

Article

Variable-Length Resolvability for General Sources and Channels[†]

Hideki Yagi ^{1,*}  and Te Sun Han ² 

¹ Department of Computer and Network Engineering, The University of Electro-Communications, Tokyo 182-8585, Japan

² The National Institute of Information and Communications Technology (NICT), Tokyo 184-8795, Japan

* Correspondence: h.yagi@uec.ac.jp

[†] This paper is an extended version of our paper published in the proceedings of 2017 IEEE International Symposium on Information Theory (ISIT2017), Aachen, Germany, titled “Variable-length resolvability for general sources”.

Abstract: We introduce the problem of variable-length (VL) source resolvability, in which a given target probability distribution is approximated by encoding a VL uniform random number, and the asymptotically minimum average length rate of the uniform random number, called the VL resolvability, is investigated. We first analyze the VL resolvability with the variational distance as an approximation measure. Next, we investigate the case under the divergence as an approximation measure. When the asymptotically exact approximation is required, it is shown that the resolvability under two kinds of approximation measures coincides. We then extend the analysis to the case of channel resolvability, where the target distribution is the output distribution via a general channel due to a fixed general source as an input. The obtained characterization of channel resolvability is fully general in the sense that, when the channel is just an identity mapping, it reduces to general formulas for source resolvability. We also analyze the second-order VL resolvability.

Keywords: random number generation; source resolvability; channel resolvability; output approximation; variable-length resolvability



Citation: Yagi, H.; Han, T.S.

Variable-Length Resolvability for General Sources and Channels.

Entropy **2023**, *25*, 1466. <https://doi.org/10.3390/e25101466>

Academic Editors: Jun Chen and Sadaf Salehkalaibar

Received: 7 September 2023

Revised: 30 September 2023

Accepted: 3 October 2023

Published: 19 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Generating a random number subject to a given probability distribution has a number of applications, such as in information security, statistical machine learning, and computer science. From the viewpoint of information theory, random number generation may be considered to be a transformation (encoding) of sequences emitted from a given source with *coin distribution* into other sequences with *target distribution* via a deterministic mapping [1–3]. Among others, there have been two major types of problems of random number generation: *intrinsic randomness* [4,5] and (source) *resolvability* [6,7]. In the former case, a fixed-length (FL) uniform random number is extracted from an arbitrary coin distribution, and we want to find the *maximum* achievable rate of such uniform random numbers. In the latter case, in contrast, an FL uniform random number used as a coin distribution is encoded to approximate a given *target distribution*, and we want to find the *minimum* achievable rate of such uniform random numbers. Thus, there is a duality between these two problems.

The problem of intrinsic randomness has been extended to the case of *variable-length* (VL) uniform random numbers, for which the length of random numbers may vary. This problem, referred to as the *VL intrinsic randomness*, was first introduced by Vembu and Verdú [5] for a finite source alphabet and later extended by Han [4] to a countably infinite alphabet. This problem was actually motivated because, in many practical situations, it is indispensable to consider cases where FL uniform random numbers are not available, and, instead, *VL uniform random numbers* are available; typically, this is in cases where we work with Elias’

universal random numbers [1]. The use of such uniform random numbers is expected generally to increase the achievable average length rate for intrinsic randomness. Then, the following natural question may be raised: Can we indeed lower the average length rate needed in the “resolvability” problem by using VL random numbers? The answer is “yes”. Despite the duality between these two kinds of problems for random number generation, the VL counterpart in the resolvability problem has not been discussed, where we focus on this problem.

We introduce the problem of *VL source/channel resolvability*, where a given target probability distribution is to be approximated by encoding a *VL uniform random number*. Distance measures between the target distribution and an approximated distribution are used to measure the fineness of the approximation. We first analyze the fundamental limit on VL source resolvabilities with the variational distance as an approximation measure in Section 3. We use the smooth Shannon entropy, which is a version of smooth Rényi entropy [8], to characterize the δ -source resolvability, which is defined as the minimum achievable length rate of uniform random numbers with an asymptotic distance of less than or equal to $\delta \in [0, 1)$. In the proof of the direct part, we will develop a simple version of *information spectrum slicing* [2], in which each “sliced” information density quantized to an integer is approximated by an FL uniform random number. Due to the simplicity of the method, the analysis with variational distance is first facilitated. As an important implication of general formulas for the δ -source resolvability, it is shown that the minimum resolvability rate of VL resolvability is equal to $(1 - \delta)$ times that of FL resolvability when the source is stationary and memoryless or is even with one-point spectrum (cf. Corollary 1). This result indicates an advantage of the use of a VL uniform random number when $\delta > 0$ because we can make the VL resolvability rate strictly smaller than an FL one. We then extend these analyses to the case under the (unnormalized) divergence as an approximation measure in Section 4. When $\delta = 0$, that is, when the asymptotically exact approximation is required, it is shown that the 0-source resolvabilities under two kinds of approximation measures coincide with each other.

In Section 5, we then consider the problem of *channel resolvability* [6,9,10], in which not only a source but also a channel is fixed, and the output distribution via the channel is now the target of approximation. This problem, also referred to as the problem of output approximation, provides a powerful tool to analyze the fundamental limits of various problems in information theory and information security. Some such examples include identification codes [11–13], distributed hypothesis testing [14], message authentication [15], secret key generation [16], and coding for secure communication [17–19]. We consider two types of problems in which either a general source (*mean-channel resolvability*) or a VL uniform random number (*VL channel resolvability*) is used as a coin distribution. It is shown that the formulas established are equal for both coin distributions. In the special case that the channel is the identity mapping, the formulas established reduce to those in source resolvability as established in Sections 3 and 4.

From Sections 3–5, the so-called *first-order* resolvability rates are analyzed, and the next important step may be the second-order analysis. Second-order analyses for various coding problems were initiated by Strassen [20] and have been studied in the past decade or so (cf. [21–31]). We also analyze the second-order fundamental limits of the VL channel/source resolvability in Section 6. In this paper, it is shown that the VL δ -source resolvability under the variational distance is equal to the minimum achievable rate of *fixed-to-variable-length source codes* with an error probability of less than or equal to δ . It is demonstrated that this close relationship provides a single-letter characterization for the first- and second-order source resolvability under the variational distance when the source is stationary and memoryless. It is worth noting that second-order analyses for the VL setting are relatively few, compared to those in the FL setting. The second-order formulas established in this paper are of importance from this perspective, too.

The remainder of the paper is organized as follows: Section 2 reviews the problem of FL source resolvability and the relations between the minimum resolvability rate and

the minimum coding rate of FL source codes. Section 3 formally introduces the problem of VL source resolvability with variational distance as an approximation measure. Then, Section 4 discusses VL source resolvability with divergence as an approximation measure, and Section 5 generalizes the settings to channel resolvability. Section 6 investigates the second-order fundamental limits of the VL channel/source resolvability. Section 7 concludes the paper with a discussion of possible extensions.

2. FL Resolvability: Review

Let $\mathcal{U} = \{1, 2, \dots, K\}$ be a finite alphabet of size K , and let \mathcal{X} be a finite or countably infinite alphabet. Let $\mathbf{X} = \{X^n\}_{n=1}^\infty$ be a *general source* [2], where P_{X^n} is a probability distribution on \mathcal{X}^n . We do not impose any assumptions, such as stationarity or ergodicity. In this paper, we identify X^n with its probability distribution P_{X^n} , and these symbols are used interchangeably.

We first review the problem of FL (source) resolvability [2] using the variational distance as an approximation measure. Let U_{M_n} denote the *uniform random number*, which is a random variable *uniformly* distributed over $\mathcal{U}_{M_n} := \{1, \dots, M_n\}$. Consider the problem of approximating the *target distribution* P_{X^n} by using U_{M_n} as a *coin distribution* via a deterministic mapping $\varphi_n : \{1, \dots, M_n\} \rightarrow \mathcal{X}^n$. Denoting $\tilde{X}^n = \varphi_n(U_{M_n})$, we want to make $P_{\tilde{X}^n}$ approximate P_{X^n} (cf. Figure 1). A standard choice of the performance measure for approximation is

$$d(P_{X^n}, P_{\tilde{X}^n}) := \frac{1}{2} \sum_{x \in \mathcal{X}^n} |P_{X^n}(x) - P_{\tilde{X}^n}(x)|, \tag{1}$$

which is referred to as the *variational distance* between P_{X^n} and $P_{\tilde{X}^n}$. It is easily seen that

$$0 \leq d(P_{X^n}, P_{\tilde{X}^n}) \leq 1. \tag{2}$$

Let us now review the problem for *source resolvability*. Throughout this paper, logarithms are of the base K .

Definition 1 (FL resolvability). *A resolution rate $R \geq 0$ is said to be FL achievable or simply f-achievable (under the variational distance) if there exists a deterministic mapping $\varphi_n : \{1, \dots, M_n\} \rightarrow \mathcal{X}^n$ satisfying*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log M_n \leq R, \tag{3}$$

$$\lim_{n \rightarrow \infty} d(P_{X^n}, P_{\tilde{X}^n}) = 0, \tag{4}$$

where $\tilde{X}^n = \varphi_n(U_{M_n})$ and U_{M_n} is the uniform random number over \mathcal{U}_{M_n} . The infimum of f-achievable rates, i.e.,

$$S_f(\mathbf{X}) := \inf\{R : R \text{ is f-achievable}\} \tag{5}$$

is called the FL resolvability or simply f-resolvability.

Then, we have the following theorem:

Theorem 1 (Han and Verdú [6]). *For any general target source \mathbf{X} ,*

$$S_f(\mathbf{X}) = \bar{H}(\mathbf{X}), \tag{6}$$

where

$$\bar{H}(\mathbf{X}) := \inf\left\{a : \lim_{n \rightarrow \infty} \Pr\left\{\frac{1}{n} \log \frac{1}{P_{X^n}(\mathbf{X}^n)} > a\right\} = 0\right\}. \tag{7}$$

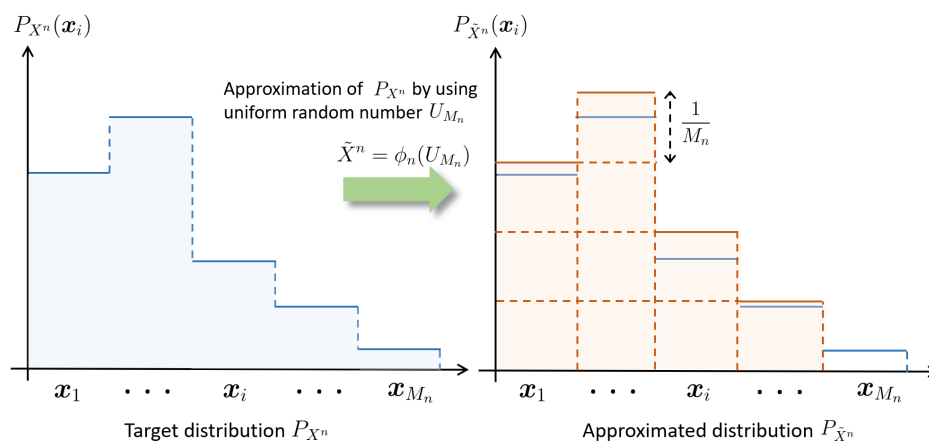


Figure 1. Illustration of the problem of FL resolvability.

Remark 1. As a dual counterpart of (7), we may define

$$\underline{H}(\mathbf{X}) := \sup \left\{ b : \lim_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} < b \right\} = 0 \right\}. \tag{8}$$

Sources such that $\overline{H}(\mathbf{X}) = \underline{H}(\mathbf{X})$ are called one-point spectrum sources (or equivalently, said to satisfy the strong converse property (cf. Han [2])), which includes stationary memoryless sources and stationary ergodic sources, etc. This class of sources is discussed later in Corollary 1.

The following problem is called the δ -resolvability problem [2,7], which relaxes the condition on the variational distance, compared to (4).

Definition 2 (FL δ -resolvability). For a fixed $\delta \in [0, 1]$, a resolution rate $R \geq 0$ is said to be FL δ -achievable or simply $f(\delta)$ -achievable (under the variational distance) if there exists a deterministic mapping $\varphi_n : \{1, \dots, M_n\} \rightarrow \mathcal{X}^n$ satisfying

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log M_n \leq R, \tag{9}$$

$$\limsup_{n \rightarrow \infty} d(P_{X^n}, P_{\tilde{X}^n}) \leq \delta, \tag{10}$$

where $\tilde{X}^n = \varphi_n(U_{M_n})$ and U_{M_n} is the uniform random number over \mathcal{U}_{M_n} . The infimum of all $f(\delta)$ -achievable rates, i.e.,

$$S_f(\delta|\mathbf{X}) := \inf \{ R : R \text{ is } f(\delta)\text{-achievable} \} \tag{11}$$

is referred to as the FL δ -resolvability or simply $f(\delta)$ -resolvability.

Then, a characterization of $S_f(\delta|\mathbf{X})$ is given by

Theorem 2 (Steinberg and Verdú [7]). For any general target source \mathbf{X} ,

$$S_f(\delta|\mathbf{X}) = \overline{H}_\delta(\mathbf{X}) \quad (\delta \in [0, 1]), \tag{12}$$

where

$$\overline{H}_\delta(\mathbf{X}) := \inf \left\{ a : \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} > a \right\} \leq \delta \right\}. \tag{13}$$

Remark 2. The FL resolvability problem is deeply related to the FL source coding problem allowing a probability of a decoding error up to ϵ . Denoting by $R_f(\epsilon|\mathbf{X})$ the minimum achievable rate for the source \mathbf{X} , there is the relationship [7]:

$$R_f(\varepsilon|\mathbf{X}) = \overline{H}_\varepsilon(\mathbf{X}) \quad (\forall \varepsilon \in [0, 1]) \tag{14}$$

and, hence, by Theorem 2,

$$S_f(\delta|\mathbf{X}) = R_f(\delta|\mathbf{X}) \quad (\forall \delta \in [0, 1]). \tag{15}$$

Formula (14) can also be shown with a smooth Rényi entropy of order zero [32].

3. VL Resolvability: Variational Distance

In this section, we introduce the problem of *variable-length* (VL) resolvability, where the target probability distribution is approximated by encoding a *VL uniform random number*. As an initial step, we analyze the fundamental limit on the VL resolvability with the variational distance as an approximation measure.

3.1. Definitions

Let \mathcal{U}^* denote the set of all sequences $\mathbf{u} \in \mathcal{U}^m$ over $m = 0, 1, 2, \dots$, where $\mathcal{U}^0 = \{\lambda\}$ (λ is the null string). Let L_n denote a random variable which takes a value in $\{0, 1, 2, \dots\}$. We define the *VL uniform random number* $U^{(L_n)}$ so that $U^{(m)}$ is uniformly distributed over \mathcal{U}^m given $L_n = m$. In other words,

$$P_{U^{(L_n)}}(\mathbf{u}, m) := \Pr\{U^{(L_n)} = \mathbf{u}, L_n = m\} = \frac{\Pr\{L_n = m\}}{K^m} \quad (\forall \mathbf{u} \in \mathcal{U}^m), \tag{16}$$

$$\Pr\{U^{(L_n)} = \mathbf{u} | L_n = m\} = \frac{P_{U^{(L_n)}}(\mathbf{u}, m)}{\Pr\{L_n = m\}} = \frac{1}{K^m} \quad (\forall \mathbf{u} \in \mathcal{U}^m), \tag{17}$$

where $K = |\mathcal{U}|$. It should be noticed that the VL sequence $\mathbf{u} \in \mathcal{U}^m$ is generated with the joint probability $P_{U^{(L_n)}}(\mathbf{u}, m)$.

We formally define the δ -resolvability problem under the variational distance using the VL random number, called the *VL δ -resolvability* or simply *v(δ)-resolvability*.

Definition 3 (VL δ -resolvability: variational distance). *A resolution rate $R \geq 0$ is said to be VL δ -achievable (under the variational distance) with $\delta \in [0, 1)$ if there exists a VL uniform random number $U^{(L_n)}$ and a deterministic mapping $\varphi_n : \mathcal{U}^* \rightarrow \mathcal{X}^n$ satisfying*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[L_n] \leq R, \tag{18}$$

$$\limsup_{n \rightarrow \infty} d(P_{X^n}, P_{\tilde{X}^n}) \leq \delta, \tag{19}$$

where $\mathbb{E}[\cdot]$ denotes the expected value and $\tilde{X}^n = \varphi_n(U^{(L_n)})$. The infimum of all v(δ)-achievable rates, i.e.,

$$S_v(\delta|\mathbf{X}) := \inf\{R : R \text{ is v}(\delta)\text{-achievable}\} \tag{20}$$

is referred to as the *VL δ -resolvability* or simply *v(δ)-resolvability*.

If $\delta = 0$, v(0)-achievable is said to be *VL achievable* or simply *v-achievable* (under the variational distance). The infimum of all v-achievable rates, i.e.,

$$S_v(\mathbf{X}) := \inf\{R : R \text{ is v-achievable}\} \tag{21}$$

is called the *VL resolvability* or simply *v-resolvability*.

