PAPER Special Section on Information Theory and Its Applications

A Fundamental Limit of Variable-Length Compression with Worst-Case Criteria in Terms of Side Information*

Sho HIGUCHI[†a)], ***Nonmember and*** **Yuta SAKAI**[††b)], ***Member***

SUMMARY In this study, we consider the data compression with side information available at both the encoder and the decoder. The information source is assigned to a variable-length code that does not have to satisfy the prefix-free constraints. We define several classes of codes whose codeword lengths and error probabilities satisfy worse-case criteria in terms of side-information. As a main result, we investigate the exact first-order asymptotics with second-order bounds scaled as $\Theta(\sqrt{n})$ as blocklength n increases under the regime of nonvanishing error probabilities. To get this result, we also derive its one-shot bounds by employing the cutoff operation.

key words: *variable-length compression, one-shot formula, conditional source coding, cutoff operation, second-order bounds*

1. Introduction

In this article, we consider the variable-length compression problem in noiseless communication channels without prefix-free constraints. When the error probability is zero, the problem becomes a naive zero-error version of variable-length compression. Wyner [1] derived an upper bound the average codeword length for the zero-error case; later Alon and Orlitsky [2] derived a lower bound on that as follows:

$$H(X) - \log(H(X)+1) - \log e \le L_X^*(0) \le H(X), \quad (1)$$

where log stands for the base-2 logarithm, $L_X^*(0)$ is the minimum average codeword length for the source X in the zero-error case, and $H(X)$ is the Shannon entropy of X.

Classically, this problem is considered without errors, but errors often occur in the practical communication channel. Therefore, it is worthwhile to consider the variable-length compression problem in the presence of errors. For the error allowing case, Kostina, Polyanskiy, and Verdú [3] derived the asymptotic analysis as shown in the following:

$$L_{X^n}^*(\varepsilon) = n(1-\varepsilon)H(X) - \sqrt{nV(X)} f_{\mathrm{G}}(\varepsilon) + O(\log n) \quad (2)$$

as $n \to \infty$, for every fixed $0 \le \varepsilon \le 1$. Here $L_{X^n}^*(\varepsilon)$ is the minimum average codeword length for the source $X^n = (X_1, \ldots, X_n)$ in which error probabilities are tolerated to be positive but at most ε, the quantity $V(X)$ is the variance of the information density $-\log P_X(X)$, and f_{G} is defined as

$$f_G(s) \triangleq \begin{cases} \phi(\Phi^{-1}(s)) & \text{if } 0 < s < 1, \\ 0 & \text{if } s = 0 \text{ or } s = 1, \end{cases} \quad (3)$$

$$\phi(t) \triangleq \frac{1}{\sqrt{2\pi}} e^{-t^2/2}, \quad (4)$$

$$\Phi(u) \triangleq \int_{-\infty}^{u} \phi(t) \mathrm{d}t. \quad (5)$$

In [3], they first extended the one-shot bounds shown in (1) from zero-error to allowing error settings by introducing the *cutoff operation:* a kind of truncating a real-valued random variable (r.v.). To obtain the second-order asymptotics shown in (2), they examined an asymptotic analysis of the expectation of the cutoff operation for information densities. Later, Sakai and Tan [4] called such information quantities the *cutoff entropies*, and the one-shot bounds in [3, Theorem 2] is a first instance of the operational characterization of cutoff entropies. Operational characterizations of cutoff entropies (including their variants) were successfully studied in several other information-theoretic problems in [4], [5]. In these studies [3]–[5], it is worth mentioning that operational characterizations and asymptotic analyses of cutoff entropies can be *independently* examined, and their combination readily provides the second-order asymptotics of coding problems**.

