y, F') \subset C_g(X, Y, F) \), it is clear that \( C_g(X, Y, F') \) is also prefix-free for every \( y \in \mathcal{Y} \). Moreover, we readily see that

\[
\mathbb{P}\{X \neq g(F'(X, Y), Y)\} = \mathbb{P}\{X \neq g(F(X, Y), Y)\} \leq \varepsilon. \tag{307}
\]

In addition, we see from (305)–(307) that

\[
\mathbb{E}[2^{\rho \ell(F'(X, Y))}] = \mathbb{E}[2^{\rho \ell(F(X, Y))}1_{\{X = g(F(X, Y), Y)\}}] + \mathbb{E}[2^{\rho \ell(\hat{b}(x, y))}1_{\{X \neq g(F(X, Y), Y)\}}] \\
\leq \frac{\mathbb{E}[2^{\rho \ell(F(X, Y))}1_{\{X = g(F(X, Y), Y)\}}]}{1 - \varepsilon} = \exp \left( \tilde{\Lambda}(X, Y, F, g \mid \rho) \left( \frac{1}{1 - \varepsilon} \right) \right), \tag{308}
\]

where the inequality follows from (304) and the fact that

\[
2^{\rho \ell(\hat{b}(x, y))} \leq \frac{\mathbb{E}[2^{\rho \ell(F(X, Y))}1_{\{X = g(F(X, Y), Y)\}}]}{1 - \varepsilon} \tag{309}
\]

for every \((x, y) \in \mathcal{X} \times \mathcal{Y}\). Therefore, we observe that

\[
\exp \left( \rho \Lambda^\ast_{\text{avg}}(X, Y \mid \rho, \varepsilon) \right) \overset{(a)}{\leq} \mathbb{E}[2^{\rho \ell(F'(X, Y))}] \overset{(b)}{\leq} \exp \left( \tilde{\Lambda}(X, Y, F, g \mid \rho) \left( \frac{1}{1 - \varepsilon} \right) \right) \overset{(c)}{\leq} \exp \left( \rho \tilde{\Lambda}_\text{avg}^\ast(X, Y \mid \rho, \varepsilon) + \zeta \left( \frac{1}{1 - \varepsilon} \right) \right) \tag{310}
\]

where

- (a) follows from (307) and the definition of \( \Lambda^\ast_{\text{avg}}(X, Y \mid \rho, \varepsilon) \) stated in (108),
- (b) follows from (308), and
- (d) follows from (303).

As \( \zeta > 0 \) is arbitrary, we obtain the right-hand inequality of (117) from (310). This completes the proof of Lemma 8. 

\[\text{APPENDIX K}\]

\text{PROOF OF LEMMA 9}\]

Firstly, we shall verify the converse bound of Lemma 9, i.e., the left-hand inequality of (121). Consider a variable-length semi-stochastic code \((F, g)\) such that \( C_g(X, Y, F) \) is prefix-free for every \( y \in \mathcal{Y} \) and

\[
\sup_{y \in \mathcal{Y}} \mathbb{P}\{X \neq g(F(X, Y), Y) \mid Y = y\} \leq \varepsilon. \tag{311}
\]

Consider the deterministic maps \( b : \mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1\}^* \) and \( \varepsilon : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1] \) as defined in (285) and (288), respectively. Since (311) implies that (89) holds, it follows from Lemma 5 and (287) that

\[
\tilde{\Lambda}(X, Y, F, g \mid \rho) \geq \tilde{H}^F_{\varepsilon/(1 + \rho)}(X \mid Y) - \log R(\varepsilon, 2^{\ell(b(\cdot, \cdot))}), \tag{312}
\]

which yields the left-hand inequality of (121) together with (292).
Finally, replacing the map $\delta^* : \mathcal{Y} \to [0,1]$ chosen in (293) by the constant $0 \leq \varepsilon < 1$, it can be verified by the same way as we did for (302) that
\[
\bar{\Lambda}_{\text{max}}^*(X,Y \| \rho, \varepsilon) < \bar{H}_{1/(1+\rho)}^r(X \mid Y) + 1 \\
= \bar{H}_{1/(1+\rho)}(X \mid Y) + 1, \tag{313}
\]
where the last inequality follows from (64). This completes the proof of Lemma 9.

\section*{Appendix L}
\section*{Proof of Lemma 10}

The left-hand inequality of (122) is clear by the definitions of $\Lambda_{\text{max}}^*(X,Y \| \rho, \varepsilon)$ and $\bar{\Lambda}_{\text{max}}^*(X,Y \| \rho, \varepsilon)$ stated in (109) and (120), respectively.

It thus remains to verify the right-hand inequality of (122). Let $\zeta$ be an arbitrary positive real number. By the definition of $\bar{\Lambda}_{\text{max}}^*(X,Y \| \rho, \varepsilon)$ stated in (120), one can choose a variable-length semi-stochastic code $(F,g)$ such that $C_{\rho}(X,Y,F)$ is prefix-free for every $y \in \mathcal{Y}$ and
\[
\rho \bar{\Lambda}_{\text{max}}^*(X,Y \| \rho, \varepsilon) \geq \bar{\Lambda}(X,Y,F,g \mid \rho) - \zeta, \tag{314}
\]
\[
\sup_{y \in \mathcal{Y}} \mathbb{P}(X \neq g(F(X,Y),Y) \mid Y = y) \leq \varepsilon. \tag{315}
\]
Consider a stochastic encoder $F^* : X \times \mathcal{Y} \to \{0,1\}^*$ and a deterministic map $\bar{b} : X \times \mathcal{Y} \to \{0,1\}^*$ defined as in (305) and (306), respectively. Then, the codeword set $C_{\rho}(X,Y,F)$ is prefix-free for every $y \in \mathcal{Y}$, and it follows from (315) that
\[
\sup_{y \in \mathcal{Y}} \mathbb{P}(X \neq g(F^*(X,Y),Y) \mid Y = y) \leq \varepsilon. \tag{316}
\]
Hence, similar to (310), we obtain
\[
\exp \left( \rho \Lambda_{\text{max}}^*(X,Y \| \rho, \varepsilon) \right) \leq \exp \left( \bar{\Lambda}(X,Y,F,g \mid \rho) \right) \frac{1}{1 - \varepsilon}
\leq \exp \left( \rho \bar{\Lambda}_{\text{max}}^*(X,Y \| \rho, \varepsilon) + \zeta \right) \frac{1}{1 - \varepsilon}, \tag{317}
\]
where the last inequality follows from (314). This completes the proof of Lemma 10.