Remark 3. One may think that condition (18) can be replaced with the condition on the sup-entropy rate:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} H(U^{(L_n)}) \leq R \tag{22}$$

as in [6], where $H(\cdot)$ denotes the Shannon entropy. Indeed, both conditions yield the same resolvability result. To see this, let us denote by $\tilde{S}_v(\delta|\mathbf{X})$ the infimum of v -achievable rates R under constraints (19) and (22). It is easily checked that

$$\begin{aligned} \mathbb{E}[L_n] &= \sum_{m=1}^{\infty} \sum_{\mathbf{u} \in \mathcal{U}^m} P_{U^{(L_n)}}(\mathbf{u}, m) \log K^m \\ &= \sum_{m=1}^{\infty} \sum_{\mathbf{u} \in \mathcal{U}^m} P_{U^{(L_n)}}(\mathbf{u}, m) \log \frac{\Pr\{L_n = m\}}{P_{U^{(L_n)}}(\mathbf{u}, m)} \\ &= H(U^{(L_n)}) - H(L_n) \leq H(U^{(L_n)}). \end{aligned} \tag{23}$$

This implies $S_v(\delta|\mathbf{X}) \leq \tilde{S}_v(\delta|\mathbf{X})$. On the other hand, by invoking the well-known relation (cf. ([33], Corollary 3.12)), it holds that

$$H(L_n) \leq \log(e \cdot \mathbb{E}[L_n]). \tag{24}$$

Consider any resolution rate $R > S_v(\delta|\mathbf{X})$. Then, (18) holds for some $U^{(L_n)}$ and φ_n and, hence, (24) leads to

$$\lim_{n \rightarrow \infty} \frac{1}{n} H(L_n) = \lim_{n \rightarrow \infty} \frac{1}{n} \log(e \cdot \mathbb{E}[L_n]) = 0. \tag{25}$$

From this equation, (23) yields that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} H(U^{(L_n)}) = \limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[L_n] \leq R \tag{26}$$

to obtain $R \geq \tilde{S}_v(\delta|\mathbf{X})$, implying that $S_v(\delta|\mathbf{X}) \geq \tilde{S}_v(\delta|\mathbf{X})$. Thus, $S_v(\delta|\mathbf{X}) = \tilde{S}_v(\delta|\mathbf{X})$.

3.2. Smooth Shannon Entropy

To establish a general formula for $S_v(\delta|\mathbf{X})$, we introduce the following quantity for a general source \mathbf{X} . Let $\mathcal{P}(\mathcal{X}^n)$ denote the set of all probability distributions on \mathcal{X}^n . For $\delta \in [0, 1)$, by defining the δ -ball using the variational distance

$$B_\delta(X^n) = \{P_{V^n} \in \mathcal{P}(\mathcal{X}^n) : d(P_{X^n}, P_{V^n}) \leq \delta\}, \tag{27}$$

we introduce the smooth Shannon entropy:

$$\begin{aligned} H_{[\delta]}(X^n) &:= \inf_{P_{V^n} \in B_\delta(X^n)} \sum_{\mathbf{x} \in \mathcal{X}^n} P_{V^n}(\mathbf{x}) \log \frac{1}{P_{V^n}(\mathbf{x})} \\ &= \inf_{P_{V^n} \in B_\delta(X^n)} H(V^n), \end{aligned} \tag{28}$$

where $H(V^n)$ denotes the Shannon entropy of P_{V^n} . The $H_{[\delta]}(X^n)$ is a nonincreasing function of δ . Based on this quantity for a general source $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$, we define

$$H_{[\delta]}(\mathbf{X}) = \limsup_{n \rightarrow \infty} \frac{1}{n} H_{[\delta]}(X^n). \tag{29}$$

Remark 4. Renner and Wolf [8] have defined the smooth Rényi entropy of order $\alpha \in (0, 1) \cup (1, \infty)$ as

$$H_{[\delta]}^\alpha(X^n) = \inf_{P_{V^n} \in B_\delta(X^n)} \frac{1}{1 - \alpha} \log \sum_{\mathbf{x} \in \mathcal{X}^n} P_{V^n}(\mathbf{x})^\alpha. \tag{30}$$

By letting $\alpha \uparrow 1$, we have

$$\lim_{\alpha \uparrow 1} H_{[\delta]}^\alpha(X^n) = H_{[\delta]}(X^n). \tag{31}$$

As for the proof, see Appendix A.

3.3. General Formula for General $\delta \in [0, 1)$

The following main theorem indicates that the $v(\delta)$ -resolvability $S_v(\delta|X)$ can be characterized by the smooth Shannon entropy for X .

Theorem 3. For any general target source X ,

$$S_v(\delta|X) = \lim_{\gamma \downarrow 0} H_{[\delta+\gamma]}(X) \quad (\delta \in [0, 1)). \tag{32}$$

Remark 5. In Formula (32), the limit $\lim_{\gamma \downarrow 0}$ of the offset term $+\gamma$ appears in the characterization of $S_v(\delta|X)$. This is because the smooth entropy $H_{[\delta]}(X^n)$ for X^n involves the infimum over the non-asymptotic δ -ball $B_\delta(X^n)$ for a given length n . Alternatively, we may consider the asymptotic δ -ball defined as

$$B_\delta(X) = \left\{ V = \{V^n\}_{n=1}^\infty : \limsup_{n \rightarrow \infty} d(P_{X^n}, P_{V^n}) \leq \delta \right\}, \tag{33}$$

and then we obtain the alternative formula

$$S_v(\delta|X) = \inf_{V \in B_\delta(X)} H(V) \quad (\delta \in [0, 1)) \tag{34}$$

without an offset term, where

$$H(V) := \limsup_{n \rightarrow \infty} \frac{1}{n} H(V^n) \tag{35}$$

is the sup-entropy rate for V with the Shannon entropy $H(V^n)$. The proof of (34) is given in Appendix B.

The same remark also applies to general formulas to be established in the subsequent sections.

Remark 6. Independently of this work, Tomita, Uyematsu, and Matsumoto [34] have investigated the following problem: the coin distribution is given by fair coin-tossing and the average number of coin tosses should be asymptotically minimized as in [3] while the variational distance between the target and approximated distributions should satisfy (19). In this case, the asymptotically minimum average number of coin tosses is also characterized by the right-hand side (r.h.s.) of (32) (cf. [34]). Since the coin distribution is restricted to that given by fair coin-tossing with a stopping algorithm, realizations of L_n must satisfy the Kraft inequality (for prefix codes), whereas the problem addressed in this paper allows the probability distribution of L_n to be an arbitrary discrete one, not necessarily implying prefix codes. In this sense, our problem is more relaxed, while the coin is constrained to be conditionally independent given L_n . Theorem 3 indicates that the $v(\delta)$ -resolvability does not differ in both problems. Later, we shall show that, even in the case where the coin distribution may be any general source X , the δ -resolvability remains the same (cf. Theorem 5 and Remark 14).

On the other hand, we now define the following information quantity (to be used in Remark 7 below) to discuss the relationship with VL source coding: For $\delta \in [0, 1)$ we define

$$G_{[\delta]}(X^n) = \inf_{\substack{\mathcal{A}_n \subseteq \mathcal{X}^n: \\ \Pr\{X^n \in \mathcal{A}_n\} \geq 1-\delta}} \sum_{x \in \mathcal{A}_n} P_{X^n}(x) \log \frac{1}{P_{X^n}(x)}. \tag{36}$$

The $G_{[\delta]}(X^n)$ is a nonincreasing function of δ . Based on this quantity, we define

$$G_{[\delta]}(X) = \limsup_{n \rightarrow \infty} \frac{1}{n} G_{[\delta]}(X^n). \tag{37}$$

Then, we have

Remark 7. There is a deep relation between the δ -resolvability problem and VL δ -source coding with the error probability asymptotically not exceeding δ . Koga and Yamamoto [35] (also, cf. Han [36]) showed that the minimum average length rate $R_v^*(\delta|\mathbf{X})$ of VL δ -source codes is given by

$$R_v^*(\delta|\mathbf{X}) = \lim_{\gamma \downarrow 0} G_{[\delta+\gamma]}(\mathbf{X}) \quad (\forall \delta \in [0, 1]). \tag{38}$$

Theorem 3 and Proposition 1 (to be shown just below) reveal that

$$S_v(\delta|\mathbf{X}) = R_v^*(\delta|\mathbf{X}) \quad (\forall \delta \in [0, 1]). \tag{39}$$

The following proposition shows a general relationship between $G_{[\delta]}(\mathbf{X})$ and $H_{[\delta]}(\mathbf{X})$.

Proposition 1. For any general source \mathbf{X} ,

$$H_{[\delta]}(\mathbf{X}) = G_{[\delta]}(\mathbf{X}) \leq (1 - \delta)\overline{H}_{\delta-\gamma}(\mathbf{X}) \quad (\forall \delta \in (0, 1), \forall \gamma \in (0, \delta]). \tag{40}$$

In particular,

$$\lim_{\gamma \downarrow 0} H_{[\delta+\gamma]}(\mathbf{X}) = \lim_{\gamma \downarrow 0} G_{[\delta+\gamma]}(\mathbf{X}) \leq (1 - \delta)\overline{H}_{\delta}(\mathbf{X}) \quad (\forall \delta \in [0, 1]). \tag{41}$$

(Proof) See Appendix C.

By plugging $\gamma = \delta$ into (40), a looser but sometimes useful bound

$$H_{[\delta]}(\mathbf{X}) = G_{[\delta]}(\mathbf{X}) \leq (1 - \delta)\overline{H}(\mathbf{X}) \tag{42}$$

can be obtained. Equation (40) has been derived by [21], which improves a bound established in [35,37]. In view of Theorems 2 and 3, (41) in Proposition 1 implies

$$S_v(\delta|\mathbf{X}) \leq (1 - \delta)S_f(\delta|\mathbf{X}) \tag{43}$$

for all $\delta \in [0, 1)$, where $S_f(\delta|\mathbf{X})$ denotes the $f(\delta)$ -resolvability. This general relationship elucidates the advantage of the use of VL uniform random numbers to make the average length rate lower. The proposition also claims that $G_{[\delta]}(\mathbf{X})$ coincides with $H_{[\delta]}(\mathbf{X})$ for all $\delta \in [0, 1)$ for any general source \mathbf{X} .

A consequence of Theorem 3 is the following corollary:

Corollary 1. Let $\mathbf{X} = \{X^n\}_{n=1}^\infty$ be a one-point spectrum source ($\overline{H}(\mathbf{X}) = \underline{H}(\mathbf{X})$) with $X^n = (X_1, X_2, \dots, X_n)$, then we have

$$S_f(\delta|\mathbf{X}) = \overline{H}_{\delta}(\mathbf{X}) = H^*(\mathbf{X}) \quad (\forall \delta \in [0, 1]), \tag{44}$$

where $H^*(\mathbf{X}) := \overline{H}(\mathbf{X}) = \underline{H}(\mathbf{X})$. Moreover, it holds that

$$S_v(\delta|\mathbf{X}) = H_{[\delta]}(\mathbf{X}) = G_{[\delta]}(\mathbf{X}) = (1 - \delta)H^*(\mathbf{X}) \quad (\forall \delta \in [0, 1]), \tag{45}$$

where we notice that $H^*(\mathbf{X}) = H(X_1)$ (entropy) for stationary memoryless sources.

(Proof) See Appendix D.

Now, we are ready to give the proof of Theorem 3:

Proof of Theorem 3.

(1) Converse Part:

Let R be $v(\delta)$ -achievable. Then, there exists $U^{(L_n)}$ and φ_n satisfying (18) and

$$\limsup_{n \rightarrow \infty} \delta_n \leq \delta, \tag{46}$$

where we define $\delta_n = d(P_{X^n}, P_{\tilde{X}^n})$ with $\tilde{X}^n = \varphi_n(U^{(L_n)})$. Equation (46) implies that, for any given $\gamma > 0$, it holds that $\delta_n \leq \delta + \gamma$ for all $n \geq n_0$ with some $n_0 > 0$, and thus we have

$$H_{[\delta+\gamma]}(X^n) \leq H_{[\delta_n]}(X^n) \quad (\forall n \geq n_0), \tag{47}$$

because $H_{[\delta]}(X^n)$ is a nonincreasing function of δ . Since $P_{\tilde{X}^n} \in B_{\delta_n}(X^n)$, we have

$$H_{[\delta_n]}(X^n) \leq H(\tilde{X}^n). \tag{48}$$

On the other hand, it follows from (23) that

$$H(\tilde{X}^n) \leq H(U^{(L_n)}) = \mathbb{E}[L_n] + H(L_n), \tag{49}$$

where the inequality is due to the fact that φ_n is a deterministic mapping and $\tilde{X}^n = \varphi_n(U^{(L_n)})$.

Combining (47)–(49) yields

$$\begin{aligned} H_{[\delta+\gamma]}(\mathbf{X}) &= \limsup_{n \rightarrow \infty} \frac{1}{n} H_{[\delta+\gamma]}(X^n) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[L_n] + \limsup_{n \rightarrow \infty} \frac{1}{n} H(L_n) \leq R, \end{aligned} \tag{50}$$

where we have used (18) and (25) for the last inequality. Since $\gamma > 0$ is arbitrary, we obtain

$$\lim_{\gamma \downarrow 0} H_{[\delta+\gamma]}(\mathbf{X}) \leq R. \tag{51}$$

(2) Direct Part:

Without loss of generality, we assume that $H^+ := \lim_{\gamma \downarrow 0} H_{[\delta+\gamma]}(\mathbf{X})$ is finite ($H^+ < +\infty$). Letting $R = H^+ + 3\gamma$, where $\gamma > 0$ is an arbitrary constant, we shall show that R is $v(\delta)$ -achievable. In what follows, we use a simpler form of *information spectrum slicing* [2], where each piece of sliced information quantized to a positive integer ℓ is approximated by the uniform random number $U^{(\ell)}$ of the length ℓ .

First, we note that

$$H^+ \geq H_{[\delta+\gamma]}(\mathbf{X}) \geq \frac{1}{n} H_{[\delta+\gamma]}(X^n) - \gamma \quad (\forall n > n_0) \tag{52}$$

because of the monotonicity of $H_{[\delta]}(\mathbf{X})$ in δ . Let V^n be a random variable subject to $P_{V^n} \in B_{\delta+\gamma}(X^n)$, which satisfies

$$H_{[\delta+\gamma]}(X^n) + \gamma \geq H(V^n). \tag{53}$$

For $\gamma > 0$, we can choose a $c_n > 0$ so large that

$$\Pr\{V^n \notin T_n\} \leq \gamma \tag{54}$$

where

$$T_n := \left\{ \mathbf{x} \in \mathcal{X}^n : \frac{1}{n} \log \frac{1}{P_{V^n}(\mathbf{x})} \leq c_n \right\}. \tag{55}$$

We also define

$$\ell(\mathbf{x}) := \begin{cases} \lceil \log \frac{1}{P_{V^n}(\mathbf{x})} + n\gamma \rceil & \text{for } \mathbf{x} \in T_n \\ 0 & \text{otherwise.} \end{cases} \tag{56}$$

For $m = 0, 1, \dots, \beta_n := \lceil n(c_n + \gamma) \rceil$, set

$$S_n(m) := \{\mathbf{x} \in \mathcal{X}^n : \ell(\mathbf{x}) = m\}, \tag{57}$$

then these sets form a partition of \mathcal{X}^n , i.e.,

$$\bigcup_{m=0}^{\beta_n} S_n(m) = \mathcal{X}^n \quad \text{and} \quad \bigcup_{m=1}^{\beta_n} S_n(m) = T_n. \tag{58}$$

We set L_n so that

$$\Pr\{L_n = m\} = \Pr\{V^n \in S_n(m)\}, \tag{59}$$

where it is obvious that $\sum_{m=0}^{\beta_n} \Pr\{L_n = m\} = 1$, and, hence, the probability distribution of the VL uniform random number $U^{(L_n)}$ is given as

$$P_{U^{(L_n)}}(\mathbf{u}, m) := \Pr\{U^{(L_n)} = \mathbf{u}, L_n = m\} = \frac{\Pr\{V^n \in S_n(m)\}}{K^m} \quad (\forall \mathbf{u} \in \mathcal{U}^m). \tag{60}$$

(a) Construction of Mapping $\varphi_n : \mathcal{U}^* \rightarrow \mathcal{X}^n$:

Index the elements in $S_n(m)$ as $x_1, x_2, \dots, x_{|S_n(m)|}$ ($m = 1, 2, \dots$), where

$$|S_n(m)| \leq K^{m-n\gamma} \tag{61}$$

since for $x \in S_n(m)$

$$\log \frac{1}{P_{V^n}(x)} \leq m - n\gamma \iff P_{V^n}(x) \geq K^{-(m-n\gamma)}, \tag{62}$$

and, therefore,

$$1 \geq \sum_{x \in S_n(m)} P_{V^n}(x) \geq \sum_{x \in S_n(m)} K^{-(m-n\gamma)} = |S_n(m)|K^{-(m-n\gamma)}. \tag{63}$$

For $i = 1, 2, \dots, |S_n(m)|$, define $\tilde{A}_i^{(m)} \subset \mathcal{U}^m$ as the set of sequences $\mathbf{u} \in \mathcal{U}^m$ so that

$$\sum_{\mathbf{u} \in \tilde{A}_i^{(m)}} P_{U^{(L_n)}}(\mathbf{u}, m) \leq P_{V^n}(x_i) < \sum_{\mathbf{u} \in \tilde{A}_i^{(m)}} P_{U^{(L_n)}}(\mathbf{u}, m) + \frac{\Pr\{V^n \in S_n(m)\}}{K^m} \tag{64}$$

and

$$\tilde{A}_i^{(m)} \cap \tilde{A}_j^{(m)} = \emptyset \quad (i \neq j). \tag{65}$$