So far, we have introduced previous studies of the variable-length compression mainly in the absence of side information. Henceforth, we consider a communication channel in which side information Y exists in addition to the source X. Slepian–Wolf coding [7] is a code such that for two correlated information sources X and Y, encoding is done independently of each other and decoding is done such that the error probability of each X and Y is as small as possible. This is an important distributed coding problem from both practical and theoretical viewpoints. Asymptotic analysis of *variable-length* Slepian–Wolf coding under vanishing error probabilities conditions has been studied [8]. On the other hand, to the best of our knowledge, its asymptotic analysis

Manuscript received February 9, 2023.
Manuscript revised May 22, 2023.
Manuscript publicized July 3, 2023.

[†]The author is with the University of Hyogo, Himeji-shi, 671-2280 Japan.
[††]The author is with Shimane University, Matsue-shi, 690-8504 Japan.

*Part of this research was presented at the 45th Symposium on Information Theory and Its Applications (SITA2022). This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grants 21K21291 and 23K16839.

a) E-mail: ei22n019@guh.u-hyogo.ac.jp
b) E-mail: yuta.sakai@mat.shimane-u.ac.jp

DOI: 10.1587/transfun.2023TAP0003

**Unfortunately, cutoff entropies are not useful to derive higher-order asymptotics of some coding problems. Actually, Sakai, Yavas, and Tan [6] refined the remainder term in (2) from $O(\log n)$ to $-((1-\varepsilon)/2)\log n + O(1)$ not via analysis of cutoff entropies but via Cramér-type strong large deviations of average codeword lengths.

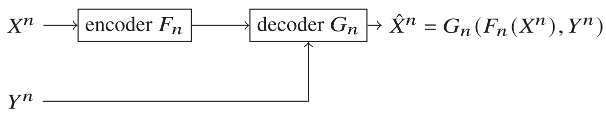

Fig. 1 Slepian–Wolf coding.

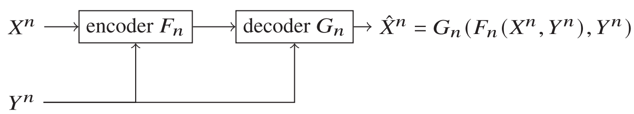

Fig. 2 conditional source coding.

under a fixed error probability conditions remains an open problem. Whereas Slepian–Wolf coding uses side information Y only for decoding, conditional source coding (see, e.g., [9]) can use side information Y for both encoding and decoding. Schemes of Slepian–Wolf coding and conditional source coding are shown in Figs. 1 and 2, respectively. It is clear that the latter is more compression efficient in general. In other words, a converse result of conditional source coding would immediately yield a hint to the fundamental limits of Slepian–Wolf coding. Since the analysis of distributed coding problems is generally difficult, some motivations of our study considering variable-length conditional source coding are to prevent such a difficulty and to establish insights for tackling Slepian–Wolf coding with variable-rate.

In [4], variable-length conditional source coding was studied under two formalisms: the average and maximum error probabilities. One-shot bounds of the fundamental limits of these coding problems were then established by different types of cutoff entropies. Especially, the latter formalism was analyzed by the *conditional* cutoff operation, while the former formalism was analyzed by the same cutoff operation as in [3]. In this study, we introduce another performance criterion to variable-length conditional source coding. Under our setting, we investigate operational characterizations and asymptotic expansions of a cutoff-operation-based entropy.

The rest of this paper is organized as follows: Section 2 introduces the notations and definitions treated in this article. Section 3 shows our main result deriving a one-shot formula and its second-order asymptotic analysis with remainder term $\Theta(\sqrt{n})$. Section 4 concludes this study.

2. Preliminaries

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathbb{E}[Z \mid W]$ be the conditional expectation of a real-valued r.v. Z given the σ-algebra generated by another r.v. W, and $\mathbb{P}\{E \mid W\} = \mathbb{E}[\mathbf{1}_E \mid W]$ the conditional probability of an event $E \in \mathcal{F}$. In addition, we introduce some conditional information quantities.