\section*{Appendix M}
\section*{Proof of Lemma 11}

Equation (153) can be verified as
\[
\mathbb{P}(\tilde{G}_{\text{avg}}^*(X,Y) = 0) = \sum_{y \in \mathcal{Y}} P_Y(y) \mathbb{P}(\tilde{G}_{\text{avg}}^*(X,Y) = 0 \mid Y = y)
= \sum_{y \in \mathcal{Y}} P_Y(y) \left( 1 - \sum_{k=1}^{\infty} P_{X \mid Y}(\xi_y(k) \mid y) \prod_{j=1}^{k} \left( 1 - \pi_{\text{avg}}(j,y) \right) \right)
\leq \sum_{y \in \mathcal{Y}} P_Y(y) \left( 1 - \sum_{k=1}^{j} P_{X \mid Y}(\xi_y(k) \mid y) - \xi \right)
\leq \varepsilon, \tag{318}
\]
where
- (a) follows from (140); see also [11, Equation (29)],
- (b) follows from (151), and
- (c) follows from (149).

Consider a guessing strategy $(g,\pi)$, and the giving-up guessing function $\tilde{G} : X \times \mathcal{Y} \to \mathbb{N} \cup \{0\}$ induced by $(g,\pi)$. Suppose that
\[
\mathbb{P}(\tilde{G}(X,Y) = 0) \leq \varepsilon. \tag{319}
\]
To prove (152), it suffices to show that
\[
\mathbb{E}[\tilde{G}(X,Y)^g] \geq \mathbb{E}[\tilde{G}_{\text{avg}}^*(X,Y)^g] \tag{320}
\]
for every positive real number \( \rho \). Now, we shall verify that
\[
\mathbb{P}\{\tilde{G}(X, Y) \geq k\} \geq \mathbb{P}\{\tilde{G}^\ast_{\text{avg}}(X, Y) \geq k\}
\] (321)
for every positive integer \( k \). It follows from (153) and (319) that
\[
\mathbb{P}\{\tilde{G}(X, Y) = 0\} \leq \mathbb{P}\{\tilde{G}^\ast_{\text{avg}}(X, Y) = 0\}.
\] (322)
In addition, since \( x \mapsto g(x, y) \) is bijective for each \( y \in \mathcal{Y} \) and \( \varsigma_y : \mathcal{N} \to \mathcal{X} \) rearranges the probability masses in \( P_{X|Y}(\cdot | y) \) in nonincreasing order (see (147)), we see that
\[
\sum_{l=1}^{k} \sum_{y \in \mathcal{Y}} P_Y(y) \mathbb{P}\{g(X, Y) = l | Y = y\} \leq \sum_{l=1}^{k} \sum_{y \in \mathcal{Y}} P_Y(y) P_{X|Y}(\varsigma_y(l) | y)
\] (323)
for every positive integer \( k \). Thus, we observe that
\[
\mathbb{P}\{\tilde{G}(X, Y) \leq k\} = \mathbb{P}\{\tilde{G}(X, Y) = 0\} + \sum_{l=1}^{k} \mathbb{P}\{\tilde{G}(X, Y) = l\}
\]
\[
\leq \mathbb{P}\{\tilde{G}(X, Y) = 0\} + \sum_{l=1}^{k} \mathbb{P}\{g(X, Y) = l\}
\]
\[
(a) \leq \mathbb{P}\{\tilde{G}(X, Y) = 0\} + \sum_{l=1}^{k} \sum_{y \in \mathcal{Y}} P_Y(y) P_{X|Y}(\varsigma_y(l) | y)
\]
\[
(b) \leq \mathbb{P}\{\tilde{G}(X, Y) = 0\} + \mathbb{P}\{1 \leq \tilde{G}^\ast_{\text{avg}}(X, Y) \leq k\}
\]
\[
(c) \leq \mathbb{P}\{\tilde{G}^\ast_{\text{avg}}(X, Y) \leq k\}
\] (324)
for every \( 0 \leq k \leq J \), where
- (a) follows from (323),
- (b) follows from (140), (150), and (151), and
- (c) follows from (153).

In addition, we get
\[
\mathbb{P}\{\tilde{G}^\ast_{\text{avg}}(X, Y) \leq J + 1\} = \sum_{l=0}^{J+1} \mathbb{P}\{\tilde{G}^\ast_{\text{avg}}(X, Y) = l\}
\]
\[
(a) = e + \sum_{l=0}^{J+1} \mathbb{P}\{\tilde{G}^\ast_{\text{avg}}(X, Y) = l\}
\]
\[
(b) = e + \sum_{l=0}^{J+1} \sum_{y \in \mathcal{Y}} P_Y(y) P_{X|Y}(\varsigma_y(l) | y) \prod_{j=1}^{l} \left(1 - \pi^\ast_{\text{avg}}(j, y)\right)
\]
\[
(c) = e + \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{l=0}^{J} P_{X|Y}(\varsigma_y(l) | y) + \xi
\]
\[
(d) = 1,
\] (325)
where
- (a) follows from (153),
- (b) follows from (140), (150), and (151),
- (c) follows from (151), and
- (d) follows from (149).

Combining (324) and (325), we have that (321) holds. Therefore, we obtain
\[
\mathbb{E}[\tilde{G}(X, Y)\rho] = \sum_{k=1}^{\infty} \left(k^\rho - (k-1)^\rho\right) \mathbb{P}\{\tilde{G}(X, Y) \geq k\}
\]
\[
\geq \sum_{k=1}^{\infty} \left(k^\rho - (k-1)^\rho\right) \mathbb{P}\{\tilde{G}^\ast_{\text{avg}}(X, Y) \geq k\}
\]
\[
= \mathbb{E}[\tilde{G}^\ast_{\text{avg}}(X, Y)\rho],
\] (326)
proving (152). This completes the proof of Lemma 11.
Appendix N

Proof of Theorem 6

A. Converse Part

To prove the left-hand inequality of (146), it suffices to consider the optimal guessing strategy $(\mathcal{G}, \pi_{\text{avg}}^*)$ given in Lemma 11. Choose the map $\delta : \mathcal{Y} \rightarrow [0, 1]$ given by the formula

$$
\delta(y) = \left[ - \sum_{k=1}^{J} P_{X|Y}(x_k | y) \sum_{l=1}^{J} P_Y(b) + \sum_{i=1}^{J} P_{X|Y}(x_{k} | y) + \varepsilon \right]_+
$$

for each $y \in \mathcal{Y}$, where $J$ is given in (148). We observe that

$$
\log(J + 1) \leq \frac{1}{\varepsilon} \sum_{y \in \mathcal{Y}} P_Y(y) \delta(y) \log(J + 1)
$$

$$
\leq \frac{1}{\varepsilon} \sum_{y \in \mathcal{Y}} P_Y(y) \delta(y) \inf \left\{ R > 0 \left| \mathbb{P} \left( \log \frac{1}{P_{X|Y}(X | Y)} > R \right| Y = y \right\} \leq \delta(y) \right\}
$$

$$
\leq \frac{1}{\varepsilon} \sum_{y \in \mathcal{Y}} P_Y(y) \delta(y) \inf \left\{ R > 0 \left| \frac{H(P_{X|Y} = y)}{R} \leq \delta(y) \right\}
$$

$$
\leq \frac{1}{\varepsilon} \sum_{y \in \mathcal{Y}} P_Y(y) H(P_{X|Y} = y)
$$

$$
= \frac{H(X | Y)}{\varepsilon},
$$

where

- (a) follows by the choice of $\delta : \mathcal{Y} \rightarrow [0, 1]$ stated in (327),

- (b) follows from the fact that

$$
\delta(y) < \sum_{k=J+1}^{\infty} P_{X|Y}(x_k | y)
$$

for each $y \in \mathcal{Y}$, and

- (c) follows by Markov’s inequality.