If

$$\sum_{i=1}^{|S_n(m)|} \sum_{\mathbf{u} \in \tilde{A}_i^{(m)}} P_{U^{(L_n)}}(\mathbf{u}, m) < \sum_{i=1}^{|S_n(m)|} P_{V^n}(x_i) = \Pr\{V^n \in S_n(m)\}, \tag{66}$$

then add a $\mathbf{u}_i \in \mathcal{U}^m \setminus (\cup_j \tilde{A}_j^{(m)})$ to obtain

$$A_i^{(m)} = \tilde{A}_i^{(m)} \cup \{\mathbf{u}_i\} \tag{67}$$

for $i = 1, 2, \dots$ in order, until it holds that with some $1 \leq c \leq |S_n(m)|$

$$\bigcup_{i=1}^c A_i^{(m)} \cup \bigcup_{i=c+1}^{|S_n(m)|} \tilde{A}_i^{(m)} = \mathcal{U}^m, \tag{68}$$

where $\mathbf{u}_1, \mathbf{u}_2, \dots$ are selected to be all different. Since $|\mathcal{U}^m| = K^m$ and

$$\sum_{\mathbf{u} \in \mathcal{U}^m} P_{U^{(L_n)}}(\mathbf{u}, m) = \sum_{\mathbf{u} \in \mathcal{U}^m} \frac{\Pr\{V^n \in S_n(m)\}}{K^m} = \Pr\{V^n \in S_n(m)\}, \tag{69}$$

such a $1 \leq c \leq |S_n(m)|$ always exists. For simplicity, we set for $i = c + 1, c + 2, \dots, |S_n(m)|$

$$A_i^{(m)} = \tilde{A}_i^{(m)} \tag{70}$$

and for $i = 1, 2, \dots, |S_n(m)|$

$$\varphi_n(\mathbf{u}) = \mathbf{x}_i \text{ for } \mathbf{u} \in A_i^{(m)}, \tag{71}$$

which defines the random variable \tilde{X}^n with values in \mathcal{X}^n such that

$$P_{\tilde{X}^n}(\mathbf{x}_i) = \sum_{\mathbf{u} \in A_i^{(m)}} P_{U^{(L_n)}}(\mathbf{u}, m) \quad (\mathbf{x}_i \in S_n(m)), \tag{72}$$

that is, $\tilde{X}^n = \varphi_n(U^{(L_n)})$, where if $\mathcal{X}^n \setminus T_n \neq \emptyset$, we choose some $\mathbf{x}_0 \in \mathcal{X}^n \setminus T_n$ and set

$$P_{\tilde{X}^n}(\mathbf{x}_0) = \Pr\{V^n \notin T_n\} \text{ and } \varphi_n(\lambda) = \mathbf{x}_0. \tag{73}$$

Notice that, by this construction, we have

$$|P_{\tilde{X}^n}(\mathbf{x}_i) - P_{V^n}(\mathbf{x}_i)| \leq \frac{\Pr\{V^n \in S_n(m)\}}{K^m} \tag{74}$$

for $i = 1, 2, \dots, |S_n(m)|$; $m = 1, 2, \dots, \beta_n$, and

$$\Pr\{\tilde{X}^n \notin T_n\} = \Pr\{V^n \notin T_n\} \leq \gamma. \tag{75}$$

(b) *Evaluation of Average Length:*

Since $m = 0$ does not contribute to the average length $\mathbb{E}[L_n]$, it is evaluated as follows:

$$\begin{aligned} \mathbb{E}[L_n] &= \sum_{m=1}^{\beta_n} \sum_{\mathbf{u} \in \mathcal{U}^m} P_{U^{(L_n)}}(\mathbf{u}, m) \cdot m \\ &= \sum_{m=1}^{\beta_n} \sum_{i=1}^{|S_n(m)|} \sum_{\mathbf{u} \in A_i^{(m)}} P_{U^{(L_n)}}(\mathbf{u}, m) \cdot m \\ &= \sum_{m=1}^{\beta_n} \sum_{\mathbf{x}_i \in S_n(m)} P_{\tilde{X}^n}(\mathbf{x}_i) \cdot m, \end{aligned} \tag{76}$$

where we have used $\mathcal{U}^m = \bigcup_{i=1}^{|S_n(m)|} A_i^{(m)}$ and (72). For $\mathbf{x}_i \in S_n(m)$, we obtain from (74)

$$\begin{aligned} P_{\tilde{X}^n}(\mathbf{x}_i) &\leq P_{V^n}(\mathbf{x}_i) + \frac{\Pr\{V^n \in S_n(m)\}}{K^m} \\ &\leq P_{V^n}(\mathbf{x}_i) \left(1 + \frac{1}{P_{V^n}(\mathbf{x}_i)K^m}\right) \\ &\leq P_{V^n}(\mathbf{x}_i) \left(1 + \frac{1}{K^{n\gamma}}\right), \end{aligned} \tag{77}$$

where, to derive the last inequality, we have used (62). Plugging the inequality

$$m \leq \log \frac{1}{P_{V^n}(\mathbf{x}_i)} + n\gamma + 1 \quad (\forall \mathbf{x}_i \in S_n(m)) \tag{78}$$

and (77) into (76), we obtain

$$\begin{aligned} \mathbb{E}[L_n] &\leq \left(1 + \frac{1}{K^{n\gamma}}\right) \sum_{m=1}^{\beta_n} \sum_{\mathbf{x}_i \in S_n(m)} P_{V^n}(\mathbf{x}_i) \left(\log \frac{1}{P_{V^n}(\mathbf{x}_i)} + n\gamma + 1\right) \\ &\leq \left(1 + \frac{1}{K^{n\gamma}}\right) (H(V^n) + n\gamma + 1), \end{aligned} \tag{79}$$

which yields

$$\begin{aligned}
 \limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[L_n] &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} H(V^n) + 2\gamma \\
 &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} H_{[\delta+\gamma]}(X^n) + 3\gamma \\
 &= H_{[\delta+\gamma]}(\mathbf{X}) + 3\gamma \\
 &\leq H^+ + 3\gamma = R,
 \end{aligned} \tag{80}$$

where the second inequality follows from (53) and the last one is due to (52).

(c) *Evaluation of Variational Distance:*

From (61) and (74), we have

$$\sum_{\mathbf{x} \in S_n(m)} |P_{\tilde{X}^n}(\mathbf{x}) - P_{V^n}(\mathbf{x})| \leq \frac{|S_n(m)| \Pr\{V^n \in S_n(m)\}}{K^m} \leq \frac{\Pr\{V^n \in S_n(m)\}}{K^{n\gamma}}, \tag{81}$$

which, in view of (58), leads to

$$\begin{aligned}
 d(P_{\tilde{X}^n}, P_{V^n}) &= \frac{1}{2} \sum_{\mathbf{x} \in T_n} |P_{\tilde{X}^n}(\mathbf{x}) - P_{V^n}(\mathbf{x})| + \frac{1}{2} \sum_{\mathbf{x} \notin T_n} |P_{\tilde{X}^n}(\mathbf{x}) - P_{V^n}(\mathbf{x})| \\
 &\leq \frac{1}{2} \sum_{m=1}^{\beta_n} \sum_{\mathbf{x} \in S_n(m)} |P_{\tilde{X}^n}(\mathbf{x}) - P_{V^n}(\mathbf{x})| \\
 &\quad + \frac{1}{2} (\Pr\{\tilde{X}^n \notin T_n\} + \Pr\{V^n \notin T_n\}) \\
 &\leq \frac{1}{2} \sum_{m=1}^{\beta_n} \frac{\Pr\{V^n \in S_n(m)\}}{K^{n\gamma}} + \gamma \leq \frac{1}{2} K^{-n\gamma} + \gamma,
 \end{aligned} \tag{82}$$

where we have used (75) to obtain the leftmost inequality in (82). By the triangle inequality, we obtain

$$d(P_{X^n}, P_{\tilde{X}^n}) \leq d(P_{X^n}, P_{V^n}) + d(P_{\tilde{X}^n}, P_{V^n}) \leq \delta + 2\gamma + \frac{1}{2} K^{-n\gamma}, \tag{83}$$

where the last inequality follows because $P_{V^n} \in B_{\delta+\gamma}(X^n)$. Thus, we obtain from (83)

$$\limsup_{n \rightarrow \infty} d(P_{X^n}, P_{\tilde{X}^n}) \leq \delta + 2\gamma. \tag{84}$$

Since $\gamma > 0$ is arbitrary and we have (80), we conclude that R is $v(\delta)$ -achievable.

□

3.4. General Formula for $\delta = 0$

In this subsection, we consider the special case with $\delta = 0$. In this case, we can elucidate the relationship between the minimum achievable rates for VL source codes with an asymptotically vanishing decoding error probability and the FL source codes.

We obtain the following corollary from Theorem 3 and Proposition 1:

Corollary 2. For any general target source \mathbf{X} ,

$$S_v(\mathbf{X}) = \lim_{\gamma \downarrow 0} G_{[\gamma]}(\mathbf{X}), \tag{85}$$

where $G_{[\gamma]}(\mathbf{X})$ is defined in (37).

It has been shown by Han [2] that any source $\mathbf{X} = \{X^n\}_{n=1}^\infty$ satisfying the uniform integrability (cf. Han [2]) satisfies

$$\lim_{\gamma \downarrow 0} G_{[\gamma]}(\mathbf{X}) = H(\mathbf{X}) := \limsup_{n \rightarrow \infty} \frac{1}{n} H(X^n), \tag{86}$$

where $H(\mathbf{X})$ is called the *sup-entropy rate*. Notice here, in particular, that the finiteness of an alphabet implies the uniform integrability [2]. Thus, we obtain the following corollary:

Corollary 3. For any finite alphabet target source \mathbf{X} ,

$$S_v(\mathbf{X}) = H(\mathbf{X}). \tag{87}$$

Remark 8. As in the case of FL resolvability and FL source coding problems, $S_v(\mathbf{X})$ is tightly related to VL source codes with vanishing decoding error probabilities. Denoting by $R_v^*(\mathbf{X})$ the minimum error-vanishing VL -achievable rate for a source \mathbf{X} , Han [36] has shown that

$$R_v^*(\mathbf{X}) = \lim_{\gamma \downarrow 0} G_{[\gamma]}(\mathbf{X}), \tag{88}$$

and, hence, from Corollary 3, it is concluded that

$$S_v(\mathbf{X}) = R_v^*(\mathbf{X}). \tag{89}$$

In addition, if a general source \mathbf{X} satisfies the uniform integrability and the strong converse property (cf. Han [2]), then equation (86) holds and hence it follows from ([2], Theorem 1.7.1) that

$$S_f(\mathbf{X}) = S_v(\mathbf{X}) = R_v^*(\mathbf{X}) = R_v(\mathbf{X}) = R_f(\mathbf{X}) = H(\mathbf{X}), \tag{90}$$

where $R_f(\mathbf{X}) := R_f(0|\mathbf{X})$ and $R_v(\mathbf{X})$ denotes the minimum achievable rate of VL source codes with zero error probabilities for all $n = 1, 2, \dots$.

Remark 9. Han and Verdú [6] have discussed the problem of mean-resolvability for the target distribution P_{X^n} . In this problem, the coin distribution may be a general source $\tilde{\mathbf{X}} = \{\tilde{X}^n\}_{n=1}^\infty$, where \tilde{X}^n is a random variable that takes values in \mathcal{X}^n with the average length rate $\frac{1}{n}\mathbb{E}[L_n]$ in (18) replaced with the entropy rate $\frac{1}{n}H(\tilde{X}^n)$. Denoting by $\bar{S}_v(\mathbf{X})$ the mean-resolvability, which is defined as the infimum of v -achievable rates for a general source \mathbf{X} (with the countably infinite alphabet), we can easily verify that any mean-resolution rate $R > \bar{S}_v(\mathbf{X})$ must satisfy $R \geq \lim_{\gamma \downarrow 0} G_{[\gamma]}(\mathbf{X})$ so that $S_v(\mathbf{X}) \leq \bar{S}_v(\mathbf{X})$. On the other hand, $S_v(\mathbf{X}) \geq \bar{S}_v(\mathbf{X})$ by definition. Thus, in view of Corollary 2, we have

Corollary 4. For any general target source \mathbf{X} ,

$$S_v(\mathbf{X}) = \bar{S}_v(\mathbf{X}) = \lim_{\gamma \downarrow 0} G_{[\gamma]}(\mathbf{X}). \tag{91}$$

4. VL Resolvability: Divergence

So far, we have considered the problem of VL resolvability, in which the approximation level is measured by the variational distance between X^n and \tilde{X}^n . It is sometimes of use to deal with another quantity as an approximation measure. In this section, we use the (unnormalized) divergence as the approximation measure.

4.1. Definitions

In this subsection, we address the following problem.

Definition 4 (VL δ -resolvability: divergence). A resolution rate $R \geq 0$ is said to be VL δ -achievable or simply $v_D(\delta)$ -achievable (under the divergence) with $\delta \geq 0$, if there exists a VL uniform random number $U^{(L_n)}$ and a deterministic mapping $\varphi_n : \mathcal{U}^* \rightarrow \mathcal{X}^n$ satisfying

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[L_n] \leq R, \tag{92}$$

$$\limsup_{n \rightarrow \infty} D(\tilde{X}^n || X^n) \leq \delta, \tag{93}$$

where $\tilde{X}^n = \varphi_n(U^{(L_n)})$ and $D(\tilde{X}^n || X^n)$ denotes the divergence between $P_{\tilde{X}^n}$ and P_{X^n} :

$$D(\tilde{X}^n || X^n) = \sum_{\mathbf{x} \in \mathcal{X}^n} P_{\tilde{X}^n}(\mathbf{x}) \log \frac{P_{\tilde{X}^n}(\mathbf{x})}{P_{X^n}(\mathbf{x})}. \tag{94}$$

The infimum of all $v_D(\delta)$ -achievable rates, i.e.,

$$S_v^D(\delta | \mathbf{X}) := \inf\{R : R \text{ is } v_D(\delta)\text{-achievable}\} \tag{95}$$

is called the VL δ -resolvability or simply the $v_D(\delta)$ -resolvability.

Remark 10. The measure of approximation is now the divergence $D(\tilde{X}^n || X^n)$ but not its reversed version $D(X^n || \tilde{X}^n)$. In the context of resolvability, divergence $D(\tilde{X}^n || X^n)$ (and its counterpart in the case of channel resolvability) is usually employed as in [6,7,9]. We also use this type of divergence in the subsequent sections.

To establish the general formula for $S_v^D(\delta | \mathbf{X})$, we introduce the following quantity for a general source $\mathbf{X} = \{X^n\}_{n=1}^\infty$. Recall that $\mathcal{P}(\mathcal{X}^n)$ denotes the set of all probability distributions on \mathcal{X}^n . For $\delta \geq 0$, defining the δ -ball using the divergence as

$$B_\delta^D(X^n) = \{P_{V^n} \in \mathcal{P}(\mathcal{X}^n) : D(V^n || X^n) \leq \delta\}, \tag{96}$$

we introduce the following quantity, referred to as the smooth entropy using the divergence:

$$H_{[\delta]}^D(X^n) := \inf_{P_{V^n} \in B_\delta^D(X^n)} H(V^n), \tag{97}$$

where $H(V^n)$ denotes the Shannon entropy of P_{V^n} . Obviously, $H_{[\delta]}^D(X^n)$ is a nonincreasing function of δ . Now, we define

$$H_{[\delta]}^D(\mathbf{X}) = \limsup_{n \rightarrow \infty} \frac{1}{n} H_{[\delta]}^D(X^n). \tag{98}$$

The following lemma is used to derive Corollary 5 of Theorem 4 below in the next subsection.

Lemma 1. For any general source \mathbf{X} ,

$$H_{[\delta]}^D(\mathbf{X}) \leq H_{[g(\delta)]}^D(\mathbf{X}) \quad (\delta \geq 0), \tag{99}$$

where we define $g(\delta) = 2\delta^2 / \ln K$, and

$$\lim_{\delta \downarrow 0} G_{[\delta]}(\mathbf{X}) = \lim_{\delta \downarrow 0} H_{[\delta]}^D(\mathbf{X}) = \lim_{\delta \downarrow 0} H_{[\delta]}^D(\mathbf{X}) \leq \bar{H}(\mathbf{X}). \tag{100}$$

(Proof) See Appendix E.

4.2. General Formula

Here, we establish another main theorem, which characterizes $S_v^D(\delta | \mathbf{X})$ for all $\delta \geq 0$ in terms of the smooth entropy using the divergence.

Theorem 4. For any general target source \mathbf{X} ,

$$S_v^D(\delta | \mathbf{X}) = \lim_{\gamma \downarrow 0} H_{[\delta+\gamma]}^D(\mathbf{X}) \quad (\delta \geq 0). \tag{101}$$

Remark 11. It should be noticed that the approximation measure considered here is not the normalized divergence

$$\frac{1}{n} D(\varphi_n(U^{(L_n)}) || X^n), \tag{102}$$

which has been used in the problem of FL δ -resolvability [7]. The achievability scheme given in the proof of the direct part of Theorem 4 can also be used in the case of this relaxed measure. Indeed, denoting the VL δ -resolvability with the normalized divergence by $\tilde{S}_v^D(\delta|\mathbf{X})$, the general formula for $\tilde{S}_v^D(\delta|\mathbf{X})$ is given in the same form as (101), if the radius of the δ -ball $B_\delta^D(X^n)$ in the definition of $H_{[\delta]}^D(X^n)$ is replaced with the normalized divergence. It generally holds that $S_v^D(\delta|\mathbf{X}) \geq \tilde{S}_v^D(\delta|\mathbf{X})$ for all $\delta \geq 0$ because the normalized divergence is smaller than the unnormalized divergence.

As we have seen in Lemma 1, we generally have $S_v^D(g(\delta)|\mathbf{X}) \geq S_v(\delta|\mathbf{X})$ for any $\delta \in [0, 1)$ with $g(\delta) = 2\delta^2 / \ln K$. In particular, in the case that $\delta = 0$, we obtain the following corollary of Theorems 3 and 4.

Corollary 5. For any general target source \mathbf{X} ,

$$S_v^D(0|\mathbf{X}) = S_v(\mathbf{X}). \tag{103}$$

Corollary 5 indicates that the $v_D(0)$ -resolvability $S_v^D(0|\mathbf{X})$ coincides with the v -resolvability $S_v(\mathbf{X})$ and is also characterized by the r.h.s. of (85). By (88), it also implies that $S_v^D(0|\mathbf{X}) = R_v^*(\mathbf{X})$, where $R_v^*(\mathbf{X})$ denotes the minimum error-vanishing achievable rate with VL source codes for \mathbf{X} .