Definition 1 ([4])**.** *Let X be a discrete r.v. and Y an arbitrary r.v. Then*

$$\iota(X \mid Y) = \iota_{X|Y}(X \mid Y) \triangleq \log \frac{1}{P_{X|Y}(X \mid Y)}, \tag{6}$$

where $P_{X|Y}(X \mid Y)$ is the conditional probability of X given Y. We define information measures of X given Y as follows:

$$\mathcal{H}(X \mid Y) \triangleq \mathbb{E}[\iota(X \mid Y) \mid Y], \tag{7}$$

$$\mathcal{V}(X \mid Y) \triangleq \mathbb{E}[(\iota(X \mid Y) - \mathcal{H}(X \mid Y))^2 \mid Y], \tag{8}$$

$$\mathcal{T}(X \mid Y) \triangleq \mathbb{E}[|\iota(X \mid Y) - \mathcal{H}(X \mid Y)|^3 \mid Y], \tag{9}$$

$$H(X \mid Y) \triangleq \mathbb{E}[\mathcal{H}(X \mid Y)], \tag{10}$$

$$V_c(X \mid Y) \triangleq \mathbb{E}[\mathcal{V}(X \mid Y)], \tag{11}$$

$$V_u(X \mid Y) \triangleq \mathbb{E}[(\iota(X \mid Y) - H(X \mid Y))^2], \tag{12}$$

$$V_{\sup}(X \mid Y) \triangleq \operatorname{ess\,sup} \mathcal{V}(X \mid Y), \tag{13}$$

$$V_{\inf}(X \mid Y) \triangleq \operatorname{ess\,inf} \mathcal{V}(X \mid Y), \tag{14}$$

where the essential supremum *of a r.v. Z is defined as*

$$\operatorname{ess\,sup} Z \triangleq \inf\{z \mid \mathbb{P}\{Z > z\} = 0\}, \tag{15}$$

and the essential infimum *is* $\operatorname{ess\,inf} Z = -\operatorname{ess\,sup}(-Z)$.

Proposition 1 (essential supremum inequality relations)**.** *For two real-valued r.v.'s X and Y, it holds that*

$$\operatorname{ess\,sup} X + \operatorname{ess\,sup} Y \geq \operatorname{ess\,sup}\,(X + Y), \tag{16}$$

$$\operatorname{ess\,sup}\,(X - Y) \geq \operatorname{ess\,sup} X - \operatorname{ess\,sup} Y. \tag{17}$$

These inequalities are quite elementary, but the proof is provided to make the paper self-contained.

Proof of Proposition 1: See Appendix A. ■

Kostina, Polyanskiy, and Verdú [3] introduced the *cutoff operation* $\langle\cdot\rangle_\varepsilon$ as follows:

Definition 2 (uncoditional cutoff operation [3])**.** *Given a real $0 \leq \varepsilon \leq 1$ and a real-valued r.v. A, define*

$$\langle A \rangle_\varepsilon \triangleq \begin{cases} A & A < \eta, \\ \eta & A = \eta \ (w.p.\ 1-\alpha), \\ 0 & A = \eta \ (w.p.\ \alpha), \\ 0 & \text{otherwise}, \end{cases} \tag{18}$$

where $\eta \in \mathbb{R}$ and $0 \leq \alpha < 1$ are chosen so that

$$\mathbb{P}\{A > \eta\} + \alpha \mathbb{P}\{A = \eta\} = \varepsilon. \tag{19}$$

To examine our problem of variable-length compressions in the presence of side-information, we now introduce the *conditional cutoff operation* as follows:

Definition 3 (conditional cutoff operation [4])**.** *Given a real $0 \leq \varepsilon \leq 1$, a real-valued r.v. Z, and an arbitrary r.v. W,*

$$\langle Z \mid W \rangle_\varepsilon \triangleq \begin{cases} Z & Z < \eta_W, \\ B_W Z & Z = \eta_W, \\ 0 & Z > \eta_W, \end{cases} \tag{20}$$

where B_W denotes a Bernoulli r.v. in which the conditional independence $B_W \perp Z \mid W$ holds and

$$\mathbb{P}\{B_W = 0 \mid W\} = \beta_W \qquad \text{(a.s.)}, \tag{21}$$

where $\eta_W \in \mathbb{R}$ and $0 \leq \beta_W < 1$ are $\sigma(W)$-measurable r.v.'s *chosen so that*

$$\mathbb{P}\{Z > \eta_W \mid W\} + \beta_W \mathbb{P}\{Z = \eta_W \mid W\} = \varepsilon \qquad \text{(a.s.)}. \tag{22}$$

The following proposition, presented in [4, Equation 31], will be used in the subsequent analysis.

Proposition 2. *Let Z be a nonnegative-valued r.v. and W an arbitrary r.v. It holds that*

$$\mathbb{E}[\langle Z \mid W\rangle_\varepsilon \mid W] = (1-\varepsilon)\mathbb{E}[Z \mid W] - \int_{\eta_W}^{\infty} \mathbb{P}\{Z > t \mid W\}\mathrm{d}t - \varepsilon(\eta_W - \mathbb{E}[Z \mid W]) \qquad \text{(a.s.)}. \tag{23}$$

Sakai and Tan [4] considered variable-length conditional source coding under two error formalisms. Let n be a positive integer and (X^n, Y^n) a sequence of independent copies of (X, Y), where $X^n = (X_1, \ldots, X_n)$ and $Y^n = (Y_1, \ldots, Y_n)$. Denote by $\{0,1\}^*$ the set of finite-length binary strings containing the empty string $\emptyset$. Given reals $L \geq 0$ and $0 \leq \varepsilon \leq 1$, two types of variable-length (n, L, ε)-codes of the source X in the presence of side information Y are defined as follows.

Definition 4 (Maximum Error Criterion [4])**.** *The length of a string $F_n(X^n, Y^n) \in \{0,1\}^*$ is written by $\ell(F_n(X^n, Y^n))$. A variable-length $(n, L, \varepsilon)_{\mathrm{avg,max}}$-code for source X with side information Y is defined as*

$$\mathbb{E}[\ell(F_n(X^n, Y^n))] \leq L, \tag{24}$$
$$\mathbb{P}\{X^n \neq G_n(F_n(X^n, Y^n), Y^n) | Y^n\} \leq \varepsilon \qquad \text{(a.s.)}, \tag{25}$$

The fundamental limit of $(n, L, \varepsilon)_{\mathrm{avg,max}}$-code is defined as

$$L^*_{\mathrm{avg,max}}(n, \varepsilon, X, Y) \triangleq \inf\{L : \text{an } (n, L, \varepsilon)_{\mathrm{avg,max}}\text{-code exist}\}. \tag{26}$$

Definition 5 (Average Error Criterion [4])**.** *A variable-length $(n, L, \varepsilon)_{\mathrm{avg,avg}}$-code for source X with side information Y is defined to satisfy*

$$\mathbb{E}[\ell(F_n(X^n, Y^n))] \leq L, \tag{27}$$
$$\mathbb{P}\{X^n \neq G_n(F_n(X^n, Y^n), Y^n)\} \leq \varepsilon, \tag{28}$$

The fundamental limit of $(n, L, \varepsilon)_{\mathrm{avg,avg}}$-code is defined as

$$L^*_{\mathrm{avg,avg}}(n, \varepsilon, X, Y) \triangleq \inf\{L : \text{an } (n, L, \varepsilon)_{\mathrm{avg,avg}}\text{-code exists}\}. \tag{29}$$

Under some mild conditions, they [4] investigated asymptotic analyses of these fundamental limits and derived

$$L^*_{\mathrm{avg,avg}}(n, \varepsilon, X, Y) = n(1-\varepsilon)H(X|Y) - \sqrt{nV_u(X|Y)} f_G(\varepsilon) + O(\log n), \tag{30}$$

$$L^*_{\mathrm{avg,max}}(n, \varepsilon, X, Y) = n(1-\varepsilon)H(X|Y) - \sqrt{nV_c(X|Y)} f_G(\varepsilon) + O(\log n). \tag{31}$$

Table 1 Code subscript naming table.