Letting

$$
\kappa(x, y) = \mathcal{G}^*(x, y),
$$

$$
\epsilon(x, y) = \prod_{k=1}^{\infty} \left( 1 - \pi_{\text{avg}}^*(k, y) \right)
$$

for each $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we have

$$
\frac{1}{\rho} \log \mathbb{E} \left[ \tilde{G}_{\text{avg}}^*(X, Y)^\rho \right] \overset{(a)}{\geq} H_0^* (X | Y) - \log \left( \sum_{k=1}^{\infty} \frac{1}{k} \right)
$$

$$
\overset{(b)}{\geq} H_0^* (X | Y) - \log(1 + \log(J + 1))
$$

$$
\overset{(c)}{\geq} H_0^* (X | Y) - \log \left( 1 + \frac{H(X | Y)}{\varepsilon} \right),
$$

where

- (a) follows from Lemma 4,

- (b) follows from the fact that

$$
\sum_{k=1}^{m} \frac{1}{k} \leq 1 + \log m,
$$

and

- (c) follows from (328).

This completes the proof of the converse bound of Theorem 6, i.e., the left-hand inequality of (146).

\[ \square \]
B. Achievability Part

We shall verify the right-hand inequality of (146). Fix a positive real number $\zeta$ arbitrarily, and choose a map $\delta^* : \mathcal{Y} \to [0, 1]$ by the same manner as (293). For each $y \in \mathcal{Y}$, choose an integer $\tilde{J}(y)$ so that

$$\tilde{J}(y) = \sup \left\{ j \geq 0 \left| \sum_{k=1}^{j} P_{X|Y}(\zeta_y(k) \mid y) < 1 - \delta^*(y) \right. \right\},$$

(334)

and choose a real number $\tilde{M}(y)$ so that

$$\tilde{M}(y) = 1 - \delta^*(y) - \sum_{k=1}^{\tilde{J}(y)} P_{X|Y}(\zeta_y(k) \mid y).$$

(335)

Consider the optimal guessing function $g^* : X \times \mathcal{Y} \to \mathbb{N}$ given in (150). If $\zeta_y^{-1}(x) \leq \tilde{J}(y)$, then

$$g^*(x, y) = \sum_{k=1}^{\zeta_y^{-1}(x)} 1$$

$$\leq \sum_{k=1}^{\tilde{J}(y)} \left( \frac{P_{X|Y}(\zeta_y(k) \mid y)}{P_{X|Y}(x \mid y)} \right)^{1/(1+p)}$$

$$\leq \sum_{k=1}^{\tilde{J}(y)} \left( \frac{P_{X|Y}(\zeta_y(k) \mid y)}{P_{X|Y}(x \mid y)} \right)^{1/(1+p)} + \left( \frac{\tilde{M}(y)}{P_{X|Y}(x \mid y)} \right)^{1/(1+p)}$$

$$= \frac{1}{Q_{X|Y}^{(1/(1+p), \zeta^*(-)(\cdot))(x \mid y)}}.$$  

(336)

where the last equality follows from (93). In addition, if $\zeta_y^{-1}(x) = \tilde{J}(y) + 1$, then

$$g^*(x, y) = \sum_{k=1}^{\tilde{J}(y)+1} 1$$

$$\leq \sum_{k=1}^{\tilde{J}(y)} \left( \frac{P_{X|Y}(\zeta_y(k) \mid y)}{\tilde{M}(y)} \right)^{1/(1+p)} + \left( \frac{\tilde{M}(y)}{\tilde{M}(y)} \right)^{1/(1+p)}$$

$$= \frac{1}{Q_{X|Y}^{(1/(1+p), \zeta^*(-)(\cdot))(x \mid y)}}.$$  

(337)

Therefore, noting that $Q_{X|Y}^{(1/(1+p), \zeta^*(-)(\cdot))(x \mid y)} = 0$ if $\zeta_y^{-1}(x) \geq \tilde{J}(y) + 2$, we observe that

$$g^*(x, y) Q_{X|Y}^{(1/(1+p), \zeta^*(-)(\cdot))(x \mid y)} \leq 1$$

(338)

for every $y \in \mathcal{Y}$.

Given a deterministic map $\pi : X \times \mathcal{Y} \to [0, 1]$, construct a deterministic map $\epsilon : X \times \mathcal{Y} \to [0, 1]$ so that

$$\epsilon(x, y) := 1 - \prod_{k=1}^{\zeta_y^{-1}(x)} \left( 1 - \pi(k, y) \right).$$  

(339)

Then, the giving-up guessing function $G : X \times \mathcal{Y} \to \{0, 1\}$ induced by $(g^*, \pi)$ can be written as

$$G(x, y) = \begin{cases} 
    g(x, y) & \text{with probability } 1 - \epsilon(x, y), \\
    0 & \text{with probability } \epsilon(x, y).
\end{cases}$$  

(340)
for each \((x, y) \in \mathcal{X} \times \mathcal{Y}\). Therefore, it follows from Lemma 6 and (338) that there exists a giving-up policy \(\pi^* : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1]\) such that the guessing strategy \((g^*, \pi^*)\) induces \(\tilde{G} : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{N} \cup \{0\}\) that satisfies
\[
P\{\tilde{G}(X, Y) = 0\} = \sum_{y \in \mathcal{Y}} P_Y(y) P\{\tilde{G}(X, Y) = 0 \mid Y = y\}
\]
\[
\leq \sum_{y \in \mathcal{Y}} P_Y(y) \delta^*(y)
\]
\[
\leq \epsilon,
\]
\[
\frac{1}{\rho} \log \mathbb{E}[\tilde{G}(X, Y)^\rho] \leq H_{1/\rho}^\rho(X \mid Y) + \zeta,
\]
\[(342)\]
where
- (a) follows from (95) of Lemma 6,
- (b) follows from the fact that \(\delta^*(\cdot) \in \mathcal{E}_0(\epsilon)\); see Lemma 3,
- (c) follows from (96) of Lemma 6, and
- (d) follows by the choice of \(\delta^*(\cdot)\) stated in (293).