Proof of Theorem 4.

(1) Converse Part:

Let R be $v_D(\delta)$ -achievable. Then, there exists $U^{(L_n)}$ and φ_n satisfying (92) and

$$\limsup_{n \rightarrow \infty} \delta_n \leq \delta, \tag{104}$$

where we define $\delta_n = D(\tilde{X}^n || X^n)$ with $\tilde{X}^n = \varphi_n(U^{(L_n)})$. Equation (104) implies that, for any given $\gamma > 0$, it holds that $\delta_n \leq \delta + \gamma$ for all $n \geq n_0$ with some $n_0 > 0$, and therefore

$$H_{[\delta+\gamma]}^D(X^n) \leq H_{[\delta_n]}^D(X^n) \quad (\forall n \geq n_0) \tag{105}$$

since $H_{[\delta]}^D(X^n)$ is a nonincreasing function of δ . Since $P_{\tilde{X}^n} \in B_{\delta_n}^D(X^n)$, we have

$$H_{[\delta_n]}^D(X^n) \leq H(\tilde{X}^n). \tag{106}$$

On the other hand, it follows from (23) that

$$H(\tilde{X}^n) \leq H(U^{(L_n)}) = \mathbb{E}[L_n] + H(L_n), \tag{107}$$

where the inequality is due to the fact that φ_n is a deterministic mapping and $\tilde{X}^n = \varphi_n(U^{(L_n)})$.

Combining (105)–(107) yields

$$\begin{aligned} H_{[\delta+\gamma]}^D(\mathbf{X}) &= \limsup_{n \rightarrow \infty} \frac{1}{n} H_{[\delta+\gamma]}^D(X^n) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[L_n] + \limsup_{n \rightarrow \infty} \frac{1}{n} H(L_n) \leq R, \end{aligned} \tag{108}$$

where we used (25) and (92) for the last inequality. Since $\gamma > 0$ is arbitrary, we have

$$\lim_{\gamma \downarrow 0} H_{[\delta+\gamma]}^D(\mathbf{X}) \leq R. \tag{109}$$

(2) Direct Part:

We modify the achievability scheme in the proof of the direct part of Theorem 3. Although the proof of this part is quite similar to that of the direct part of Theorem 3, we give here the full proof in order to avoid subtle possible confusions. We may

assume that $H^+ := \lim_{\gamma \downarrow 0} H_{[\delta+\gamma]}^D(\mathbf{X})$ is finite ($H^+ < +\infty$). Letting $R = H^+ + \mu$, where $\mu > 0$ is an arbitrary constant, we shall show that R is $v_D(\delta)$ -achievable. Let V^n be a random variable subject to $P_{V^n} \in B_{\delta+\gamma}^D(\mathbf{X}^n)$ satisfying

$$H_{[\delta+\gamma]}^D(\mathbf{X}^n) + \gamma \geq H(V^n) \tag{110}$$

for any fixed $\gamma \in (0, \frac{1}{2}]$. We can choose a $c_n > 0$ so large that

$$\gamma_0 := \Pr\{V^n \notin T_n\} \leq \gamma \tag{111}$$

where

$$T_n := \left\{ \mathbf{x} \in \mathcal{X}^n : \frac{1}{n} \log \frac{1}{P_{V^n}(\mathbf{x})} \leq c_n \right\}. \tag{112}$$

We also define

$$\ell(\mathbf{x}) := \left\lceil \log \frac{1}{P_{V^n}(\mathbf{x})} + n\gamma \right\rceil \text{ for } \mathbf{x} \in T_n. \tag{113}$$

Letting, for $m = 1, 2, \dots, \beta_n := \lceil n(c_n + \gamma) \rceil$,

$$S_n(m) := \{\mathbf{x} \in \mathcal{X}^n : \ell(\mathbf{x}) = m\}, \tag{114}$$

these sets form a partition of T_n :

$$\bigcup_{m=1}^{\beta_n} S_n(m) = T_n. \tag{115}$$

We set L_n so that

$$\Pr\{L_n = m\} = \frac{\Pr\{V^n \in S_n(m)\}}{\Pr\{V^n \in T_n\}} = \frac{\Pr\{V^n \in S_n(m)\}}{1 - \gamma_0}, \tag{116}$$

which satisfies

$$\sum_{m=1}^{\beta_n} \Pr\{L_n = m\} = \frac{\Pr\{V^n \in T_n\}}{1 - \gamma_0} = 1, \tag{117}$$

and, hence, the probability distribution of $U^{(L_n)}$ is given as

$$P_{U^{(L_n)}}(\mathbf{u}, m) := \Pr\{U^{(L_n)} = \mathbf{u}, L_n = m\} = \frac{\Pr\{V^n \in S_n(m)\}}{(1 - \gamma_0)K^m} \quad (\forall \mathbf{u} \in \mathcal{U}^m). \tag{118}$$

(a) *Construction of Mapping $\varphi_n : \mathcal{U}^* \rightarrow \mathcal{X}^n$:*

Index the elements in $S_n(m)$ as $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{|S_n(m)|}$ ($m = 1, 2, \dots, \beta_n$), where it holds that

$$|S_n(m)| \leq K^{m-n\gamma} \tag{119}$$

(cf. (61)–(63)). For $i = 1, 2, \dots, |S_n(m)|$, define $\tilde{A}_i^{(m)} \subset \mathcal{U}^m$ as the set of sequences $\mathbf{u} \in \mathcal{U}^m$ so that

$$\sum_{\mathbf{u} \in \tilde{A}_i^{(m)}} P_{U^{(L_n)}}(\mathbf{u}, m) \leq \frac{P_{V^n}(\mathbf{x}_i)}{1 - \gamma_0} < \sum_{\mathbf{u} \in \tilde{A}_i^{(m)}} P_{U^{(L_n)}}(\mathbf{u}, m) + \frac{\Pr\{L_n = m\}}{K^m} \tag{120}$$

and

$$\tilde{A}_i^{(m)} \cap \tilde{A}_j^{(m)} = \emptyset \quad (i \neq j). \tag{121}$$

If

$$\sum_{i=1}^{|S_n(m)|} \sum_{\mathbf{u} \in \tilde{A}_i^{(m)}} P_{U^{(L_n)}}(\mathbf{u}, m) < \frac{1}{1 - \gamma_0} \sum_{i=1}^{|S_n(m)|} P_{V^n}(\mathbf{x}_i) = \Pr\{L_n = m\}, \tag{122}$$

then add a $\mathbf{u}_i \in \mathcal{U}^m \setminus (\cup_j \tilde{A}_j^{(m)})$ to obtain

$$A_i^{(m)} = \tilde{A}_i^{(m)} \cup \{\mathbf{u}_i\} \tag{123}$$

for $i = 1, 2, \dots$ in order, until it holds that with some $1 \leq c \leq |S_n(m)|$

$$\bigcup_{i=1}^c A_i^{(m)} \cup \bigcup_{i=c+1}^{|S_n(m)|} \tilde{A}_i^{(m)} = \mathcal{U}^m, \tag{124}$$

where $\mathbf{u}_1, \mathbf{u}_2, \dots$ are selected to be all distinct. Since $|\mathcal{U}^m| = K^m$ and

$$\sum_{\mathbf{u} \in \mathcal{U}^m} P_{U^{(L_n)}}(\mathbf{u}, m) = \sum_{\mathbf{u} \in \mathcal{U}^m} \frac{\Pr\{V^n \in S_n(m)\}}{(1 - \gamma_0)K^m} = \Pr\{L_n = m\}, \tag{125}$$

such a $1 \leq c \leq |S_n(m)|$ always exists. For simplicity, we set for $i = c + 1, c + 2, \dots, |S_n(m)|$

$$A_i^{(m)} = \tilde{A}_i^{(m)} \tag{126}$$

and for $i = 1, 2, \dots, |S_n(m)|$

$$\varphi_n(\mathbf{u}) = \mathbf{x}_i \text{ for } \mathbf{u} \in A_i^{(m)}, \tag{127}$$

which defines the random variable \tilde{X}^n with values in \mathcal{X}^n such that

$$P_{\tilde{X}^n}(\mathbf{x}_i) = \sum_{\mathbf{u} \in A_i^{(m)}} P_{U^{(L_n)}}(\mathbf{u}, m) \quad (\mathbf{x}_i \in S_n(m)), \tag{128}$$

that is, $\tilde{X}^n = \varphi_n(U^{(L_n)})$. Notice that, by this construction, we have

$$\left| P_{\tilde{X}^n}(\mathbf{x}_i) - \frac{P_{V^n}(\mathbf{x}_i)}{1 - \gamma_0} \right| \leq \frac{\Pr\{L_n = m\}}{K^m} = \frac{\Pr\{V^n \in S_n(m)\}}{(1 - \gamma_0)K^m} \tag{129}$$

for $i = 1, 2, \dots, |S_n(m)|; m = 1, 2, \dots, \beta_n$, and

$$\Pr\{\tilde{X}^n \notin T_n\} = 0 \text{ and } \Pr\{V^n \notin T_n\} \leq \gamma. \tag{130}$$

(b) *Evaluation of Average Length:*

The average length $\mathbb{E}[L_n]$ is evaluated as follows:

$$\begin{aligned} \mathbb{E}[L_n] &= \sum_{m=1}^{\beta_n} \sum_{\mathbf{u} \in \mathcal{U}^m} P_{U^{(L_n)}}(\mathbf{u}, m) \cdot m \\ &= \sum_{m=1}^{\beta_n} \sum_{i=1}^{|S_n(m)|} \sum_{\mathbf{u} \in A_i^{(m)}} P_{U^{(L_n)}}(\mathbf{u}, m) \cdot m \\ &= \sum_{m=1}^{\beta_n} \sum_{\mathbf{x}_i \in S_n(m)} P_{\tilde{X}^n}(\mathbf{x}_i) \cdot m, \end{aligned} \tag{131}$$

where we have used $\mathcal{U}^m = \bigcup_{i=1}^{|S_n(m)|} A_i^{(m)}$ and (128). For $\mathbf{x}_i \in S_n(m)$ we obtain from (129) and the right inequality of (130)

$$\begin{aligned}
 P_{\tilde{X}^n}(\mathbf{x}_i) &\leq \frac{P_{V^n}(\mathbf{x}_i)}{1 - \gamma_0} + \frac{\Pr\{V^n \in S_n(m)\}}{(1 - \gamma_0)K^m} \\
 &= \frac{1}{1 - \gamma_0} \left(P_{V^n}(\mathbf{x}_i) + \frac{\Pr\{V^n \in S_n(m)\}}{K^m} \right) \\
 &\leq \frac{1}{1 - \gamma_0} \left(1 + \frac{1}{P_{V^n}(\mathbf{x}_i)K^m} \right) P_{V^n}(\mathbf{x}_i) \\
 &\leq (1 + 2\gamma) \left(1 + \frac{1}{K^{n\gamma}} \right) P_{V^n}(\mathbf{x}_i),
 \end{aligned} \tag{132}$$

where, to derive the last inequality, we have used the fact $0 \leq \gamma_0 \leq \gamma \leq \frac{1}{2}$ and

$$P_{V^n}(\mathbf{x}_i) \geq K^{-(m-n\gamma)} \quad (\forall \mathbf{x}_i \in S_n(m)). \tag{133}$$

It should be noticed that (132) also implies that

$$P_{\tilde{X}^n}(\mathbf{x}) \leq (1 + 2\gamma) \left(1 + \frac{1}{K^{n\gamma}} \right) P_{V^n}(\mathbf{x}) \quad (\forall \mathbf{x} \in \mathcal{X}^n) \tag{134}$$

since $P_{\tilde{X}^n}(\mathbf{x}) = 0$ for $\mathbf{x} \notin T_n = \bigcup_{m=1}^{\beta_n} S_n(m)$. Plugging the inequality

$$m \leq \log \frac{1}{P_{V^n}(\mathbf{x}_i)} + n\gamma + 1 \quad (\forall \mathbf{x}_i \in S_n(m)) \tag{135}$$

and (132) into (131), we obtain

$$\begin{aligned}
 \mathbb{E}[L_n] &\leq (1 + 2\gamma) \left(1 + \frac{1}{K^{n\gamma}} \right) \\
 &\quad \cdot \sum_{m=1}^{\beta_n} \sum_{\mathbf{x}_i \in S_n(m)} P_{V^n}(\mathbf{x}_i) \left(\log \frac{1}{P_{V^n}(\mathbf{x}_i)} + n\gamma + 1 \right) \\
 &\leq (1 + 2\gamma) \left(1 + \frac{1}{K^{n\gamma}} \right) (H(V^n) + n\gamma + 1).
 \end{aligned} \tag{136}$$

Thus, we obtain from (136)

$$\begin{aligned}
 \limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[L_n] &\leq (1 + 2\gamma) \left\{ \limsup_{n \rightarrow \infty} \frac{1}{n} H(V^n) + \gamma \right\} \\
 &\leq (1 + 2\gamma) \left\{ \limsup_{n \rightarrow \infty} \frac{1}{n} H_{[\delta+\gamma]}^D(X^n) + 2\gamma \right\} \\
 &\leq (1 + 2\gamma)(H^+ + 2\gamma),
 \end{aligned} \tag{137}$$

where the second inequality follows from (110). Since we have assumed that H^+ is finite and $\gamma \in (0, \frac{1}{2}]$ is arbitrary, the r.h.s. of (137) can be made as close to H^+ as desired. Therefore, for all sufficiently small $\gamma > 0$, we obtain

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[L_n] \leq H^+ + \mu = R \tag{138}$$

(c) *Evaluation of Divergence:*

The divergence $D(\tilde{X}^n || X^n)$ can be rewritten as

$$D(\tilde{X}^n || X^n) = D(\tilde{X}^n || V^n) + \mathbb{E} \left[\log \frac{P_{V^n}(\tilde{X}^n)}{P_{X^n}(\tilde{X}^n)} \right]. \tag{139}$$

In view of (132), we obtain

$$\begin{aligned}
 D(\tilde{X}^n || V^n) &= \sum_{m=1}^{\beta_n} \sum_{\mathbf{x} \in \mathcal{S}_n(m)} P_{\tilde{X}^n}(\mathbf{x}) \log \frac{P_{\tilde{X}^n}(\mathbf{x})}{P_{V^n}(\mathbf{x})} \\
 &\leq \sum_{m=1}^{\beta_n} \sum_{\mathbf{x} \in \mathcal{S}_n(m)} P_{\tilde{X}^n}(\mathbf{x}) \log \left\{ (1 + 2\gamma) \left(1 + \frac{1}{K^{n\gamma}} \right) \right\} \\
 &\leq \frac{2\gamma}{\ln K} + \log \left(1 + \frac{1}{K^{n\gamma}} \right)
 \end{aligned} \tag{140}$$

and

$$\begin{aligned}
 \mathbb{E} \left[\log \frac{P_{V^n}(\tilde{X}^n)}{P_{X^n}(\tilde{X}^n)} \right] &= \sum_{\mathbf{x} \in \mathcal{X}^n} P_{\tilde{X}^n}(\mathbf{x}) \log \frac{P_{V^n}(\mathbf{x})}{P_{X^n}(\mathbf{x})} \\
 &\leq (1 + 2\gamma) \left(1 + \frac{1}{K^{n\gamma}} \right) D(V^n || X^n) \\
 &\leq (1 + 2\gamma)(\delta + \gamma) \left(1 + \frac{1}{K^{n\gamma}} \right),
 \end{aligned} \tag{141}$$

where to obtain the last inequality we used the fact that $P_{V^n} \in B_{\delta+\gamma}^D(X^n)$. Plugging (140) and (141) into (139) yields

$$\begin{aligned}
 \limsup_{n \rightarrow \infty} D(\tilde{X}^n || X^n) &\leq \frac{2\gamma}{\ln K} + (1 + 2\gamma)(\delta + \gamma) \\
 &\leq \delta + \gamma(2\delta + 5),
 \end{aligned} \tag{142}$$

where we have used the fact that $\frac{2\gamma}{\ln K} \leq 3\gamma$ for all $K \geq 2$ and the assumption $0 < \gamma \leq \frac{1}{2}$ to derive the last inequality. Since $\gamma \in (0, \frac{1}{2}]$ is arbitrary and we have (138), R is $v_D(\delta)$ -achievable.

□

5. Mean and VL Channel Resolvability

So far, we have studied the problem of *source resolvability*, whereas the problem of *channel resolvability* has been introduced by Han and Verdú [6] to investigate the capacity of identification codes [11]. In a conventional problem of this kind, a target *output distribution* P_{Y^n} via a channel W^n due to an input X^n is approximated by encoding the FL uniform random number U_{M_n} as a channel input. In this section, we generalize the problem of such channel resolvability to that in the variable-length setting.

5.1. Definitions

Let \mathcal{X} and \mathcal{Y} be finite or countably infinite alphabets. Let $W = \{W^n\}_{n=1}^\infty$ be a general channel, where $W^n : \mathcal{X}^n \rightarrow \mathcal{Y}^n$ denotes a stochastic mapping. We denote by $Y = \{Y^n\}_{n=1}^\infty$ the output process via W due to an input process $X = \{X^n\}_{n=1}^\infty$, where X^n and Y^n take values in \mathcal{X}^n and \mathcal{Y}^n , respectively. Again, we do not impose any assumptions such as stationarity or ergodicity on either X or W . As in the previous sections, we will identify X^n and Y^n with their probability distributions P_{X^n} and P_{Y^n} , respectively, and these symbols are used interchangeably.

In this section, we consider several types of problems of approximating a target output distribution P_{Y^n} . The first one is the problem of *mean-resolvability* [6], in which the channel input is allowed to be an arbitrary general source.