		Error probability	
		avg	max
codeword length	avg	avg,avg	avg,max
	max	max,avg	max,max

These asymptotic equations are the same in the first-order term. By the law of total variance, we see that

$$V_u(X \mid Y) = \mathbb{E}[(\mathcal{H}(X \mid Y) - H(X \mid Y))^2] + V_c(X \mid Y). \tag{32}$$

implying that $V_u(X \mid Y) \geq V_c(X \mid Y)$. Namely, $L^*_{\mathrm{avg,avg}}$ is not greater than $L^*_{\mathrm{avg,max}}$ in the $\sqrt{n}$-scale.

3. One-Shot and Second-Order Bounds

3.1 Statement of Main Result

Definition 6 (Our Proposed Criterion)**.** *A variable-length $(n, L, \varepsilon)_{\mathrm{max,max}}$-code for source X with side information Y is defined to satisfy*

$$\mathbb{E}[\ell(F_n(X^n, Y^n)) \mid Y^n] \leq L \qquad \text{(a.s.)}, \tag{33}$$
$$\mathbb{P}\{X^n \neq G_n(F_n(X^n, Y^n), Y^n) | Y^n\} \leq \varepsilon \qquad \text{(a.s.)}. \tag{34}$$

Table 1 shows the subscript correspondence to each definition of codeword length and error probability. The subscript "max" corresponds to (33) and (34) meaning the worst case with respect to Y; whereas "avg" corresponds to (27) and (28) meaning that the criteria are averaged over Y in the sense that the law of total expectation, i.e.,

$$\mathbb{E}[\ell(F_n(X^n, Y^n))] = \mathbb{E}[\mathbb{E}[\ell(F_n(X^n, Y^n)) \mid Y^n]] \leq \operatorname{ess\,sup} \mathbb{E}[\ell(F_n(X^n, Y^n)) \mid Y^n] \tag{35}$$

and

$$\mathbb{P}\{X^n \neq G_n(F_n(X^n, Y^n), Y^n)\} = \mathbb{E}[\mathbb{P}\{X^n \neq G_n(F_n(X^n, Y^n), Y^n) | Y^n\}] \leq \operatorname{ess\,sup} \mathbb{P}\{X^n \neq G_n(F_n(X^n, Y^n), Y^n) | Y^n\}. \tag{36}$$

The fundamental limit of $(n, L, \varepsilon)_{\mathrm{max,max}}$ code is

$$L^*_{\mathrm{max,max}}(n, \varepsilon, X, Y) \triangleq \inf\{L : \textit{an } (n, L, \varepsilon)_{\mathrm{max,max}} \text{ code exists}\}. \tag{37}$$

As a special case, it should be mentioned that $L^*_{\mathrm{max,max}}(1, \varepsilon, X, Y) = 0$, provided that the tolerated probability of error ε is sufficiently large so that $\varepsilon \geq 1 - \max_x P_{X|Y}(x \mid Y)$ almost surely. This situation means that all source symbols are encoded to the empty string $\emptyset$ and the decoder always produces a most likely source symbol $\arg\max_x P_{X|Y}(x \mid Y)$. In other words, all information of the source X is removed except for the side-information Y, and one just executes a maximum a posteriori (MAP) estimator of X given Y only. On the other hand, in asymptotic analysis under the regime of nonvanishing error probabilities, the fundamental limit $L^*_{\mathrm{max,max}}(n, \varepsilon, X, Y)$ with fixed $0 \le \varepsilon < 1$ must be strictly positive for every nondegenerate source distribution of (X, Y) and for sufficiently large n, because $\max_{\boldsymbol{x}} P_{X^n|Y^n}(\boldsymbol{x} \mid Y^n)$ vanishes with positive probability as blocklength n increases.