As \(\zeta > 0\) is arbitrary, this proves the right-hand inequality of (146), completing the proof of the achievability bound of Theorem 6.

\[\hfill\]

**APPENDIX O**

**PROOF OF LEMMA 12**

Equation (159) can be verified by the same way as (318). Consider a guessing strategy \((g, \pi)\), and the giving-up guessing function \(\tilde{G} : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{N} \cup \{0\}\) induced by \((g, \pi)\). Suppose that
\[
P\{\tilde{G}(X, Y) = 0 \mid Y = y\} \leq \epsilon
\]
\[(343)\]
for every \(y \in \mathcal{Y}\). To prove (158), it suffices to show that
\[
\mathbb{E}[\tilde{G}(X, Y)^\rho] \geq \mathbb{E}[\tilde{G}^*_{\max}(X, Y)^\rho]
\]
\[(344)\]
for every positive real number \(\rho\). Since \(x \mapsto g(x, y)\) is bijective for each \(y \in \mathcal{Y}\) and \(\varsigma_y : \mathbb{N} \rightarrow \mathcal{X}\) rearranges the probability masses in \(P_{X|Y} (\cdot \mid y)\) in nonincreasing order (see (147)), we see that
\[
\sum_{l=1}^k P\{g(X, Y) = l \mid Y = y\} \leq \sum_{l=1}^k P_{X|Y}(\varsigma_y(l) \mid y)
\]
\[(345)\]
for every \(y \in \mathcal{Y}\). Thus, in the same way as we proved (321), we may observe that
\[
P\{\tilde{G}(X, Y) \geq k \mid Y = y\} \geq P\{\tilde{G}^*_{\max}(X, Y) \geq k \mid Y = y\}
\]
\[(346)\]
for every \(y \in \mathcal{Y}\) and every positive integer \(k\). Therefore, we obtain
\[
\mathbb{E}[\tilde{G}(X, Y)^\rho \mid Y = y] = \sum_{k=1}^\infty \left(k^\rho - (k-1)^\rho \right) P\{\tilde{G}(X, Y) \geq k \mid Y = y\}
\]
\[
\geq \sum_{k=1}^\infty \left(k^\rho - (k-1)^\rho \right) P\{\tilde{G}^*_{\max}(X, Y) \geq k \mid Y = y\}
\]
\[
= \mathbb{E}[\tilde{G}^*_{\max}(X, Y)^\rho \mid Y = y]
\]
\[(347)\]
for every \(y \in \mathcal{Y}\), proving (344). Therefore, we have (152). This completes the proof of Lemma 12.
APPENDIX P
PROOF OF THEOREM 7

A. Converse Part

To prove the left-hand inequality of (154), it suffices to consider the optimal guessing strategy \((g^*, \pi_{\text{max}}^*)\) given in Lemma 12. We observe that

\[
\log(J(y) + 1) \leq \inf \left\{ R > 0 \left| \mathbb{P}\left\{ \log \frac{1}{P_{X|Y}(X|Y)} > R \right| Y = y \right\} \leq \epsilon \right\}
\]

for every \(y \in \mathcal{Y}\), where

- (a) follows from the fact that

\[
\epsilon < \sum_{k=J(y)+1}^{\infty} P_{X|Y}(\mathcal{S}_y(k) | y)
\]

for each \(y \in \mathcal{Y}\), and

- (b) follows by Markov’s inequality.

Letting

\[
\kappa(x, y) = g^*(x, y), \quad \epsilon(x, y) = \prod_{k=1}^{J(y)+1} \left( 1 - \pi_{\text{max}}^*(k, y) \right)
\]

for each \((x, y) \in X \times \mathcal{Y}\), we have

\[
\frac{1}{\rho} \log \mathbb{E}[\hat{G}_{\text{max}}(X, Y)^{\rho}] \overset{(a)}{\geq} \tilde{H}^\epsilon(X | Y) - \sup_{y \in \mathcal{Y}} \log \left( \sum_{k=1}^{J(y)+1} \frac{1}{k} \right)
\]

\[
\overset{(b)}{\geq} \tilde{H}^\epsilon(X | Y) - \sup_{y \in \mathcal{Y}} \log(1 + \log(J(y) + 1))
\]

\[
\overset{(c)}{\geq} \tilde{H}^\epsilon(X | Y) - \log \left( 1 + \frac{\sup_{y \in \mathcal{Y}} H(P_{X|Y} | y)}{\epsilon} \right),
\]

where

- (a) follows from Lemma 5,
- (b) follows from (333), and
- (c) follows from (348).

This completes the proof of the converse bound of Theorem 7, i.e., the left-hand inequality of (154).

B. Achievability Part

We shall verify the right-hand inequality of (154). Choose a real number \(\tilde{M}(y)\) so that

\[
\tilde{M}(y) = 1 - \epsilon - \sum_{k=1}^{J(y)} P_{X|Y}(\mathcal{S}_y(k) | y).
\]

(353)
Consider the optimal guessing function \( g^* : X \times Y \to \mathbb{N} \) given in (150). If \( \varsigma_{g^*}^{-1}(x) \leq J(y) \), then

\[
g^*(x, y) = \sum_{k=1}^{\varsigma_{g^*}^{-1}(x)} 1 \leq \sum_{k=1}^{\varsigma_{g^*}^{-1}(x)} \left( \frac{P_X Y (\varsigma_y (k) \mid y)}{P_X Y (x \mid y)} \right)^{1/(1+p)} \leq \sum_{k=1}^{J(y)} \left( \frac{P_X Y (\varsigma_y (k) \mid y)}{P_X Y (x \mid y)} \right)^{1/(1+p)} + \left( \frac{\bar{M}(y)}{\bar{M}(y)} \right)^{1/(1+p)} = \frac{1}{Q_{X|Y}^{(1/(1+p), \epsilon)}(x \mid y)},
\]

where the last equality follows from (93). In addition, if \( \varsigma_{g^*}^{-1}(x) = J(y) + 1 \), then

\[
g^*(x, y) = \sum_{k=1}^{J(y)+1} 1 \leq \sum_{k=1}^{J(y)} \left( \frac{P_X Y (\varsigma_y (k) \mid y)}{\bar{M}(y)} \right)^{1/(1+p)} + \left( \frac{\bar{M}(y)}{\bar{M}(y)} \right)^{1/(1+p)} = \frac{1}{Q_{X|Y}^{(1/(1+p), \epsilon)}(x \mid y)}.
\]

Therefore, noting that \( Q_{X|Y}^{(1/(1+p), \epsilon)}(x \mid y) = 0 \) if \( \varsigma_{g^*}^{-1}(x) \geq J(y) + 2 \), we observe that

\[
g^*(x, y) Q_{X|Y}^{(1/(1+p), \epsilon)}(x \mid y) \leq 1
\]

for every \( y \in Y \).