Definition 5 (mean δ -channel resolvability: variational distance). *Let $\delta \in [0, 1)$ be fixed arbitrarily. A resolution rate $R \geq 0$ is said to be mean δ -achievable for X (under the variational distance) if there exists a general source $\tilde{X} = \{\tilde{X}^n\}_{n=1}^\infty$ satisfying*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} H(\tilde{X}^n) \leq R, \tag{143}$$

$$\limsup_{n \rightarrow \infty} d(P_{Y^n}, P_{\tilde{Y}^n}) \leq \delta, \tag{144}$$

where \tilde{Y}^n denotes the output via W^n due to input \tilde{X}^n . The infimum of all mean δ -achievable rates for \mathbf{X} , i.e.,

$$\bar{S}_v(\delta|\mathbf{X}, \mathbf{W}) := \inf\{R : R \text{ is mean } \delta\text{-achievable for } \mathbf{X}\} \tag{145}$$

is referred to as the mean δ -resolvability for \mathbf{X} . We also define the mean δ -resolvability for the worst input as

$$\bar{S}_v(\delta|\mathbf{W}) := \sup_{\mathbf{X}} \bar{S}_v(\delta|\mathbf{X}, \mathbf{W}). \tag{146}$$

On the other hand, we may also consider the problem of VL channel resolvability. Here, the VL uniform random number $U^{(L_n)}$ is defined as in the foregoing sections. Consider the problem of approximating the target output distribution P_{Y^n} via W^n due to X^n by using another input $\tilde{X}^n = \varphi_n(U^{(L_n)})$ with a deterministic mapping $\varphi_n : \mathcal{U}^* \rightarrow \mathcal{X}^n$.

Definition 6 (VL δ -channel resolvability: variational distance). Let $\delta \in [0, 1)$ be fixed arbitrarily. A resolution rate $R \geq 0$ is said to be VL δ -achievable or simply $v(\delta)$ -achievable for \mathbf{X} (under the variational distance) if there exists a VL uniform random number $U^{(L_n)}$ and a deterministic mapping $\varphi_n : \mathcal{U}^* \rightarrow \mathcal{X}^n$ satisfying

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[L_n] \leq R, \tag{147}$$

$$\limsup_{n \rightarrow \infty} d(P_{Y^n}, P_{\tilde{Y}^n}) \leq \delta, \tag{148}$$

where $\mathbb{E}[\cdot]$ denotes the expected value and \tilde{Y}^n denotes the output via W^n due to input $\tilde{X}^n = \varphi_n(U^{(L_n)})$. The infimum of all $v(\delta)$ -achievable rates for \mathbf{X} , i.e.,

$$S_v(\delta|\mathbf{X}, \mathbf{W}) := \inf\{R : R \text{ is } v(\delta)\text{-achievable for } \mathbf{X}\} \tag{149}$$

is called the VL δ -channel resolvability or simply $v(\delta)$ -channel resolvability for \mathbf{X} . We also define the VL δ -channel resolvability or simply $v(\delta)$ -channel resolvability for the worst input as

$$S_v(\delta|\mathbf{W}) := \sup_{\mathbf{X}} S_v(\delta|\mathbf{X}, \mathbf{W}). \tag{150}$$

When W^n is the identity mapping, the problem of channel resolvability reduces to that of source resolvability, which has been investigated in the foregoing sections. In this sense, the problem of channel resolvability is a generalization of the problem of source resolvability.

Similarly to the problem of source resolvability, we may also use the divergence between the target output distribution P_{Y^n} and the approximated output distribution $P_{\tilde{Y}^n}$ as an approximation measure.

Definition 7 (mean δ -channel resolvability: divergence). Let $\delta \geq 0$ be fixed arbitrarily. A resolution rate $R \geq 0$ is said to be mean δ -achievable for \mathbf{X} (under the divergence) if there exists a general source $\tilde{\mathbf{X}} = \{\tilde{X}^n\}_{n=1}^\infty$ satisfying

$$\limsup_{n \rightarrow \infty} \frac{1}{n} H(\tilde{X}^n) \leq R, \tag{151}$$

$$\limsup_{n \rightarrow \infty} D(\tilde{Y}^n || Y^n) \leq \delta, \tag{152}$$

where \tilde{Y}^n denotes the output via W^n due to input \tilde{X}^n . The infimum of all mean δ -achievable rates for \mathbf{X} , i.e.,

$$\bar{S}_v^D(\delta|\mathbf{X}, \mathbf{W}) := \inf\{R : R \text{ is mean } \delta\text{-achievable for } \mathbf{X}\} \tag{153}$$

is referred to as the mean δ -channel resolvability for \mathbf{X} . We also define the mean δ -channel resolvability for the worst input as

$$\bar{S}_v^D(\delta|\mathbf{W}) := \sup_{\mathbf{X}} \bar{S}_v^D(\delta|\mathbf{X}, \mathbf{W}). \tag{154}$$

Definition 8 (VL δ -channel resolvability: divergence). Let $\delta \geq 0$ be fixed arbitrarily. A resolution rate $R \geq 0$ is said to be VL δ -achievable or simply $v_D(\delta)$ -achievable for \mathbf{X} (under the divergence) if there exists a VL uniform random number $U^{(L_n)}$ and a deterministic mapping $\varphi_n : \mathcal{U}^* \rightarrow \mathcal{X}^n$ satisfying

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[L_n] \leq R, \tag{155}$$

$$\limsup_{n \rightarrow \infty} D(\tilde{Y}^n || Y^n) \leq \delta, \tag{156}$$

where $\mathbb{E}[\cdot]$ denotes the expected value and \tilde{Y}^n denotes the output via W^n due to input $\tilde{X}^n = \varphi_n(U^{(L_n)})$. The infimum of all $v_D(\delta)$ -achievable rates for \mathbf{X} , i.e.,

$$S_v^D(\delta|\mathbf{X}, \mathbf{W}) := \inf\{R : R \text{ is } v_D(\delta)\text{-achievable for } \mathbf{X}\} \tag{157}$$

is called the VL δ -channel resolvability or simply $v_D(\delta)$ -channel resolvability for \mathbf{X} . We also define the VL δ -channel resolvability or simply $v_D(\delta)$ -channel resolvability for the worst input as

$$S_v^D(\delta|\mathbf{W}) := \sup_{\mathbf{X}} S_v^D(\delta|\mathbf{X}, \mathbf{W}). \tag{158}$$

Remark 12. Since the outputs of a deterministic mapping $\tilde{X}^n = \varphi_n(U^{(L_n)})$ form a general source $\tilde{\mathbf{X}}$, it holds that

$$\bar{S}_v(\delta|\mathbf{X}, \mathbf{W}) \leq S_v(\delta|\mathbf{X}, \mathbf{W}) \quad (\delta \in [0, 1)), \tag{159}$$

$$\bar{S}_v^D(\delta|\mathbf{X}, \mathbf{W}) \leq S_v^D(\delta|\mathbf{X}, \mathbf{W}) \quad (\delta \geq 0) \tag{160}$$

for any general source \mathbf{X} and general channel \mathbf{W} . These relations lead to the analogous relation for the mean/VL δ -channel resolvability for the worst input:

$$\bar{S}_v(\delta|\mathbf{W}) \leq S_v(\delta|\mathbf{W}) \quad (\delta \in [0, 1)), \tag{161}$$

$$\bar{S}_v^D(\delta|\mathbf{W}) \leq S_v^D(\delta|\mathbf{W}) \quad (\delta \geq 0). \tag{162}$$

5.2. General Formulas

For a given general source $\mathbf{X} = \{X^n\}_{n=1}^\infty$ and a general channel $\mathbf{W} = \{W^n\}_{n=1}^\infty$, let $\mathbf{Y} = \{Y^n\}_{n=1}^\infty$ be the channel output via \mathbf{W} due to input \mathbf{X} . We define

$$H_{[\delta], W^n}(X^n) = \inf_{P_{V^n} \in B_\delta(X^n, W^n)} H(V^n), \tag{163}$$

$$H_{[\delta], W^n}^D(X^n) = \inf_{P_{V^n} \in B_\delta^D(X^n, W^n)} H(V^n) \tag{164}$$

where $H(V^n)$ denotes the Shannon entropy of V^n and $B_\delta(X^n, W^n)$ and $B_\delta^D(X^n, W^n)$ are defined as

$$B_\delta(X^n, W^n) = \{P_{V^n} \in \mathcal{P}(\mathcal{X}^n) : d(P_{Y^n}, P_{Z^n}) \leq \delta\}, \tag{165}$$

$$B_\delta^D(X^n, W^n) = \{P_{V^n} \in \mathcal{P}(\mathcal{X}^n) : D(Z^n || Y^n) \leq \delta\}, \tag{166}$$

respectively, with Z^n defined as the output via W^n due to input V^n . Both $H_{[\delta], W^n}(X^n)$ and $H_{[\delta], W^n}^D(X^n)$ are nonincreasing functions of δ . In addition, we define

$$H_{[\delta],W}(\mathbf{X}) = \limsup_{n \rightarrow \infty} \frac{1}{n} H_{[\delta],W^n}(X^n), \tag{167}$$

$$H_{[\delta],W}^D(\mathbf{X}) = \limsup_{n \rightarrow \infty} \frac{1}{n} H_{[\delta],W^n}^D(X^n), \tag{168}$$

which play an important role in characterizing the mean/VL δ -channel resolvability. We show the general formulas for the mean/VL δ -channel resolvability.

Theorem 5 (with variational distance). *For any input process \mathbf{X} and any general channel \mathbf{W} ,*

$$\bar{S}_v(\delta|\mathbf{X}, \mathbf{W}) = S_v(\delta|\mathbf{X}, \mathbf{W}) = \lim_{\gamma \downarrow 0} H_{[\delta+\gamma],W}(\mathbf{X}) \quad (\delta \in [0, 1)). \tag{169}$$

In particular,

$$\bar{S}_v(\delta|\mathbf{W}) = S_v(\delta|\mathbf{W}) = \sup_{\mathbf{X}} \lim_{\gamma \downarrow 0} H_{[\delta+\gamma],W}(\mathbf{X}) \quad (\delta \in [0, 1)). \tag{170}$$

Theorem 6 (with divergence). *For any input process \mathbf{X} and any general channel \mathbf{W} ,*

$$\bar{S}_v^D(\delta|\mathbf{X}, \mathbf{W}) = S_v^D(\delta|\mathbf{X}, \mathbf{W}) = \lim_{\gamma \downarrow 0} H_{[\delta+\gamma],W}^D(\mathbf{X}) \quad (\delta \geq 0). \tag{171}$$

In particular,

$$\bar{S}_v^D(\delta|\mathbf{W}) = S_v^D(\delta|\mathbf{W}) = \sup_{\mathbf{X}} \lim_{\gamma \downarrow 0} H_{[\delta+\gamma],W}^D(\mathbf{X}) \quad (\delta \geq 0). \tag{172}$$

Remark 13. *It can be easily verified that the variational distance satisfies*

$$d(P_{Y^n}, P_{Z^n}) \leq d(P_{X^n}, P_{V^n}), \tag{173}$$

and, therefore, we have $B_\delta(X^n) \subseteq B_\delta(X^n, W^n)$. This relation and formulas (32) and (169) indicate that

$$S_v(\delta|\mathbf{X}, \mathbf{W}) \leq S_v(\delta|\mathbf{X}) \quad (\delta \in [0, 1)) \tag{174}$$

for any given channel \mathbf{W} . Likewise, it is well known that the divergence satisfies the data processing inequality $D(\tilde{Y}^n||Y^n) \leq D(\tilde{X}^n||X^n)$ [33], and formulas (101) and (171) lead to

$$S_v^D(\delta|\mathbf{X}, \mathbf{W}) \leq S_v^D(\delta|\mathbf{X}) \quad (\delta \geq 0), \tag{175}$$

regardless of channel \mathbf{W} .

Remark 14. *It is obvious that Theorems 5 and 6 reduce to Theorems 3 and 4, respectively, when the channel \mathbf{W} is the identity mapping. Precisely, for the identity mapping $\mathbf{W} = \mathbf{I}$, the mean δ -resolvability and the $v(\delta)$ -channel resolvability for \mathbf{X} are given by*

$$\bar{S}_v(\delta|\mathbf{X}) = S_v(\delta|\mathbf{X}) = \lim_{\gamma \downarrow 0} H_{[\delta+\gamma]}(\mathbf{X}), \tag{176}$$

where $\bar{S}_v(\delta|\mathbf{X})$ denotes the mean δ -resolvability $\bar{S}_v(\delta|\mathbf{X}, \mathbf{W})$ for the identity mapping \mathbf{W} . The analogous relationship holds under the divergence:

$$\bar{S}_v^D(\delta|\mathbf{X}) = S_v^D(\delta|\mathbf{X}) = \lim_{\gamma \downarrow 0} H_{[\delta+\gamma]}^D(\mathbf{X}), \tag{177}$$

where $\bar{S}_v^D(\delta|\mathbf{X})$ denotes the mean δ -resolvability $\bar{S}_v^D(\delta|\mathbf{X}, \mathbf{W})$ for the identity mapping $\mathbf{W} = \mathbf{I}$. Thus, it turns out that Theorems 5 and 6 are indeed generalizations of Theorems 3 and 4.

Proof of Theorems 5 and 6.

(1) **Converse Part:**

Because of the general relationship (159), to prove the converse part of Theorem 5, it suffices to show that

$$\bar{S}_v(\delta|X, W) \geq \lim_{\gamma \downarrow 0} H_{[\delta+\gamma], W}(X). \tag{178}$$

Let R be mean δ -achievable for X under the variational distance. Then, there exists a general source $\tilde{X} = \{\tilde{X}^n\}_{n=1}^\infty$ satisfying (143) and

$$\limsup_{n \rightarrow \infty} \delta_n \leq \delta, \tag{179}$$

where $\delta_n := d(P_{\gamma^n}, P_{\tilde{\gamma}^n})$. Fixing $\gamma > 0$ arbitrarily, we have $\delta_n \leq \delta + \gamma$ for all $n \geq n_0$ with some $n_0 > 0$ and then

$$H_{[\delta+\gamma], W^n}(X^n) \leq H_{[\delta_n], W^n}(X^n) \quad (n \geq n_0) \tag{180}$$

since $H_{[\delta], W^n}(X^n)$ is a nonincreasing function of δ . Since $P_{\tilde{X}^n} \in B_{\delta_n}(X^n, W^n)$, we have $H_{[\delta_n], W^n}(X^n) \leq H(\tilde{X}^n)$. Thus, we obtain from (143)

$$H_{[\delta+\gamma], W}(X) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} H(\tilde{X}^n) \leq R. \tag{181}$$

Since $\gamma > 0$ is an arbitrary constant, this implies (178).

The converse part of Theorem 6 can be proven in an analogous way with due modifications.

(2) Direct Part:

Because of the general relationship (159), to prove the direct part (achievability) of Theorem 5, it suffices to show that, for any fixed $\gamma > 0$, the resolution rate

$$R = \lim_{\gamma \downarrow 0} H_{[\delta+\gamma], W}(X) + 3\gamma \tag{182}$$

is $v(\delta)$ -achievable for X under the variational distance.

Let $P_{V^n} \in B_{\delta+\gamma}(X^n, W^n)$ be a source satisfying

$$H(V^n) \leq H_{[\delta+\gamma], W^n}(X^n) + \gamma. \tag{183}$$

Then, by the same argument to derive (80) and (82) as developed in the proof of the direct part of Theorem 3, we can construct a VL uniform random number $U^{(L_n)}$ and a deterministic mapping $\varphi_n : \mathcal{U}^* \rightarrow \mathcal{X}^n$ satisfying

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[L_n] \leq \lim_{\gamma \downarrow 0} H_{[\delta+\gamma], W}(X) + 3\gamma = R \tag{184}$$

and

$$d(P_{\tilde{X}^n}, P_{V^n}) \leq \frac{1}{2} K^{-n\gamma} + \gamma, \tag{185}$$

where $\tilde{X}^n = \varphi_n(U^{(L_n)})$. Let Z^n denote the output random variable via W^n due to input V^n . Then, letting \tilde{Y}^n be the output via channel W^n due to input \tilde{X}^n , we can evaluate $d(P_{\tilde{Y}^n}, P_{Z^n})$ as

$$\begin{aligned}
 d(P_{\tilde{Y}^n}, P_{Z^n}) &= \frac{1}{2} \sum_{\mathbf{y} \in \mathcal{Y}^n} |P_{\tilde{Y}^n}(\mathbf{y}) - P_{Z^n}(\mathbf{y})| \\
 &= \frac{1}{2} \sum_{\mathbf{y} \in \mathcal{Y}^n} \left| \sum_{\mathbf{x} \in \mathcal{X}^n} W(\mathbf{y}|\mathbf{x}) (P_{\tilde{X}^n}(\mathbf{x}) - P_{V^n}(\mathbf{x})) \right| \\
 &\leq \frac{1}{2} \sum_{\mathbf{y} \in \mathcal{Y}^n} \sum_{\mathbf{x} \in \mathcal{X}^n} W(\mathbf{y}|\mathbf{x}) |P_{\tilde{X}^n}(\mathbf{x}) - P_{V^n}(\mathbf{x})| \\
 &= d(P_{\tilde{X}^n}, P_{V^n}) \leq \frac{1}{2} K^{-n\gamma} + \gamma.
 \end{aligned} \tag{186}$$

Thus, we obtain

$$\begin{aligned}
 \limsup_{n \rightarrow \infty} d(P_{Y^n}, P_{\tilde{Y}^n}) &\leq \limsup_{n \rightarrow \infty} d(P_{Y^n}, P_{Z^n}) + \limsup_{n \rightarrow \infty} d(P_{\tilde{Y}^n}, P_{Z^n}) \\
 &\leq \delta + 2\gamma,
 \end{aligned} \tag{187}$$

where we have used the fact $P_{V^n} \in B_{\delta+\gamma}(X^n, W^n)$ to derive the last inequality. Since $\gamma > 0$ is an arbitrary constant, we can conclude that R is $v(\delta)$ -achievable for \mathbf{X} .