Theorem 1. *Suppose the following two hypotheses hold:*
(a) *$\mathcal{V}(X \mid Y)$ is bounded away from zero almost surely,*
(b) *$\mathcal{T}(X \mid Y)$ is bounded away from infinity almost surely.*
Given $0 \le \varepsilon \le 1$, it holds that

$$L^*_{\mathrm{max,max}}(n, \varepsilon, X, Y) = n(1-\varepsilon)\operatorname{ess\,sup} \mathcal{H}(X \mid Y) + \theta_n, \tag{38}$$

where the remainder term θ_n is asymptotically bounded as

$$\begin{aligned} &-\sqrt{n\, V_{\sup}(X \mid Y)}\, f_{\mathrm{G}}(\varepsilon) + O(\log n) \\ &\quad \le \theta_n \le -\sqrt{n\, V_{\inf}(X \mid Y)}\, f_{\mathrm{G}}(\varepsilon) + O(1). \end{aligned} \tag{39}$$

Comparing the first terms of (30), (31) and (38), we can see that $L^*_{\mathrm{max,max}}$ is greater than $L^*_{\mathrm{avg,avg}}$ and $L^*_{\mathrm{avg,max}}$ in the n-scale. Equation (39) tells us that the remainder term θ_n in the right-hand side of (38) is roughly $\Theta(\sqrt{n})$.

3.2 Proof of Theorem 1

To prove Theorem 1, we show the following three lemmas. The first one derived a one-shot formula of the fundamental limit with $n = 1$.

Lemma 1. *Let $L^*_{\mathrm{max,max}}(\varepsilon, X, Y) \triangleq L^*_{\mathrm{max,max}}(1, \varepsilon, X, Y)$. Given $0 \le \varepsilon \le 1$, it holds that*

$$L^*_{\mathrm{max,max}}(\varepsilon, X, Y) = \operatorname{ess\,sup} \mathbb{E}[\langle \lfloor \log \varsigma_Y^{-1}(X) \rfloor \mid Y \rangle_\varepsilon \mid Y], \tag{40}$$

where ς_Y is a $\sigma(Y)$-measurable random permutation on $\mathcal{X} \triangleq \{1, 2, \dots\}$ satisfying

$$P_{X|Y}(\varsigma_Y(1) \mid Y) \ge P_{X|Y}(\varsigma_Y(2) \mid Y) \ge \cdots \qquad \text{(a.s.)}. \tag{41}$$

Namely, the permutation ς_Y rearranges the probability masses in $P_{X|Y}(\cdot \mid Y)$ in non-increasing order.

Proof: Lemma 1 can be proved in a similar way to the one-shot formula under the criterion of Definition 4. Hence, we give a proof sketch based on [4, Appendix C]. Consider a pair (F, G) of encoder and decoder that fulfills

$$\mathbb{E}[\ell(F(X, Y)) \mid Y] \le L \qquad \text{(a.s.)}, \tag{42}$$

$$\mathbb{P}\{X \neq G(F(X, Y), Y) \mid Y\} \le \varepsilon \qquad \text{(a.s.)}. \tag{43}$$

By the same way as [4, Equations (114) and (115)], we can construct a better variable-length stochastic code (F_1, g_1) than an arbitrarily given (F, G). In fact, as shown in [4, Equations (116) and (119)], the average codeword length and the maximum error probability of (F_1, g_1) are not longer than that of (F, G). By the majorization relation as in [4, Equations (120)–(130)], we can bound the average codeword length from below via the conditional cutoff operation $\langle \cdot \mid \cdot \rangle_\varepsilon$ as

$$\mathbb{E}[\ell(F_1(X, Y)) \mid Y] \ge \mathbb{E}[\langle \lfloor \log \varsigma_Y^{-1}(X) \rfloor \mid Y \rangle_\varepsilon \mid Y], \tag{44}$$

proving

$$L \ge \mathbb{E}[\langle \lfloor \log \varsigma_Y^{-1}(X) \rfloor \mid Y \rangle_\varepsilon \mid Y], \tag{45}$$

which corresponds to the converse bound of Lemma 1.