Given a deterministic map \( \pi : X \times Y \to [0, 1] \), construct another deterministic map \( \epsilon : X \times Y \to [0, 1] \) so that

\[
\epsilon(x, y) := 1 - \prod_{k=1}^{\varsigma_{g^*}^{-1}(x)} \left( 1 - \pi(k, y) \right).
\]

Then, the giving-up guessing function \( \bar{G} : X \times Y \to \mathbb{N} \cup \{0\} \) induced by \((g^*, \pi)\) can be written as

\[
\bar{G}(x, y) = \begin{cases} g(x, y) & \text{with probability } 1 - \epsilon(x, y), \\ 0 & \text{with probability } \epsilon(x, y). \end{cases}
\]

for each \((x, y) \in X \times Y \). Therefore, it follows from Lemma 6 and (356) that there exists a giving-up policy \( \pi^* : X \times Y \to [0, 1] \) such that the guessing strategy \((g^*, \pi^*)\) induces \( \bar{G} : X \times Y \to \mathbb{N} \cup \{0\} \) satisfying

\[
P\{\bar{G}(X, Y) = 0 \mid Y = y\} \overset{(a)}{=} \epsilon
\]

for every \( y \in Y \), and

\[
\frac{1}{\rho} \log \mathbb{E}[\bar{G}(X, Y)^\rho] \overset{(b)}{=} \bar{H}^{\epsilon \rho}_{1/(1+p)}(X \mid Y) \overset{(c)}{=} \bar{H}_{1/(1+p)}^\rho(X \mid Y),
\]

where

- (a) follows from (95) of Lemma 6,
- (b) follows from (96) of Lemma 6, and
- (d) follows from (64).

This proves the right-hand inequality of (146), completing the proof of the achievability bound of Theorem 6. ■
APPENDIX Q

PROOF OF (179)

Consider an assignment function \( f : X \times Y \rightarrow \{0, 1, 2, \ldots, M\} \) and a stochastic map \( \map : 2^X \times Y \rightarrow 2^X \) satisfying (172) and

\[
P\{X \not\in \map(f(X), Y)\} \leq \varepsilon.
\] (361)

For each \( y \in Y \), denote by \( \{\mathcal{L}(m, y)\}_{m=1}^M \) the sub-partition of \( X \) induced by the assignment \( x \mapsto f(x, y) \); see (171). Define

\[
\mathcal{L}(y) := \bigcup_{m=1}^M \mathcal{L}(m, y)
\] (362)

for each \( y \in Y \).

Consider the stochastic sub-partition \( \mathcal{L} : \{0, 1, 2, \ldots, M\} \times Y \rightarrow 2^X \) induced by the pair \((f, \map)\); see (173). Since

\[
m_1 \neq m_2 \implies \mathcal{L}(m_1, y) \cap \mathcal{L}(m_2, y) = \emptyset
\] (363)

and

\[
\mathbb{E}[|\mathcal{L}(f(X), Y)| \mid Y = y] = \sum_{m=1}^M P\{f(X) = \map(f(X), Y) = m \mid Y = y\} |\mathcal{L}(m, y)|
\] (364)

for every \( y \in Y \), we see that

\[
\min_{m \in \{1, \ldots, M\} : \mathbb{P}\{f(X) = \map(f(X), Y) = m \mid Y = y\} > 0} |\mathcal{L}(m, y)| \leq \mathbb{E}[|\mathcal{L}(f(X), Y)| \mid Y = y] \leq |\mathcal{L}(y)|
\] (365)

for every \( y \in Y \). Thus, the task sub-partitioning \( \rho \)-th moment \( \mathbb{E}[|\mathcal{L}(f(X), Y)|^\rho] \) is finite if and only if \( \mathcal{L}(y) \) is finite for every \( y \in Y \). Therefore, to prove the converse bound stated in (179), it suffices to assume that \( \mathcal{L}(y) \) is finite for every \( y \in Y \).

Since \( \mathcal{L}(y) \) is finite, it follows from [6, Proposition III.1] that for every \( y \in Y \),

\[
\sum_{x \in \mathcal{L}(y) : \mathbb{P}\{X \not\in \map(f(X), Y) \mid (X, Y) = (x, y)\} < 1} \frac{1}{|\mathcal{L}(f(x, y), y)|} \leq M
\] (366)

Choose two deterministic maps \( \epsilon : X \times Y \rightarrow [0, 1] \) and \( \kappa : X \times Y \rightarrow (0, \infty) \) so that

\[
\epsilon(x, y) = \mathbb{P}\{X \not\in \map(f(X), Y) \mid (X, Y) = (x, y)\},
\] (367)

\[
\kappa(x, y) = \begin{cases} 1 & \text{if } f(x, y) = 0, \\ \left|\mathcal{L}(f(x, y), y)\right| & \text{if } f(x, y) \neq 0, \end{cases}
\] (368)

for each \((x, y) \in X \times Y\). Then, it follows from (361) that

\[
\mathbb{E}[\epsilon(X, Y)] \leq \varepsilon.
\] (369)

Moreover, we have

\[
\frac{1}{\rho} \log \mathbb{E}[|\mathcal{L}(f(X, Y)|^\rho)] \overset{(a)}{=} \frac{1}{\rho} \log \mathbb{E}[K(X, Y)^\rho]
\]

\[
\overset{(b)}{\geq} H_{1/(1+\rho)}^\epsilon(X \mid Y) - \sup_{y \in Y} \log \left( \sum_{x \in \mathcal{L}(y) : \mathbb{P}\{X \not\in L(f(X, Y), Y) \mid (X, Y) = (x, y)\} < 1} \frac{1}{|\mathcal{L}(f(x, y), y)|} \right)
\]

\[
\overset{(c)}{\geq} H_{1/(1+\rho)}^\epsilon(X \mid Y) - \log M
\] (370)

- (a) follows by the definition of \( K : X \times Y \rightarrow [0, \infty) \) stated in (85),
- (b) follows from Lemma 4, and
- (c) follows from (366).

This completes the proof of (179).
In the proof, we employ the following technical result:

**Lemma 15** (Bunte and Lapidoth [6, Proposition III.2]). Let $S$ be a finite set, $\lambda : S \to \mathbb{N} \cup \{\infty\}$ a function, and $M$ a positive integer. If

$$M \geq 2 \sum_{s \in S: \lambda(s) \leq \infty} \frac{1}{\lambda(s)} + \log |S| + 2,$$

then there exists a partition $\{L_m\}_{m=1}^M$ of $S$ such that

$$s \in L_m \implies |L_m| \leq \lambda(s).$$

For each $y \in \mathcal{Y}$, define

$$S_y := \{s_y(k) \mid 1 \leq k \leq J + 1\},$$

where the bijection $s_y : \mathbb{N} \to \mathcal{X}$ and the number $J$ are defined in (147) and (148), respectively. Since $|S_y| = J + 1$, it follows from (328) that

$$\log |S_y| \leq \frac{H(X \mid Y)}{\varepsilon}$$

for every $y \in \mathcal{Y}$. Letting $\delta : \mathcal{Y} \to [0, 1]$ be a deterministic map given as

$$\delta(y) = 1 - \left( \sum_{k=1}^J P_{X \mid Y}(s_y^{-1}(k) \mid y) + \nu P_{X \mid Y}(s_y^{-1}(J + 1) \mid y) \right)$$

for each $y \in \mathcal{Y}$, it follows by the definition of $Q_{X \mid Y}^{(1/(1+\nu), \delta(-))}$ stated in (93) that