The direct part of Theorem 6 can be proven in the same way as Theorem 4 with due modifications. Fixing $P_{V^n} \in B_{\delta+\gamma}^D(X^n, W^n)$ and using the encoding scheme as developed in the proof of Theorem 4, the evaluation of the average length rate is exactly the same, and we can obtain (138). A key step is to evaluate the divergence $D(\tilde{Y}^n || Y^n)$, which can be rewritten as

$$D(\tilde{Y}^n || Y^n) = D(\tilde{Y}^n || Z^n) + \mathbb{E} \left[\log \frac{P_{Z^n}(\tilde{Y}^n)}{P_{Y^n}(\tilde{Y}^n)} \right]. \tag{188}$$

The first term on the r.h.s. can be bounded as

$$D(\tilde{Y}^n || Z^n) \leq D(\tilde{X}^n || V^n) \leq \frac{2\gamma}{\ln K} + \log \left(1 + \frac{1}{K^{n\gamma}} \right) \tag{189}$$

as in (140), where the left inequality is due to the data processing inequality. Similarly to the derivation of (141), the second term can be bounded as

$$\begin{aligned}
 \mathbb{E} \left[\log \frac{P_{Z^n}(\tilde{Y}^n)}{P_{Y^n}(\tilde{Y}^n)} \right] &= \sum_{\mathbf{y} \in \mathcal{Y}^n} \sum_{\mathbf{x} \in \mathcal{X}^n} P_{\tilde{X}^n}(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x}) \log \frac{P_{Z^n}(\mathbf{y})}{P_{Y^n}(\mathbf{y})} \\
 &\leq (1 + 2\gamma) \left(1 + \frac{1}{K^{n\gamma}} \right) \sum_{\mathbf{y} \in \mathcal{Y}^n} \sum_{\mathbf{x} \in \mathcal{X}^n} P_{V^n}(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x}) \log \frac{P_{Z^n}(\mathbf{y})}{P_{Y^n}(\mathbf{y})} \\
 &= (1 + 2\gamma) \left(1 + \frac{1}{K^{n\gamma}} \right) D(Z^n || Y^n),
 \end{aligned} \tag{190}$$

where we have used (134). Here, $D(Z^n || Y^n) \leq \delta + \gamma$ because Z^n is the output via W^n due to input $V^n \in B_{\delta+\gamma}^D(X^n, W^n)$. The rest of the steps are the same as in the proof of Theorem 4.

□

6. Second-Order VL Channel Resolvability

So far, we have analyzed the first-order VL resolvabilities and established various first-order resolvability theorems. One of the next important steps is the second-order analysis, and so, in this section, we generalize VL resolvabilities in the second-order setting.

6.1. Definitions

We now turn to considering the *second-order* resolution rates [26,27,29]. First, we consider the VL resolvability based on the variational distance.

Definition 9 (VL (δ, R) -channel resolvability: variational distance). A second-order resolution rate $L \in (-\infty, +\infty)$ is said to be VL (δ, R) -achievable (under the variational distance) for \mathbf{X} with $\delta \in [0, 1)$ if there exist a VL uniform random number $U^{(L_n)}$ and a deterministic mapping $\varphi_n : \mathcal{U}^* \rightarrow \mathcal{X}^n$ satisfying

$$\limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} (\mathbb{E}[L_n] - nR) \leq L, \tag{191}$$

$$\limsup_{n \rightarrow \infty} d(P_{Y^n}, P_{\tilde{Y}^n}) \leq \delta, \tag{192}$$

where \tilde{Y}^n denotes the output via W^n due to input $\tilde{X}^n = \varphi_n(U^{(L_n)})$. The infimum of all VL (δ, R) -achievable rates for \mathbf{X} is denoted by

$$T_v(\delta, R|\mathbf{X}, \mathbf{W}) := \inf\{L : L \text{ is VL } (\delta, R)\text{-achievable for } \mathbf{X}\}. \tag{193}$$

When \mathbf{W} is the identity mapping \mathbf{I} , $T_v(\delta, R|\mathbf{X}, \mathbf{W})$ is simply denoted by $T_v(\delta, R|\mathbf{X})$ (source resolvability).

Next, we may consider the VL resolvability with the divergence instead of the variational distance.

Definition 10 (VL (δ, R) -channel resolvability: divergence). A second-order resolution rate $L \in (-\infty, +\infty)$ is said to be VL (δ, R) -achievable for \mathbf{X} (with the divergence) where $\delta \geq 0$ if there exists a VL uniform random number $U^{(L_n)}$ and a deterministic mapping $\varphi_n : \mathcal{U}^* \rightarrow \mathcal{X}^n$ satisfying

$$\limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} (\mathbb{E}[L_n] - nR) \leq L, \tag{194}$$

$$\limsup_{n \rightarrow \infty} D(\tilde{Y}^n || Y^n) \leq \delta, \tag{195}$$

where \tilde{Y}^n denotes the output via W^n due to input $\tilde{X}^n = \varphi_n(U^{(L_n)})$. The infimum of all VL (δ, R) -achievable rates for \mathbf{X} is denoted as

$$T_v^D(\delta, R|\mathbf{X}, \mathbf{W}) := \inf\{L : L \text{ is VL } (\delta, R)\text{-achievable for } \mathbf{X}\}. \tag{196}$$

When \mathbf{W} is the identity mapping \mathbf{I} , $T_v^D(\delta, R|\mathbf{X}, \mathbf{W})$ is simply denoted by $T_v^D(\delta, R|\mathbf{X})$ (source resolvability).

Remark 15. It is easily verified that

$$T_v(\delta, R|\mathbf{X}, \mathbf{W}) = \begin{cases} +\infty & \text{for } R < S_v(\delta|\mathbf{X}, \mathbf{W}) \\ -\infty & \text{for } R > S_v(\delta|\mathbf{X}, \mathbf{W}). \end{cases} \tag{197}$$

Hence, only the case $R = S_v(\delta|\mathbf{X}, \mathbf{W})$ is of interest to us. The same remark also applies to $T_v^D(\delta, R|\mathbf{X}, \mathbf{W})$.

6.2. General Formulas

We establish general formulas for the second-order resolvability. The proofs of the following theorems are given below subsequently to Remark 17.

Theorem 7 (with variational distance). For any input process \mathbf{X} and general channel \mathbf{W} ,

$$T_v(\delta, R|\mathbf{X}, \mathbf{W}) = \lim_{\gamma \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \left(H_{[\delta+\gamma], W^n}(X^n) - nR \right) \quad (\delta \in [0, 1), R \geq 0). \tag{198}$$

In particular, in the case where \mathbf{W} is the identity mapping \mathbf{I} ,

$$T_v(\delta, R|\mathbf{X}) = \lim_{\gamma \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \left(H_{[\delta+\gamma]}(X^n) - nR \right) \quad (\delta \in [0, 1), R \geq 0). \tag{199}$$

Theorem 8 (with divergence). For any input process \mathbf{X} and general channel \mathbf{W} ,

$$T_v^D(\delta, R|\mathbf{X}, \mathbf{W}) = \lim_{\gamma \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \left(H_{[\delta+\gamma], \mathbf{W}^n}^D(X^n) - nR \right) \quad (\delta \geq 0, R \geq 0). \tag{200}$$

In particular, in the case where \mathbf{W} is the identity mapping \mathbf{I} ,

$$T_v^D(\delta, R|\mathbf{X}) = \lim_{\gamma \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \left(H_{[\delta+\gamma]}^D(X^n) - nR \right) \quad (\delta \geq 0, R \geq 0). \tag{201}$$

Remark 16. As discussed in Section 5, we may also consider using a general source $\tilde{\mathbf{X}}$ as an input to channel \mathbf{W} , and we can define L to be a mean (δ, R) -achievable rate for \mathbf{X} by replacing (191) and (194) with

$$\limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} (H(\tilde{X}^n) - nR) \leq L. \tag{202}$$

Let $\bar{T}_v(\delta, R|\mathbf{X}, \mathbf{W})$ and $\bar{T}_v^D(\delta, R|\mathbf{X}, \mathbf{W})$ denote the infimum of all mean (δ, R) -achievable rates for \mathbf{X} under the variational distance and the divergence, respectively. Then, it is not difficult to verify that

$$\bar{T}_v(\delta, R|\mathbf{X}, \mathbf{W}) = T_v(\delta, R|\mathbf{X}, \mathbf{W}) \quad (\delta \in [0, 1]), \tag{203}$$

$$\bar{T}_v^D(\delta, R|\mathbf{X}, \mathbf{W}) = T_v^D(\delta, R|\mathbf{X}, \mathbf{W}) \quad (\delta \geq 0). \tag{204}$$

Thus, there is no loss in the (δ, R) -achievable resolution rate even if the channel input $\tilde{\mathbf{X}}$ is restricted to be generated by the VL uniform random number $U^{(L_n)}$.

Remark 17. As in the first-order case, when the channel \mathbf{W} is the identity mapping \mathbf{I} , $T_v(\delta, R|\mathbf{X})$ coincides with the minimum second-order length rate of the VL source codes. More precisely, we denote by $R_v^*(\delta, R|\mathbf{X})$ the minimum second-order length rate of a sequence of VL source codes with the first-order average length rate R and an average error probability asymptotically not exceeding δ . Yagi and Nomura [31] have shown that

$$R_v^*(\delta, R|\mathbf{X}) = \lim_{\gamma \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \left(G_{[\delta+\gamma]}(X^n) - nR \right) \quad (\delta \in [0, 1], R \geq 0). \tag{205}$$

Modifying the proof of Proposition 1 (cf. Appendix C), we can show that the r.h.s. of (199) coincides with that of (205), and, therefore, it generally holds that

$$T_v(\delta, R|\mathbf{X}) = R_v^*(\delta, R|\mathbf{X}) \quad (\delta \in [0, 1], R \geq 0). \tag{206}$$

As a special case, suppose that \mathbf{X} is a stationary and memoryless source X with the finite third absolute moment of $\log \frac{1}{P_X(X)}$. In this case, Kostina et al. [25] have recently given a single-letter characterization for $R_v^*(\delta, R|\mathbf{X})$ with $R = H_{[\delta]}(X) = (1 - \delta)H(X)$ as

$$R_v^*(\delta, R|\mathbf{X}) = -\sqrt{\frac{V(X)}{2\pi}} e^{-\frac{(Q^{-1}(\delta))^2}{2}}, \tag{207}$$

where $V(X)$ denotes the variance of $\log \frac{1}{P_X(X)}$ (varentropy) and Q^{-1} is the inverse of the complementary cumulative distribution function of the standard Gaussian distribution. In view of the general relation (206), we can also obtain the single-letter characterization for $T_v(\delta, R|\mathbf{X})$:

$$T_v(\delta, R|\mathbf{X}) = -\sqrt{\frac{V(X)}{2\pi}} e^{-\frac{(Q^{-1}(\delta))^2}{2}}. \tag{208}$$

It has not yet been made clear whether we can also have a single-letter formula for $T_v(\delta, R|\mathbf{X}, \mathbf{W})$ when the channel \mathbf{W} is memoryless but not necessarily the identity mapping.

Proof of Theorems 7 and 8.

(1) Converse Part:

We will show the converse part of Theorem 7. The converse part of Theorem 8 can be proved in an analogous way.

Let L be VL (δ, R) -achievable for X under the variational distance. Then, there exists $U^{(L_n)}$ and φ_n satisfying (191) and

$$\limsup_{n \rightarrow \infty} \delta_n \leq \delta, \tag{209}$$

where we define $\delta_n = d(P_{Y^n}, P_{\tilde{Y}^n})$, and \tilde{Y}^n is the output via W^n due to input $\tilde{X}^n = \varphi_n(U^{(L_n)})$. Equation (209) implies that, for any given $\gamma > 0$, it holds that $\delta_n \leq \delta + \gamma$ for all $n \geq n_0$ with some $n_0 > 0$, and, therefore,

$$H_{[\delta+\gamma], W^n}(X^n) \leq H_{[\delta_n], W^n}(X^n) \quad (\forall n \geq n_0). \tag{210}$$

Since $P_{\tilde{X}^n} \in B_{\delta_n}(X^n, W^n)$, we have

$$H_{[\delta_n], W^n}(X^n) \leq H(\tilde{X}^n). \tag{211}$$

On the other hand, it follows from (23) that

$$H(\tilde{X}^n) \leq H(U^{(L_n)}) = \mathbb{E}[L_n] + H(L_n), \tag{212}$$

where the inequality is due to the fact that φ_n is a deterministic mapping and $\tilde{X}^n = \varphi_n(U^{(L_n)})$.

Combining (210)–(212) yields

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \left(H_{[\delta+\gamma], W^n}(X^n) - nR \right) \\ & \leq \limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \left(H(\tilde{X}^n) - nR \right) \\ & \leq \limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \left(\mathbb{E}[L_n] - nR \right) + \limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} H(L_n) \leq L, \end{aligned} \tag{213}$$

where we have used (24) and (191) for the last inequality. Since $\gamma > 0$ is arbitrary, we have

$$\lim_{\gamma \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \left(H_{[\delta+\gamma], W^n}(X^n) - nR \right) \leq L. \tag{214}$$

(2) Direct Part:

We will show the direct part (achievability) of Theorem 7 by modifying the argument of Theorems 3 and 5, whereas the direct part of Theorem 8 can be proved in a similar manner by modifying that of Theorem 4 instead of Theorem 3.

Letting

$$L = \lim_{\gamma \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \left(H_{[\delta+\gamma], W^n}(X^n) - nR \right) + 2\gamma, \tag{215}$$

where $\gamma > 0$ is an arbitrary constant, we shall show that L is VL (δ, R) -achievable for X under the variational distance.

We use the same achievability scheme as in the proof of Theorem 3 with slightly different parameter settings. For $\gamma > 0$, we choose a $c_n > 0$ so that

$$\Pr\{V^n \notin T_n\} \leq \gamma \tag{216}$$

where $P_{V^n} \in B_{\delta+\gamma}(X^n, W^n)$ with $H_{[\delta+\gamma], W^n}(X^n) + \gamma \geq H(V^n)$ and

$$T_n := \left\{ \mathbf{x} \in \mathcal{X}^n : \frac{1}{n} \log \frac{1}{P_{V^n}(\mathbf{x})} \leq c_n \right\}. \tag{217}$$

We here define

$$\ell(\mathbf{x}) := \begin{cases} \lceil \log \frac{1}{P_{V^n}(\mathbf{x})} + \sqrt{n}\gamma \rceil & \text{for } \mathbf{x} \in T_n \\ 0 & \text{otherwise} \end{cases} \tag{218}$$

and $\beta_n = \lceil (nc_n + \sqrt{n}\gamma) \rceil$. Arguing similarly to the proof of Theorems 3 and 5, we can show that there exist $\varphi_n : \mathcal{U}^* \rightarrow \mathcal{X}^n$ and $U^{(L_n)}$ such that

$$d(P_{Y^n}, P_{\tilde{Y}^n}) \leq \delta + 2\gamma + \frac{1}{2}K^{-\sqrt{n}\gamma} \tag{219}$$

and

$$\mathbb{E}[L_n] \leq \left(1 + \frac{1}{K\sqrt{n}\gamma} \right) \left(H_{[\delta+\gamma], W^n}(X^n) + 2\sqrt{n}\gamma + 1 \right). \tag{220}$$

Therefore, we obtain

$$\lim_{n \rightarrow \infty} d(P_{Y^n}, P_{\tilde{Y}^n}) \leq \delta + 2\gamma \tag{221}$$

and

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} (\mathbb{E}[L_n] - nR) \\ & \leq \limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \left(H_{[\delta+\gamma], W^n}(X^n) - nR \right) + 2\gamma \\ & \leq \lim_{\gamma \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \left(H_{[\delta+\gamma], W^n}(X^n) - nR \right) + 2\gamma = L. \end{aligned} \tag{222}$$

Since $\gamma > 0$ is arbitrary, L is VL (δ, R) -achievable for \mathbf{X} .

□

7. Conclusions

We have investigated the problem of *VL source/channel resolvability*, in which a given target probability distribution is approximated by transforming *VL uniform random numbers*. Table 1 summarizes various first-order resolvabilities and their characterizations in terms of information quantities. In this table, the theorem numbers that contain the corresponding characterization are also indicated.

In this paper, we have first analyzed the fundamental limits on the VL δ -source resolvability with the variational distance in Theorem 3. The VL δ -source resolvability is essentially characterized in terms of smooth Shannon entropies. In the proof of the direct part, we have developed a simple method for information spectrum slicing, in which sliced information densities quantized to the same integer are approximated by an FL uniform random number of the same length. Next, we have extended the analysis to the δ -source resolvability under the unnormalized divergence in Theorem 4. The smoothed entropy with the divergence again plays an important role in characterizing the δ -source resolvability.

Table 1. Summary of First-Order Resolvability and Information Quantities.