We finally show the existence of a $(1, L, \varepsilon)_{\mathrm{max,max}}$-code achieving (45) with equality. Define the $\sigma(Y)$-measurable r.v.'s κ_Y and γ_Y as follows

$$\kappa_Y \triangleq \sup\left\{ k \ge 0 \;\middle|\; \sum_{x=1}^{k} P_{X|Y}(\varsigma_Y(x) \mid Y) \le 1 - \varepsilon \right\}, \tag{46}$$

$$\gamma_Y \triangleq 1 - \varepsilon - \sum_{x=1}^{\kappa_Y} P_{X|Y}(\varsigma_Y(x) \mid Y). \tag{47}$$

In addition, we define a code $(F^*_{\sup}, g^*)$ as

$$F^*_{\sup}(x, Y) \triangleq \begin{cases} b_{\varsigma_Y^{-1}(x)} & \text{if } 1 \le \varsigma_Y^{-1}(x) \le \kappa_Y, \\ B_{\sup} & \text{if } \varsigma_Y^{-1}(x) = \kappa_Y + 1, \\ \varnothing & \text{if } \kappa_Y + 1 < \varsigma_Y^{-1}(x) < \infty, \end{cases} \tag{48}$$

$$g^*(b, Y) \triangleq x \text{ if } b = b_{\varsigma_Y^{-1}(x)} \text{ for some } x \in \mathcal{X}, \tag{49}$$

where $B_{\sup}$ is a $\{0,1\}^*$-valued r.v. conditionally independent of X given Y, and[†]

$$\begin{aligned} \mathbb{P}\{B_{\sup} = \varnothing \mid Y\} &= 1 - \mathbb{P}\{B_{\sup} = b_{\kappa_Y + 1} \mid Y\} \\ &= 1 - \frac{\gamma_Y}{P_{X|Y}(\varsigma_Y(\kappa_Y + 1) \mid Y)} \end{aligned} \tag{50}$$

a.s. In [4, Equations (136) and (137)], it was shown that $(F^*_{\sup}, g^*)$ is a $(1, L, \varepsilon)_{\mathrm{avg,max}}$-code, and a similar calculations readily proves that it is also a $(1, L, \varepsilon)_{\mathrm{max,max}}$-code. This completes the proof of Lemma 1. ■

Lemma 2. *For every $0 \le \varepsilon \le 1$, it holds that*

$$\begin{aligned} &\operatorname{ess\,sup} \mathbb{E}[\langle \iota(X \mid Y) \mid Y \rangle_\varepsilon \mid Y] \\ &\qquad - \operatorname{ess\,sup} \log(1 + \mathcal{H}(X \mid Y)) - \log e \\ &\quad \le L^*_{\mathrm{max,max}}(\varepsilon, X, Y) \\ &\quad \le \operatorname{ess\,sup} \mathbb{E}[\langle \iota(X \mid Y) \mid Y \rangle_\varepsilon \mid Y] \end{aligned} \tag{51}$$

Proof: The following proposition, proved in Appendix B, is given for proving Lemma 2.

Proposition 3. *For a nonnegative-valued r.v. Z and an arbirary r.v. W, it holds that*

[†]In [4, Equation (135)], the r.v. $B_{\sup}$ was wrongly defined. We fixed this issue in (50).

$$\mathbb{E}[\langle Z \mid W\rangle_\varepsilon \mid W] = \min_{\epsilon:\mathbb{E}[\epsilon(Z,W)|W]\le\varepsilon(\text{a.s.})} \mathbb{E}[(1-\epsilon(Z,W))Z \mid W], \tag{52}$$

where the minimization in (52) *is taken over the measurable maps* $\epsilon : [0,\infty)\times\mathcal{W} \to [0,1]$ *satisfying* $\mathbb{E}[\epsilon(Z,W) \mid W] \le \varepsilon$ *almost surely.*