$$Q_{X \mid Y}^{(1/(1+\nu), \delta(-))}(x \mid y) = \begin{cases} 
\frac{P_{X \mid Y}(x \mid y)^{1/(1+\nu)}}{\sum_{k=1}^J P_{X \mid Y}(s_y^{-1}(k) \mid y)^{1/(1+\nu)} + \nu^{1/(1+\nu)} P_{X \mid Y}(s_y^{-1}(J + 1) \mid y)^{1/(1+\nu)}} & \text{if } 1 \leq s_y^{-1}(x) \leq J, \\
\frac{\nu^{1/(1+\nu)} P_{X \mid Y}(s_y^{-1}(J + 1) \mid y)^{1/(1+\nu)}}{\sum_{k=1}^J P_{X \mid Y}(s_y^{-1}(k) \mid y)^{1/(1+\nu)} + \nu^{1/(1+\nu)} P_{X \mid Y}(s_y^{-1}(J + 1) \mid y)^{1/(1+\nu)}} & \text{if } s_y^{-1}(x) = J + 1, \\
0 & \text{if } J + 2 \leq s_y^{-1}(x) < \infty 
\end{cases}$$

for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Note that

$$Q_{X \mid Y}^{(1/(1+\nu), \delta(-))}(x \mid y) > 0 \implies x \in S_y.$$  

In addition, for each $y \in \mathcal{Y}$, define the function $\lambda_y : \mathcal{X} \to \mathbb{N} \cup \{\infty\}$ by

$$\lambda_y(x) := \begin{cases} 
2 \varepsilon & \text{if } Q_{X \mid Y}^{(1/(1+\nu), \delta(-))}(x \mid y) > 0, \\
\infty & \text{if } Q_{X \mid Y}^{(1/(1+\nu), \delta(-))}(x \mid y) = 0
\end{cases}$$

where $M$ is a positive integer satisfying (183). Then, a direct calculation shows

$$\sum_{x \in \mathcal{X}: \lambda_y(x) < \infty} \frac{1}{\lambda_y(x)} = \sum_{x \in S_y: \lambda_y(x) < \infty} \frac{1}{\lambda_y(x)} \leq \frac{\varepsilon (M - 2) - H(X \mid Y)}{2 \varepsilon} \sum_{x \in S_y} Q_{X \mid Y}^{(1/(1+\nu), \delta(-))}(x \mid y)$$

for every $y \in \mathcal{Y}$. Therefore, it follows from Lemma 15 and (374) that there exists an assignment function $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1, 2, \ldots, M\}$ satisfying

$$1 \leq s_y^{-1}(x) \leq J + 1 \iff f(x, y) \neq 0$$
and
\[ |L(f(x, y))| \leq \lambda_y(x) \quad (381) \]
for every \((x, y) \in X \times Y\), provided that \((183)\) holds, where \(L : \{0, 1, 2, \ldots, M\} \times Y \to 2^X\) is the sub-partition induced by \(f\); see \((171)\).

On the other hand, consider a stochastic map \(E : 2^X \times Y \to 2^X\) satisfying \((172)\) and
\[ \mathbb{P}(E(L(f(c_y(k), y), y)) = 0) = \begin{cases} 0 & \text{if } 0 \leq k \leq J, \\ 1 - \nu & \text{if } k = J + 1, \\ 1 & \text{if } J + 2 \leq k < \infty \end{cases} \quad (382) \]
for each \((k, y) \in \mathbb{N} \times Y\), where \(\nu\) is defined in \((181)\). We see from \((382)\) that
\[ \mathbb{P}(X \notin L(f(X, Y), Y) | (X, Y) = (x, y)) = \begin{cases} 0 & \text{if } 0 \leq s_y^{-1}(x) \leq J, \\ 1 - \nu & \text{if } s_y^{-1}(x) = J + 1, \\ 1 & \text{if } J + 2 \leq s_y^{-1}(x) < \infty \end{cases} \quad (383) \]
for every \((x, y) \in X \times Y\), where the stochastic sub-partition \(L : \{0, 1, 2, \ldots, M\} \times Y \to 2^X\) induced by the pair \((f, E)\) is defined in \((173)\). Therefore, we have
\[ \mathbb{E}[|L(f(X, Y), Y)|^\rho] \leq^{(a)} \sum_{y \in Y} P_Y(y) \left( \sum_{k=1}^{J} P_{X|Y}(\varsigma_y(k) | y) |L(f(\varsigma_y(k), y), y)|^\rho + \nu P_{X|Y}(\varsigma_y(J + 1) | y) |L(f(\varsigma_y(J + 1), y), y)|^\rho \right) \]
\[ \leq^{(b)} \sum_{y \in Y} P_Y(y) \left( \sum_{k=1}^{J} P_{X|Y}(\varsigma_y(k) | y) \left( \frac{2\epsilon}{(\epsilon - M - 2 - H(X | Y)) Q_X^{\rho/(1+p)}(\varsigma_y(k) | y)} \right)^\rho + \nu P_{X|Y}(\varsigma_y(J + 1) | y) \left( \frac{2\epsilon}{(\epsilon - M - 2 - H(X | Y)) Q_X^{\rho/(1+p)}(\varsigma_y(J + 1) | y)} \right)^\rho \right) \]
\[ \leq^{(c)} \left( \frac{4\epsilon}{\epsilon - M - 2 - H(X | Y)} \right)^\rho \sum_{y \in Y} P_Y(y) \left( \sum_{k=1}^{J} P_{X|Y}(\varsigma_y(k) | y) \left( \frac{1}{Q_X^{\rho/(1+p)}(\varsigma_y(k) | y)} \right)^\rho \right) + (1 - \epsilon) \]
\[ \leq^{(d)} \left( \frac{4\epsilon}{\epsilon - M - 2 - H(X | Y)} \right)^\rho \sum_{y \in Y} P_Y(y) \left( \sum_{k=1}^{J} P_{X|Y}(\varsigma_y(k) | y)^{1/(1+p)} \right) \left( \frac{1}{Q_X^{\rho/(1+p)}(\varsigma_y(J + 1) | y)} \right)^\rho + (1 - \epsilon) \]
\[ \leq^{(e)} \left( \frac{4\epsilon}{\epsilon - M - 2 - H(X | Y)} \right)^\rho \exp \left( \rho R_{1/(1+p)}^\epsilon(X | Y) \right) + (1 - \epsilon) \]
\[ \leq \exp \left( \rho R_{1/(1+p)}^\epsilon(X | Y) \right) - \rho \log \left( \frac{\epsilon - M - 2 - H(X | Y)}{4\epsilon} \right) + 1, \quad (384) \]
where
- (a) follows from \((383)\),
- (b) follows from \((381)\),
- (c) follows from the fact that
\[ |u|^\rho < 1 + 2^\rho u^\rho \quad (385) \]
for every \(u \geq 0\); cf. \([6, \text{Equation } (26)]\),
- (d) follows from \((376)\), and
- (e) follows by the definition of \(R_{1/(1+p)}^\epsilon(X | Y)\) stated in \((182)\). Finally, noting that
\[ R_{1/(1+p)}^\epsilon(X | Y) \leq \log \left( \frac{\epsilon - M - 2 - H(X | Y)}{4\epsilon} \right) \iff \exp \left( \rho R_{1/(1+p)}^\epsilon(X | Y) \right) - \rho \log \left( \frac{\epsilon - M - 2 - H(X | Y)}{4\epsilon} \right) \leq 1, \quad (386) \]
we obtain \((184)\) from \((384)\). This completes the proof of Lemma 13.
APPENDIX S
PROOF OF LEMMA 14
Let \( \delta : \mathcal{Y} \to [0,1] \) be given by (375). After some algebra, we observe that
\[
\tilde{H}_\varepsilon^\alpha(X \mid Y) = \bar{H}_\delta^\alpha(X \mid Y),
\]
where the right-hand side is defined in (53). Moreover, it follows by the definitions of \( J, \xi, \) and \( \nu \) stated in (148), (149), and (181) that
\[
\sum_{y \in \mathcal{Y}} P_Y(y) \delta(y) = \varepsilon,
\]
implying that \( \delta(\cdot) \in \mathcal{E}_0(\varepsilon) \). Therefore, it follows from Lemma 3 that the left-hand inequality of (185) holds.