Approximation Measure	Resolvability	Characterization	Theorem #
Fixed-Length Resolvability			
Variational Distance	$S_f(\mathbf{X})$	$\bar{H}(\mathbf{X})$	Theorem 1 [6]
	$S_f(\delta\mathbf{X})$	$\bar{H}_\delta(\mathbf{X})$	Theorem 2 [7]
Variable-Length Resolvability			
Variational Distance	$S_v(\delta\mathbf{X})$	$\lim_{\gamma \downarrow 0} H_{[\delta+\gamma]}(\mathbf{X})$	Theorem 3
	$S_v(\delta\mathbf{X}, \mathbf{W})$	$\lim_{\gamma \downarrow 0} H_{[\delta+\gamma], \mathbf{W}}(\mathbf{X})$	Theorem 5
Divergence	$S_v^D(\delta\mathbf{X})$	$\lim_{\gamma \downarrow 0} H_{[\delta+\gamma]}^D(\mathbf{X})$	Theorem 4
	$S_v^D(\delta\mathbf{X}, \mathbf{W})$	$\lim_{\gamma \downarrow 0} H_{[\delta+\gamma], \mathbf{W}}^D(\mathbf{X})$	Theorem 6

Then, we have addressed the problem of δ -channel resolvability. It has been revealed in Theorems 5 and 6 that using an arbitrary general source as a coin distribution (mean-resolvability problem) cannot go beyond the fundamental limits of the VL resolvability, in which only VL uniform random numbers are allowed to be a coin distribution. As in the case of source resolvability, we have discussed the δ -channel resolvability under the variational distance and the unnormalized divergence. The second-order channel resolvability has been characterized in Theorems 7 and 8 as well as the first-order case. We notice here that a counterpart of the VL uniform random number is the problem of VL source coding, for which the general treatment, focused on overflow/underflow probabilities, is found in [38]. Indeed, when the variational distance is used as an approximation measure, it turns out that the δ -source resolvability is equal to the minimum achievable rate of VL source codes with an error probability of less than or equal to δ . This is a parallel relationship between *FL source resolvability* and the minimum achievable rate of *FL source codes* [6,7]. It is of interest to investigate whether there is a coding problem to which the δ -channel resolvability is closely related.

When $\delta = 0$, asymptotically exact approximation is required. In the case where the channel \mathbf{W} is the identity mapping \mathbf{I} , it turned out that the source resolvability under the variational distance and the unnormalized divergence coincides and is given by $\lim_{\gamma \downarrow 0} H_{[\gamma]}(\mathbf{X})$, where \mathbf{X} is the general target source. This result is analogous to the dual problem of *VL intrinsic randomness* [5,36], in which the maximum achievable rates of VL uniform random numbers extracted from a given source \mathbf{X} are the same under two kinds of approximation measures. It should be emphasized that in the case of VL intrinsic randomness, the use of *normalized* divergence as an approximation measure results in the same general formula as with the variational distance and the unnormalized divergence, which does not necessarily hold in the case of mean/VL resolvability (cf. Remark 11). It is also noteworthy that whereas only the case of $\delta = 0$ has been completely solved for the VL intrinsic randomness, we have also dealt with the case of a $\delta > 0$ for the VL source/channel resolvability.

When \mathbf{X} is a stationary and memoryless source or is even with a one-point spectrum (cf. Corollary 1), the formulas established reduce to a single-letter characterization for the first- and second-order source resolvability under the variational distance. In the case where the divergence is an approximation measure and/or the channel \mathbf{W} is a non-identity mapping, however, it has not yet been made clear whether we can derive a single-letter characterization for the δ -source/channel resolvability. This question remains to be studied.

As noted in Remark 10, the order of arguments in divergence $D(\tilde{X}^n || X^n)$ is important, and it seems difficult to extend the analyses in this paper to the case with the reversed

$D(X^n || \tilde{X}^n)$ used as an approximation measure. In the context of intrinsic randomness, the reversed divergence is also discussed [23]. Investigating the problem of source/channel resolvability using such divergence is an interesting research topic.

Author Contributions: The author H.Y. contributed to the conceptualization of the research goals and aims, the visualization/presentation of the works, the formal analysis of the results, and the writing of the paper. The author T.S.H. contributed to the formal analysis of the results, the validation of obtained results, and the review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by JSPS KAKENHI Grant Numbers JP16K06340, JP18H01438, and JP20K04462.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Equation (31)

(i) We first show $\lim_{\alpha \uparrow 1} H_{[\delta]}^\alpha(X^n) \leq H_{[\delta]}(X^n)$.

Fix $\gamma > 0$ arbitrarily. We choose $P_{V^n} \in B_\delta(X^n)$ satisfying

$$H_{[\delta]}(X^n) + \gamma \geq H(V^n). \quad (\text{A1})$$

It is well known that the Rényi entropy of order α defined as

$$H^\alpha(V^n) = \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}^n} P_{V^n}(x)^\alpha \quad (\forall \alpha \in (0, 1) \cup (1, +\infty))$$

satisfies

$$H(V^n) = \lim_{\alpha \rightarrow 1} H^\alpha(V^n). \quad (\text{A2})$$

By definition, we have

$$H^\alpha(V^n) \geq H_{[\delta]}^\alpha(X^n) \quad (\forall \alpha \in (0, 1) \cup (1, +\infty)), \quad (\text{A3})$$

leading to

$$H(V^n) = \lim_{\alpha \uparrow 1} H^\alpha(V^n) \geq \lim_{\alpha \uparrow 1} H_{[\delta]}^\alpha(X^n). \quad (\text{A4})$$

Combining (A1) and (A4) yields

$$H_{[\delta]}(X^n) + \gamma \geq H(V^n) \geq \lim_{\alpha \uparrow 1} H_{[\delta]}^\alpha(X^n). \quad (\text{A5})$$

Since $\gamma > 0$ is an arbitrary constant, we obtain $\lim_{\alpha \uparrow 1} H_{[\delta]}^\alpha(X^n) \leq H_{[\delta]}(X^n)$.

(ii) Next, we shall show $\lim_{\alpha \uparrow 1} H_{[\delta]}^\alpha(X^n) \geq H_{[\delta]}(X^n)$.

Fix $\gamma > 0$ arbitrarily. We choose some $\alpha_0 \in (0, 1)$ satisfying

$$\lim_{\alpha \uparrow 1} H_{[\delta]}^\alpha(X^n) + \gamma \geq H_{[\delta]}^{\alpha_0}(X^n). \quad (\text{A6})$$

For this α_0 we choose $P_{V^n} \in B_\delta(X^n)$ satisfying

$$H_{[\delta]}^{\alpha_0}(X^n) + \gamma \geq H^{\alpha_0}(V^n). \quad (\text{A7})$$

Since $H^\alpha(V^n)$ is a nonincreasing function of α , we have

$$H^{\alpha_0}(V^n) \geq H(V^n), \tag{A8}$$

and it follows from (A6)–(A8) that

$$\lim_{\alpha \uparrow 1} H_{[\delta]}^{\alpha}(X^n) + 2\gamma \geq H^{\alpha_0}(V^n) \geq H(V^n). \tag{A9}$$

Since $H(V^n) \geq H_{[\delta]}(X^n)$, due to $P_{V^n} \in B_{\delta}(X^n)$, and $\gamma > 0$ is arbitrarily fixed, we obtain the desired inequality.

Appendix B. Proof of Equation (34)

To prove the alternative formula (34) for the $v(\delta)$ -resolvability $S_v(\delta|X)$, we shall show

$$\lim_{\gamma \downarrow 0} H_{[\delta+\gamma]}(X) = \inf_{V \in B_{\delta}(X)} H(V) \quad (\delta \in [0, 1)). \tag{A10}$$

(i) We first show $\lim_{\gamma \downarrow 0} H_{[\delta+\gamma]}(X) \leq \inf_{V \in B_{\delta}(X)} H(V)$.

Fix $\gamma > 0$ arbitrarily. We choose $\tilde{V} = \{\tilde{V}^n\}_{n=1}^{\infty} \in B_{\delta}(X)$ satisfying

$$H(\tilde{V}) \leq \inf_{V \in B_{\delta}(X)} H(V) + \gamma. \tag{A11}$$

For $\tilde{V} \in B_{\delta}(X)$, we have $d(X^n, \tilde{V}^n) \leq \delta + \gamma$ for all $n \geq n_0$ with some $n_0 > 0$, yielding

$$H_{[\delta+\gamma]}(X^n) \leq H(\tilde{V}^n) \quad (\forall n \geq n_0). \tag{A12}$$

Thus, it follows from (A11) and (A12) that

$$H_{[\delta+\gamma]}(X) \leq \inf_{V \in B_{\delta}(X)} H(V) + \gamma. \tag{A13}$$

Since $\gamma > 0$ is an arbitrary constant, letting $\gamma \downarrow 0$ on both sides yields the desired inequality.

(ii) Next, we shall show $\lim_{\gamma \downarrow 0} H_{[\delta+\gamma]}(X) \geq \inf_{V \in B_{\delta}(X)} H(V)$.

Fix $\lambda > 0$ arbitrarily. We choose an arbitrary decreasing sequence of positive numbers $\{\gamma_i\}_{i=1}^{\infty}$ satisfying $\gamma_1 > \gamma_2 > \dots \rightarrow 0$. Then, we have

$$\lim_{\gamma \downarrow 0} H_{[\delta+\gamma]}(X) = \lim_{i \rightarrow \infty} H_{[\delta+\gamma_i]}(X). \tag{A14}$$

Additionally, by the definition of the limit superior, for each $i = 1, 2, \dots$ we have

$$\frac{1}{n} H_{[\delta+\gamma_i]}(X^n) \leq H_{[\delta+\gamma_i]}(X) + \lambda \quad (\forall n \geq n_i) \tag{A15}$$

with some $0 < n_1 < n_2 < \dots$. Now, for each $n = 1, 2, \dots$, we denote by i_n the index i satisfying

$$n_i \leq n < n_{i+1}. \tag{A16}$$

Then, from (A15), we obtain

$$\frac{1}{n} H_{[\delta+\gamma_{i_n}]}(X^n) \leq H_{[\delta+\gamma_{i_n}]}(X) + \lambda \quad (\forall n \geq n_1). \tag{A17}$$

On the other hand, by the definition of $H_{[\delta+\gamma_{i_n}]}(X^n)$, for each $n = 1, 2, \dots$, we can choose some $V_{i_n}^n \in B_{\delta+\gamma_{i_n}}(X^n)$ satisfying

$$\frac{1}{n} H(V_{i_n}^n) \leq \frac{1}{n} H_{[\delta+\gamma_{i_n}]}(X^n) + \lambda. \tag{A18}$$

We now construct the general source $\tilde{V} = \{V_{i_n}^n\}_{n=1}^{\infty}$. Since $V_{i_n}^n \in B_{\delta+\gamma_{i_n}}(X^n)$ for all $n \geq n_1$ indicates that

$$\limsup_{n \rightarrow \infty} d(X^n, V_{i_n}^n) \leq \delta + \lim_{n \rightarrow \infty} \gamma_{i_n} = \delta, \tag{A19}$$

the general source satisfies $\tilde{V} \in B_\delta(\mathbf{X})$.

From (A17) and (A18), we obtain

$$\frac{1}{n} H(V_{i_n}^n) \leq H_{[\delta + \gamma_{i_n}]}(\mathbf{X}) + 2\lambda \quad (\forall n \geq n_1). \tag{A20}$$

In view of (A14) and the fact $\tilde{V} \in B_\delta(\mathbf{X})$, taking $\limsup_{n \rightarrow \infty}$ on both sides yields

$$\begin{aligned} \inf_{V \in B_\delta(\mathbf{X})} H(V) &\leq H(\tilde{V}) \leq \limsup_{n \rightarrow \infty} H_{[\delta + \gamma_{i_n}]}(\mathbf{X}) + 2\lambda \\ &= \lim_{\gamma \downarrow 0} H_{[\delta + \gamma]}(\mathbf{X}) + 2\lambda. \end{aligned} \tag{A21}$$

Since $\lambda > 0$ is an arbitrary constant, letting $\lambda \downarrow 0$ yields the desired inequality.

Appendix C. Proof of Proposition 1

We shall prove the equality and inequality in (40). Equation (41) is an immediate consequence of (40) because $\bar{H}_\delta(\mathbf{X})$ is a right-continuous function [2].

(i) We first shall show the equality in (40): $H_{[\delta]}(\mathbf{X}) = G_{[\delta]}(\mathbf{X})$.

We show $H_{[\delta]}(\mathbf{X}) \leq G_{[\delta]}(\mathbf{X})$. For any given $\gamma > 0$ and P_{X^n} , let $\mathcal{A}_n^* \subseteq \mathcal{X}^n$ be a subset of \mathcal{X}^n which satisfies

$$\Pr\{X^n \in \mathcal{A}_n^*\} \geq 1 - \delta \tag{A22}$$

and

$$G_{[\delta]}(X^n) + \gamma \geq \sum_{x \in \mathcal{A}_n^*} P_{X^n}(x) \log \frac{1}{P_{X^n}(x)} =: F(\mathcal{A}_n^*). \tag{A23}$$

Choose $x_0 \in \mathcal{X}^n \setminus \mathcal{A}_n^*$ arbitrarily. Set $P_{\tilde{X}^n}$ so that

$$P_{\tilde{X}^n}(x) = \begin{cases} P_{X^n}(x) & \text{for } x \in \mathcal{A}_n^* \\ \alpha_0 & \text{for } x = x_0 \\ 0 & \text{otherwise,} \end{cases} \tag{A24}$$

where we define $\alpha_0 = \Pr\{X^n \notin \mathcal{A}_n^*\}$.

The variational distance between P_{X^n} and $P_{\tilde{X}^n}$ satisfies

$$\begin{aligned} d(P_{X^n}, P_{\tilde{X}^n}) &= \frac{1}{2} \sum_{x \notin \mathcal{A}_n^*} |P_{X^n}(x) - P_{\tilde{X}^n}(x)| \\ &\leq \frac{1}{2} \sum_{x \notin \mathcal{A}_n^*} (P_{X^n}(x) + P_{\tilde{X}^n}(x)) \leq \alpha_0 = 1 - \Pr\{X^n \in \mathcal{A}_n^*\} \leq \delta, \end{aligned} \tag{A25}$$

where the last inequality is due to (A22). Therefore, $P_{\tilde{X}^n} \in B_\delta(X^n)$, and this implies

$$\begin{aligned} H_{[\delta]}(X^n) &\leq H(\tilde{X}^n) = F(\mathcal{A}_n^*) + \alpha_0 \log \frac{1}{\alpha_0} \\ &\leq G_{[\delta]}(X^n) + \alpha_0 \log \frac{1}{\alpha_0} + \gamma \\ &\leq G_{[\delta]}(X^n) + \frac{\log e}{e} + \gamma, \end{aligned} \tag{A26}$$

where the first inequality is due to (A23) and the last inequality is due to the inequality $x \log x \geq -\frac{\log e}{e}$ for all $x > 0$. Thus, we obtain the desired inequality: $H_{[\delta]}(\mathbf{X}) \leq G_{[\delta]}(\mathbf{X})$.

Next, we shall show $H_{[\delta]}(\mathbf{X}) \geq G_{[\delta]}(\mathbf{X})$. Assume, without loss of generality, that the elements of \mathcal{X}^n are indexed as $\mathbf{x}_1, \mathbf{x}_2, \dots \in \mathcal{X}^n$ so that

$$P_{X^n}(\mathbf{x}_i) \geq P_{X^n}(\mathbf{x}_{i+1}) \quad (\forall i = 1, 2, \dots). \tag{A27}$$

For a given $\delta \in [0, 1)$, let j^* denote the integer satisfying

$$\sum_{i=1}^{j^*-1} P_{X^n}(\mathbf{x}_i) < 1 - \delta, \quad \sum_{i=1}^{j^*} P_{X^n}(\mathbf{x}_i) \geq 1 - \delta. \tag{A28}$$

Let V_δ^n be a random variable taking values in \mathcal{X}^n whose probability distribution is given by

$$P_{V_\delta^n}(\mathbf{x}_i) = \begin{cases} P_{X^n}(\mathbf{x}_i) + \delta & \text{for } i = 1 \\ P_{X^n}(\mathbf{x}_i) & \text{for } i = 2, 3, \dots, j^* - 1 \\ P_{X^n}(\mathbf{x}_i) - \varepsilon_n & \text{for } i = j^* \\ 0 & \text{otherwise,} \end{cases} \tag{A29}$$

where $\varepsilon_n := \delta - \sum_{i \geq j^*+1} P_{X^n}(\mathbf{x}_i)$. It is easily checked that $0 \leq \varepsilon_n \leq P_{X^n}(\mathbf{x}_{j^*})$ and the probability distribution $P_{V_\delta^n}$ majorizes (for a sequence $\mathbf{u} = (u_1, u_2, \dots, u_L)$ of length L , we denote by $\tilde{\mathbf{u}} = (\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_L)$ a permuted version of \mathbf{u} satisfying $\tilde{u}_i \geq \tilde{u}_{i+1}$ for all $i = 1, 2, \dots, L - 1$, where ties are arbitrarily broken. We say $\mathbf{u} = (u_1, u_2, \dots, u_L)$ majorizes $\mathbf{v} = (v_1, v_2, \dots, v_L)$ if $\sum_{i=1}^j \tilde{u}_i \geq \sum_{i=1}^j \tilde{v}_i$ for all $j = 1, 2, \dots, L$.) any $P_{V^n} \in B_\delta(X^n)$ [39]. Since the Shannon entropy is a Schur concave function (The function $f(\mathbf{u})$ is said to be Schur concave if $f(\mathbf{u}) \leq f(\mathbf{v})$ for any pair (\mathbf{u}, \mathbf{v}) , where \mathbf{v} is majorized by \mathbf{u} .) According to [40], we immediately obtain the following lemma, which is of use to compute $H_{[\delta]}(X^n)$.

Lemma A1 (Ho and Yeung [39]).

$$H_{[\delta]}(X^n) = H(V_\delta^n) \quad (\forall \delta \in [0, 1)). \tag{A30}$$

By the definition of $G_{[\delta]}(X^n)$, we obtain

$$G_{[\delta]}(X^n) \leq \sum_{i=1}^{j^*} P_{X^n}(\mathbf{x}_i) \log \frac{1}{P_{X^n}(\mathbf{x}_i)} \tag{A31}$$

$$\leq H(V_\delta^n) + P_{X^n}(\mathbf{x}_1) \log \frac{1}{P_{X^n}(\mathbf{x}_1)} + P_{X^n}(\mathbf{x}_{j^*}) \log \frac{1}{P_{X^n}(\mathbf{x}_{j^*})} \tag{A32}$$

$$\leq H(V_\delta^n) + \frac{2 \log e}{e}, \tag{A33}$$

where the last inequality is due to $x \log x \geq -\frac{\log e}{e}$ for all $x > 0$. Thus, it follows from Lemma A1 that

$$G_{[\delta]}(\mathbf{X}) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} H(V_\delta^n) = H_{[\delta]}(\mathbf{X}), \tag{A34}$$

which is the desired inequality.