To prove the right-hand inequality of (185), we shall revisit the guessing problem discussed in Appendix N-A. Consider the giving-up guessing function \( G^*_\text{avg} : \mathcal{X} \times \mathcal{Y} \to \mathbb{N} \cup \{0\} \) induced by the optimal guessing strategy \((g^*, \pi^\text{avg}_\delta)\) given in Lemma 11. Similar to (274), it follows from Hölder’s inequality that
\[
\mathbb{E}[G^*_\text{avg}(X,Y)^\rho \mid Y = y] \geq \exp \left( \rho H_{1/1+\rho}(P_{X|Y=y}) - \rho \log \left( \frac{1}{\varepsilon} \right) \right),
\]
where \( \delta : \mathcal{Y} \to [0,1] \) is given by (375). Hence, it follows from (387) that
\[
\mathbb{E}[G^*_\text{avg}(X,Y)^\rho] \geq \exp \left( \rho H_{1/1+\rho}(X \mid Y) - \rho \log \left( \sum_{k=1}^{J} \frac{1}{k} \right) \right). \tag{390}
\]
Therefore, it holds that
\[
G^*_\text{avg}(X,Y \parallel \rho, \varepsilon) \overset{(a)}{=} G^*_\text{avg}(X,Y \parallel \rho, \varepsilon) \overset{(b)}{\geq} \tilde{H}_\varepsilon^\alpha(X \mid Y) - \log \left( \frac{1}{\varepsilon} \right) \overset{(c)}{\geq} \tilde{H}_\varepsilon^\alpha(X \mid Y) - \log \left( 1 + \frac{H(X \mid Y)}{\varepsilon} \right), \tag{391}
\]
where
- (a) follows from Lemma 11,
- (b) follows from (390), and
- (c) follows as in Steps (b) and (c) of (332).
Combining the right-hand inequality of (146) and (391), we obtain the right-hand inequality of (185). This completes the proof of Lemma 14.

APPENDIX T
PROOF OF THEOREM 9
A. Proof of (186)
We can prove (186) in the same way as we did in Appendix Q. Replacing (361) by
\[
\sup_{y \in \mathcal{Y}} \mathbb{P}\{X \notin L(f(X,Y),Y) \mid Y = y\} \leq \varepsilon, \tag{392}
\]
Equation (369) can be strengthened to
\[
\mathbb{E}[\varepsilon(X,Y) \mid Y = y] \leq \varepsilon. \tag{393}
\]
Thus, by using Lemma 5 in Step (b) of (370) instead on Lemma 4, we obtain
\[
\frac{1}{\rho} \log \mathbb{E}[L(f(X,Y),Y)^\rho] \geq \tilde{H}_\varepsilon^{1/(1+\rho)}(X \mid Y) - \log M, \tag{394}
\]
as desired.
B. Proof of (187)

Recall that the numbers \( J(y) \) and \( \xi(y) \) are defined in (155) and (156), respectively, for each \( y \in \mathcal{Y} \). In addition, define

\[
\nu(y) \coloneqq \frac{\xi(y)}{P_{X|Y}(S_y(J(y) + 1) \mid y)},
\]

for each \( y \in \mathcal{Y} \), where the bijection \( S_y : \mathbb{N} \to X \) is defined in (147).

For each \( y \in \mathcal{Y} \), define

\[
\tilde{S}_y \coloneqq \{ S_y(k) \mid 1 \leq k \leq J(y) + 1 \}.
\]

Since \( |\tilde{S}_y| = J(y) + 1 \), it follows from (348) that

\[
\log |\tilde{S}_y| \leq \sup_{y \in \mathcal{Y}} H(P_{X|Y}(x \mid y)) \epsilon.
\]

Letting \( \delta(y) = \epsilon \) for each \( y \in \mathcal{Y} \), i.e., the deterministic map \( \delta : \mathcal{Y} \to [0,1] \) is constant, it follows by the definition of \( Q_{X|Y}^{(1/(1+p))\epsilon} \) stated in (93) that

\[
Q_{X|Y}^{(1/(1+p))\epsilon}(x, y) = \left\{ \begin{array}{ll}
\frac{P_{X|Y}(x \mid y)^{1/(1+p)}}{\sum_{k=1}^{J(y)} P_{X|Y}(S_y^{-1}(k) \mid y)^{1/(1+p)} + \nu(y)^{1/(1+p)} P_{X|Y}(S_y^{-1}(J(y) + 1 \mid y)^{1/(1+p)})} & \text{if } 1 \leq S_y^{-1}(x) \leq J(y), \\
0 & \text{if } S_y^{-1}(x) = J(y) + 1,
\end{array} \right.
\]

for every \((x, y) \in X \times \mathcal{Y}\). In addition, for each \( y \in \mathcal{Y} \), define the function \( \tilde{\lambda}_y : X \to \mathbb{N} \cup \{ \infty \} \) by

\[
\tilde{\lambda}_y(x) \coloneqq \left\{ \begin{array}{ll}
e \frac{2 \epsilon}{(M - 2) - \sup_{y \in \mathcal{Y}} H(P_{X|Y}(x \mid y))} & \text{if } Q_{X|Y}^{(1/(1+p))\epsilon}(x \mid y) > 0, \\
\infty & \text{if } Q_{X|Y}^{(1/(1+p))\epsilon}(x \mid y) = 0,
\end{array} \right.
\]

where \( M \) is a positive integer satisfying

\[
M > 2 + \frac{\sup_{y \in \mathcal{Y}} H(P_{X|Y}(x \mid y))}{\epsilon}.
\]

Then, a similar calculation to (379) yields

\[
\sum_{x \in X \colon \tilde{\lambda}_y(x) < \infty} \frac{1}{\tilde{\lambda}_y(x)} \leq \frac{\epsilon (M - 2) - \sup_{y \in \mathcal{Y}} H(P_{X|Y}(x \mid y))}{2 \epsilon}
\]

for every \( y \in \mathcal{Y} \). Therefore, it follows from Lemma 15 and (397) that there exists an assignment function \( f : X \times \mathcal{Y} \to \{0, 1, 2, \ldots, M\} \) satisfying

\[
1 \leq S_y^{-1}(x) \leq J(y) + 1 \iff f(x, y) \neq 0
\]

and

\[
|\mathcal{L}(f(x, y), y)| \leq \tilde{\lambda}_y(x)
\]

for every \((x, y) \in X \times \mathcal{Y}\), provided that (183) holds, where \( \mathcal{L} : \{0, 1, 2, \ldots, M\} \times \mathcal{Y} \to 2^X \) is the sub-partition induced by \( f \); see (171).

On the other hand, consider a stochastic map \( \mathcal{E} : 2^X \times \mathcal{Y} \to 2^X \) satisfying (172) and

\[
\mathbb{P}(\mathcal{E}(\mathcal{L}(f(S_y(k), y), y)) = \emptyset) = \left\{ \begin{array}{ll}
0 & \text{if } 0 \leq k \leq J(y), \\
1 - \nu(y) & \text{if } k = J(y) + 1,
1 & \text{if } J(y) + 2 \leq k < \infty
\end{array} \right.
\]

for each \((k, y) \in \mathbb{N} \times \mathcal{Y}\). We see from (404) that

\[
\mathbb{P}(X \notin \mathcal{L}(f(X, Y), Y) \mid (X, Y) = (x, y)) = \left\{ \begin{array}{ll}
0 & \text{if } 0 \leq S_y^{-1}(x) \leq J(y), \\
1 - \nu(y) & \text{if } S_y^{-1}(x) = J(y) + 1,
1 & \text{if } J(y) + 2 \leq S_y^{-1}(x) < \infty
\end{array} \right.
\]
for every \((x, y) \in X \times Y\), where the stochastic sub-partition \(L : \{0, 1, 2, \ldots, M\} \times Y \to 2^X\) induced by the pair \((f, E)\) is defined in (173). Therefore, we have

\[
\mathbb{E}[(f(X, Y), Y)^μ] \leq \sum_{y \in Y} \mathbb{P}_Y(y) \bigg( \sum_{k=1}^{J(y)} \mathbb{P}_{X\mid Y}(\mathcal{S}_g(k) \mid y) \mathcal{L}(\mathbb{P}_{f}(\mathcal{S}_g(k), y), y) \bigg)^μ \\
+ ν(y) \mathbb{P}_{X\mid Y}(\mathcal{S}_g(J(y) + 1) \mid y) \mathcal{L}(\mathbb{P}_{f}(\mathcal{S}_g(J(y) + 1), y), y) \bigg)^μ \\
\leq \sum_{y \in Y} \mathbb{P}_Y(y) \bigg( \frac{2 ε}{(M - 2) - \sup_{y \in Y} H(P_{X\mid Y = y})} \bigg)^μ \sum_{k=1}^{J(y)} \mathbb{P}_{X\mid Y}(\mathcal{S}_g(k) \mid y) \mathcal{L}(\mathbb{P}_{f}(\mathcal{S}_g(k), y), y) \bigg)^μ \\
+ ν(y) \mathbb{P}_{X\mid Y}(\mathcal{S}_g(J(y) + 1) \mid y) \mathcal{L}(\mathbb{P}_{f}(\mathcal{S}_g(J(y) + 1), y), y) \bigg)^μ \bigg) + (1 - ε) \\
\leq \sum_{y \in Y} \mathbb{P}_Y(y) \bigg( \frac{4 ε}{(M - 2) - \sup_{y \in Y} H(P_{X\mid Y = y})} \bigg)^μ \sum_{k=1}^{J(y)} \mathbb{P}_{X\mid Y}(\mathcal{S}_g(k) \mid y) \mathcal{L}(\mathbb{P}_{f}(\mathcal{S}_g(k), y), y) \bigg)^μ \bigg) \bigg) \bigg) + (1 - ε) \\
\leq \exp \left( ρ H_{1/(1+μ)}^ε(X \mid Y) - ρ \log \left( \frac{(M - 2) - \sup_{y \in Y} H(P_{X\mid Y = y})}{4 ε} \right) \right) + 1, \tag{406}
\]

where

- (a) follows from (405),
- (b) follows from (403),
- (c) follows from (385),
- (d) follows from (398), and
- (e) follows from (64).

Finally, noting that

\[
H_{1/(1+μ)}^ε(X \mid Y) \leq \log \left( \frac{(M - 2) - \sup_{y \in Y} H(P_{X\mid Y = y})}{4 ε} \right) \Longleftrightarrow \exp \left( ρ H_{1/(1+μ)}^ε(X \mid Y) - ρ \log \left( \frac{(M - 2) - \sup_{y \in Y} H(P_{X\mid Y = y})}{4 ε} \right) \right) \leq 1, \tag{407}
\]

we obtain (187) from (406).

\section*{REFERENCES}

[1] A. Rényi, “On measures of entropy and information,” in \textit{Proc. 4th Berkeley Symp. Math. Statist. Probab.}, 1961, pp. 574–561.
[2] L. L. Campbell, “A coding theorem and Rényi’s entropy,” \textit{Inf. Control}, vol. 8, no. 4, pp. 423–429, 1965.
[3] T. A. Courtade and S. Verdú, “Cumulant generating function of codeword lengths in optimal lossless compression,” in \textit{Proc. IEEE Int. Symp. Inf. Theory}, Honolulu, HI, USA, June–July 2014, pp. 2494–2498.
[4] J. L. Massey, “Guessing and entropy,” in \textit{Proc. IEEE Int. Symp. Inf. Theory}, Chicago, IL, USA, Jun. 1994, p. 204.
[5] E. Arikan, “An inequality on guessing and its application to sequential decoding,” \textit{IEEE Trans. Inf. Theory}, vol. 42, no. 1, pp. 99–105, Jan. 1996.
[6] C. Bunte and A. Lapidoth, “Encoding tasks and Rényi entropy,” \textit{IEEE Trans. Inf. Theory}, vol. 60, no. 9, pp. 5065–5076, Sept. 2014.
[7] R. Renner and S. Wolf, “Smooth Rényi entropy and applications,” in \textit{Proc. IEEE Int. Symp. Inf. Theory}, Chicago, USA, June–July 2004, p. 232.
[8] ——, “Simple and tight bounds for information reconciliation and privacy amplification,” in \textit{Advances in Cryptology—ASIACRYPT 2005}, Chennai, India, Dec. 2005, pp. 199–216.
[9] R. König, R. Renner, and C. Schaffner, “The operational meaning of min- and max-entropy,” \textit{IEEE Trans. Inf. Theory}, vol. 55, no. 9, pp. 4337–4347, Sept. 2009.