(ii) Next, we show the inequality in (40): $H_{[\delta]}(\mathbf{X}) \leq (1 - \delta)\bar{H}_{\delta-\gamma}(\mathbf{X})$. By the definition of $\bar{H}_{\delta-\gamma}(\mathbf{X})$, for any $\eta > 0$, there exists some $n_0 > 0$ such that

$$\Pr\{X^n \in T_n\} \geq 1 - \delta \quad (\forall n \geq n_0), \tag{A35}$$

where we define

$$T_n = \left\{ \mathbf{x} \in \mathcal{X}^n : \frac{1}{n} \log \frac{1}{P_{X^n}(\mathbf{x})} \leq \bar{H}_{\delta-\gamma}(\mathbf{X}) + \eta \right\}. \tag{A36}$$

Choose a sequence $\mathbf{x}_0 \in T_n$ arbitrarily. Set P_{V^n} so that

$$P_{V^n}(x) = \begin{cases} \alpha_n P_{X^n}(x) + \delta & \text{for } x = x_0 \\ \alpha_n P_{X^n}(x) & \text{for } x \neq x_0, x \in T_n \\ 0 & \text{otherwise,} \end{cases} \tag{A37}$$

where we define $\alpha_n = (1 - \delta) / \Pr\{X^n \in T_n\}$. Then, the variational distance between P_{X^n} and P_{V^n} satisfies

$$\begin{aligned} d(P_{X^n}, P_{V^n}) &\leq \frac{\delta}{2} + \frac{1}{2} \sum_{x \in T_n} P_{X^n}(x) |1 - \alpha_n| + \frac{1}{2} \sum_{x \in T_n^c} P_{X^n}(x) \\ &= \frac{\delta}{2} + \frac{1}{2} \Pr\{X^n \in T_n\} (1 - \alpha_n) + \Pr\{X^n \in T_n^c\} \\ &= \delta, \end{aligned} \tag{A38}$$

which indicates that $P_{V^n} \in B_\delta(X^n)$. The normalized Shannon entropy $\frac{1}{n}H(V^n)$ can be upper bounded as

$$\begin{aligned} \frac{1}{n}H(V^n) &= \frac{1}{n}(\alpha_n P_{X^n}(x_0) + \delta) \log \frac{1}{\alpha_n P_{X^n}(x_0) + \delta} \\ &\quad + \frac{1}{n} \sum_{x \in T_n \setminus \{x_0\}} P_{V^n}(x) \log \frac{1}{P_{V^n}(x)} \\ &\leq \frac{\delta}{n} \log \frac{1}{\delta} + \frac{1}{n} \sum_{x \in T_n} \alpha_n P_{X^n}(x) \log \frac{1}{\alpha_n P_{X^n}(x)} \\ &= \frac{\delta}{n} \log \frac{1}{\delta} + \frac{\alpha_n}{n} \sum_{x \in T_n} P_{X^n}(x) \left(\log \frac{1}{\alpha_n} + \log \frac{1}{P_{X^n}(x)} \right) \\ &\leq \frac{\delta}{n} \log \frac{1}{\delta} + \frac{(1 - \delta)}{n} \log \frac{1}{1 - \delta} + \alpha_n \sum_{x \in T_n} P_{X^n}(x) (\overline{H}_{\delta-\gamma}(\mathbf{X}) + \eta) \\ &= \frac{1}{n}h_2(\delta) + (1 - \delta) (\overline{H}_{\delta-\gamma}(\mathbf{X}) + \eta), \end{aligned} \tag{A39}$$

where we have used the fact that $\frac{1}{n} \log \frac{1}{P_{X^n}(x)} \leq \overline{H}_{\delta-\gamma}(\mathbf{X}) + \eta$ for $x \in T_n$ to obtain the second inequality, and $h_2(\delta) := -\delta \log \delta - (1 - \delta) \log(1 - \delta)$ denotes the binary entropy function. On the other hand, because $P_{V^n} \in B_\delta(X^n)$, there exists some $n_1 > 0$ such that

$$H_{[\delta]}(\mathbf{X}) \leq \frac{1}{n}H(V^n) + \eta \quad (\forall n > n_1). \tag{A40}$$

Combining (A39) and (A40) yields

$$H_{[\delta]}(\mathbf{X}) \leq (1 - \delta) \overline{H}_{\delta-\gamma}(\mathbf{X}) + 3\eta, \tag{A41}$$

and since $\eta > 0$ is arbitrarily fixed, (A41) means the desired inequality.

Appendix D. Proof of Corollary 1

We first show (44), i.e.,

$$S_f(\delta|\mathbf{X}) \stackrel{(a)}{=} \overline{H}_\delta(\mathbf{X}) \stackrel{(b)}{=} H^*(\mathbf{X}). \tag{A42}$$

Equality (a) is due to Theorem 2. To prove equality (b), we notice that

$$\underline{H}(\mathbf{X}) \leq \overline{H}_\delta(\mathbf{X}) \leq \overline{H}(\mathbf{X}) \quad (\forall \delta \in [0, 1)) \tag{A43}$$

for general source X by definition. If source X is with one-point spectrum, the left-hand side (l.h.s.) and the r.h.s. in (A43) match, i.e., $H^*(X) := \underline{H}(X) = \overline{H}(X)$, and the squeeze theorem indicates equality (b).

Next, we show (45), i.e.,

$$S_v(\delta|X) \stackrel{(c)}{=} H_{[\delta]}(X) \stackrel{(d)}{=} G_{[\delta]}(X) \stackrel{(e)}{=} (1 - \delta)H^*(X). \tag{A44}$$

Equality (d) is a direct consequence of Proposition 1. To prove equality (e), we notice a general relationship

$$(1 - \delta)\underline{H}(X) \leq G_{[\delta]}(X) \leq (1 - \delta)\overline{H}(X) \quad (\forall \delta \in [0, 1]) \tag{A45}$$

for general source X , where the first inequality is due to ([35], Theorem 4) and the second one is due to Proposition 1 (cf. Equation (42)). In view of $H^*(X) = \underline{H}(X) = \overline{H}(X)$ for source X with one-point spectrum, the squeeze theorem for (A45) indicates equality (e), i.e.,

$$G_{[\delta]}(X) = (1 - \delta)H^*(X) \quad (\forall \delta \in [0, 1]). \tag{A46}$$

The r.h.s. of (A46) is obviously right-continuous in $\delta \in [0, 1)$, and so is the l.h.s. side $G_{[\delta]}(X) = H_{[\delta]}(X)$ if source X is with one-point spectrum. The right-continuity of $H_{[\delta]}(X)$ and Theorem 3 indicate equality (c).

Appendix E. Proof of Lemma 1

We first show (99). For two general sources $X = \{X^n\}_{n=1}^\infty$ and $\tilde{X} = \{\tilde{X}^n\}_{n=1}^\infty$, the following well-known inequality (cf. ([33], Problem 3.18)) holds between the variational distance and the divergence:

$$\frac{2(d(P_{X^n}, P_{\tilde{X}^n}))^2}{\ln K} \leq D(\tilde{X}^n || X^n). \tag{A47}$$

This inequality implies that any $P_{V^n} \in B_{g(\delta)}^D(X^n)$ satisfies $P_{V^n} \in B_\delta(X^n)$. Thus, we have

$$H_{[\delta]}(X^n) \leq H_{[g(\delta)]}^D(X^n). \tag{A48}$$

Now, we shall show the rightmost equality of (100). It obviously follows from (99) that

$$\lim_{\delta \downarrow 0} H_{[\delta]}(X) \leq \lim_{\delta \downarrow 0} H_{[\delta]}^D(X). \tag{A49}$$

To show the opposite inequality, in view of (41), it suffices to show

$$\lim_{\delta \downarrow 0} G_{[\delta]}(X) \geq \lim_{\delta \downarrow 0} H_{[\delta]}^D(X). \tag{A50}$$

Fix $\delta \in (0, 1)$ and $\gamma > 0$ arbitrarily. We choose $\mathcal{A}_n \subseteq \mathcal{X}^n$ satisfying

$$G_{[\delta]}(X^n) + \gamma \geq \sum_{x \in \mathcal{A}_n} P_{X^n}(x) \log \frac{1}{P_{X^n}(x)}, \tag{A51}$$

$$\alpha_0 := \Pr\{X^n \in \mathcal{A}_n\} \geq 1 - \delta. \tag{A52}$$

We arrange a new random variable V^n subject to

$$P_{V^n}(x) = \begin{cases} \frac{P_{X^n}(x)}{\alpha_0} & \text{if } x \in \mathcal{A}_n \\ 0 & \text{otherwise.} \end{cases} \tag{A53}$$

Then, we obtain

$$D(V^n || X^n) = \sum_{x \in \mathcal{A}_n} P_{V^n}(x) \log \frac{P_{V^n}(x)}{P_{X^n}(x)} = \log \frac{1}{\alpha_0} \leq \log \frac{1}{1 - \delta}, \tag{A54}$$

and, thus, letting $h(\delta) = \log \frac{1}{1 - \delta}$, it holds that $P_{V^n} \in B_{h(\delta)}^D(X^n)$. We can expand (A51) as

$$\begin{aligned}
G_{[\delta]}(X^n) + \gamma &\geq \alpha_0 \sum_{\mathbf{x} \in \mathcal{A}_n} P_{V^n}(\mathbf{x}) \log \frac{1}{\alpha_0 P_{V^n}(\mathbf{x})} \\
&\geq \alpha_0 H(V^n) \\
&\geq (1 - \delta) H_{[h(\delta)]}^D(X^n),
\end{aligned} \tag{A55}$$

where the last inequality is due to (A52) and $P_{V^n} \in B_{h(\delta)}^D(X^n)$. Thus, as $\gamma > 0$ is arbitrary,

$$G_{[\delta]}(\mathbf{X}) \geq (1 - \delta) H_{[h(\delta)]}^D(\mathbf{X}). \tag{A56}$$

Since $\delta \in (0, 1)$ is arbitrary, in view of $\lim_{\delta \downarrow 0} h(\delta) = 0$, we obtain (A50).

References

- Elias, P. The efficient construction of an unbiased random sequence. *Ann. Math. Statist.* **1972**, *43*, 865–870. [\[CrossRef\]](#)
- Han, T.S. *Information-Spectrum Methods in Information Theory*; Springer: Berlin/Heidelberg, Germany, 2003.
- Knuth, D.E.; Yao, A.C.-C. The complexity of nonuniform random number generation. In *Algorithms and Complexity: New Directions and Recent Results*; Academic Press: New York, NY, USA, 1976; pp. 357–428.
- Han, T.S. Theorems on variable-length intrinsic randomness. *IEEE Trans. Inf. Theory* **2000**, *46*, 2108–2116.
- Vembu, S.; Verdú, S. Generating random bits from an arbitrary source: Fundamental limits. *IEEE Trans. Inf. Theory* **1995**, *41*, 1322–1332. [\[CrossRef\]](#)
- Han, T.S.; Verdú, S. Approximation theory of output statistics. *IEEE Trans. Inf. Theory* **1993**, *39*, 752–771. [\[CrossRef\]](#)
- Steinberg, Y.; Verdú, S. Simulation of random processes and rate-distortion theory. *IEEE Trans. Inf. Theory* **1996**, *42*, 63–86. [\[CrossRef\]](#)
- Renner, R.; Wolf, S. Smooth Rényi entropy and applications. In Proceedings of the International Symposium on Information Theory, ISIT 2004, Chicago, IL, USA, 27 June–2 July 2004; p. 232.
- Hou, J.; Kramer, G. Informational divergence approximations to product distributions. In Proceedings of the 13th Canadian Workshop on Information Theory, Toronto, ON, Canada, 18–21 June 2013.
- Yagi, H. Channel resolvability theorems for general sources and channels. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017.
- Ahlsweide, R.; Dueck, G. Identification via channels. *IEEE Trans. Inf. Theory* **1989**, *35*, 15–29. [\[CrossRef\]](#)
- Oohama, Y. Converse coding theorems for identification via channels. *IEEE Trans. Inf. Theory* **2013**, *59*, 744–759. [\[CrossRef\]](#)
- Steinberg, Y. New converses in the theory of identification via channels. *IEEE Trans. Inf. Theory* **1998**, *44*, 984–998. [\[CrossRef\]](#)
- Sreekumar, S.; Cohen, A.; Gündüz, D. Privacy-aware distributed hypothesis testing. *Entropy* **2020**, *22*, 665. [\[CrossRef\]](#)
- Zheng, F.; Xiao, Z.; Zhou, S.; Wang, J.; Huang, L. Message authentication over noisy channels. *Entropy* **2015**, *17*, 368–383. [\[CrossRef\]](#)
- Lin, P.-H.; Janda, C.R.; Jorswieck, E.A.; Schaefer, R.F. Stealthy secret key generation. *Entropy* **2020**, *22*, 679. [\[CrossRef\]](#)
- Hayashi, M. Nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel. *IEEE Trans. Inf. Theory* **2006**, *52*, 1562–1575. [\[CrossRef\]](#)
- Bloch, M.R.; Laneman, J.N. Strong secrecy from channel resolvability. *IEEE Trans. Inf. Theory* **2013**, *59*, 8077–8098. [\[CrossRef\]](#)
- Frèche, G.; Bloch, M.R.; Barret, M. Polar codes for covert communications over asynchronous discrete memoryless channels. *Entropy* **2022**, *22*, 679.
- Strassen, V. Asymptotische abschätzungen in Shannon’s informationstheorie. In Proceedings of the Transactions of the Third Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Prague, Czech Republic, 5–13 June 1962; pp. 689–723.
- Han, T.S. Second-order information theory and hypothesis testing. In Proceedings of the 2015 IEEE Information Theory Workshop, Jeju, Republic of Korea, 11–15 October 2015.
- Han, T.S.; Nomura, R. First- and second-order hypothesis testing for mixed memoryless sources. *Entropy* **2018**, *20*, 174. [\[CrossRef\]](#)
- Hayashi, M. Second-order asymptotics in fixed-length source coding and intrinsic randomness. *IEEE Trans. Inf. Theory* **2008**, *54*, 4619–4637. [\[CrossRef\]](#)
- Hayashi, M. Information spectrum approach to second-order coding rate in channel coding. *IEEE Trans. Inf. Theory* **2009**, *55*, 4947–4966. [\[CrossRef\]](#)
- Kostina, V.; Polyanskiy, Y.; Verdú, S. Variable-length compression allowing errors. *IEEE Trans. Inf. Theory* **2015**, *61*, 4316–4330. [\[CrossRef\]](#)
- Nomura, R. Source resolvability and intrinsic randomness: Two random number generation problems with respect to a subclass of f -divergences. *IEEE Trans. Inf. Theory* **2020**, *66*, 7588–7601. [\[CrossRef\]](#)
- Nomura, R.; Han, T.S. Second-order resolvability, intrinsic randomness, and fixed-length source coding for mixed sources: Information spectrum approach. *IEEE Trans. Inf. Theory* **2013**, *59*, 1–16. [\[CrossRef\]](#)

28. Polyanskiy, Y.; Poor, H.V.; Verdú, S. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theory* **2010**, *59*, 2307–2358. [[CrossRef](#)]
29. Watanabe, S.; Hayashi, M. Strong converse and second-order asymptotics of channel resolvability. In Proceedings of the 2014 IEEE International Symposium on Information Theory, Honolulu, HI, USA, 29 June–4 July 2014.
30. Yagi, H.; Han, T.S.; Nomura, R. First- and second-order coding theorems for mixed memoryless channels with general mixture. *IEEE Trans. Inf. Theory* **2016**, *62*, 4395–4412. [[CrossRef](#)]
31. Yagi, H.; Nomura, R. Variable-length coding with cost allowing non-vanishing error probability. *IEICE Trans. Fundam.* **2017**, *E100-A*, 1683–1692. [[CrossRef](#)]
32. Uyematsu, R. A new unified method for fixed-length source coding problems of general sources. *IEICE Trans. Fundam.* **2010**, *E93-A*, 1868–1877. [[CrossRef](#)]
33. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2011.
34. Tomita, T.; Uyematsu, T.; Matsumoto, R. Bounds on a variable-to-fixed type δ -resolvability problem. In Proceedings of the Poster Presentation at 2016 Symposium on Information Theory and its Applications, Gifu, Japan, 30 October–2 November 2016.
35. Koga, H.; Yamamoto, H. Asymptotic properties on codeword length of an optimal FV code for general sources. *IEEE Trans. Inf. Theory* **2005**, *51*, 1546–1555. [[CrossRef](#)]
36. Han, T.S. Weak variable-length source coding theorem. *IEEE Trans. Inf. Theory* **2000**, *46*, 1217–1226.
37. Kuzuoka, S.; Watanabe, S. An information-spectrum approach to weak variable-length source coding with side information. *IEEE Trans. Inf. Theory* **2015**, *61*, 3559–3573. [[CrossRef](#)]
38. Uchida, O.; Han, T.S. The optimal overflow and underflow probabilities of variable-length coding for general source. *IEICE Trans. Fundam.* **2001**, *E84-A*, 2457–2465.
39. Ho, S.-W.; Yeung, R.W. The interplay between entropy and variational distance. *IEEE Trans. Inf. Theory* **2010**, *56*, 5906–5929. [[CrossRef](#)]
40. Marshall, A.W.; Olkin, I.; Arnold, B.C. *Inequalities: Theory of Majorization and Its Applications*, 2nd ed.; Springer: New York, NY, USA, 2011.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.