For each integer $k \ge 1$, define C_k as follows

$$C_k \triangleq \left\{\lfloor \log x\rfloor \le \log \frac{1}{P_{X|Y}(\varsigma_Y(x) \mid Y)} \text{ for all } 1 \le x \le k\right\}, \tag{53}$$

Since $\mathbb{P}(C_k) = 1$ and $\{C_k\}_k$ is a decreasing sequence of events, we observe that

$$\mathbb{P}\left\{\lfloor \log \varsigma_Y^{-1}(x)\rfloor \le \log \frac{1}{P_{X|Y}(x \mid Y)} \text{ for all } x \in \mathcal{X}\right\} = 1. \tag{54}$$

From (54), we observe that

$$\mathbb{P}\left\{\log \frac{1}{P_{X|Y}(X \mid Y)} \le t \,\middle|\, Y\right\} \le \mathbb{P}\left\{\lfloor \log \varsigma_Y^{-1}(X)\rfloor \le t \mid Y\right\} \tag{55}$$

a.s., for all $t > 0$. We know that the following equation holds for the two nonnegative-valued r.v.'s Z_1 and Z_2:

$$\mathbb{P}\{Z_1 \le t \mid W\} \le \mathbb{P}\{Z_2 \le t \mid W\} \text{ (a.s. } \forall t > 0) \\ \Rightarrow \mathbb{E}[\langle Z_1 \mid W\rangle_\varepsilon \mid W] \ge \mathbb{E}[\langle Z_2 \mid W\rangle_\varepsilon \mid W] \text{ (a.s.)}. \tag{56}$$

It follows from (56) that

$$\operatorname{ess\,sup} \mathbb{E}[\langle\lfloor \log \varsigma_Y^{-1}(X)\rfloor \mid Y\rangle_\varepsilon \mid Y] \le \operatorname{ess\,sup} \mathbb{E}[\langle \iota(X \mid Y) \mid Y\rangle_\varepsilon \mid Y]. \tag{57}$$

Combined with Lemma 1 and (57), one derives that the following equation

$$L^*_{\max,\max}(\varepsilon, X, Y) \le \operatorname{ess\,sup} \mathbb{E}[\langle \iota(X \mid Y) \mid Y\rangle_\varepsilon \mid Y]. \tag{58}$$

We derive a left-hand inequality of (51) as follows

$$\begin{aligned}
&\operatorname{ess\,sup} \mathbb{E}[\langle\lfloor \log \varsigma_Y^{-1}(X)\rfloor \mid Y\rangle_\varepsilon \mid Y] \\
&\overset{\text{(a)}}{=} \operatorname{ess\,sup} \min_{\epsilon:\mathbb{E}[\epsilon(\lfloor \log \varsigma_Y^{-1}(X)\rfloor, Y)|Y]\le\varepsilon(\text{a.s.})} \\
&\qquad \mathbb{E}[(1-\epsilon(\lfloor \log \varsigma_Y^{-1}(X)\rfloor, Y))\lfloor \log \varsigma_Y^{-1}(X)\rfloor \mid Y] \\
&\overset{\text{(b)}}{\ge} \operatorname{ess\,sup} (\mathcal{H}(X \mid Y) - \log(\mathcal{H}(X \mid Y)+1) - \log e \\
&\qquad - \max_{\epsilon:\mathbb{E}[\epsilon(\lfloor \log \varsigma_Y^{-1}(X)\rfloor, Y)|Y]\le\varepsilon(\text{a.s.})} \\
&\qquad \mathbb{E}[\epsilon(\lfloor \log \varsigma_Y^{-1}(X)\rfloor, Y)\lfloor \log \varsigma_Y^{-1}(X)\rfloor \mid Y]) \\
&\overset{\text{(c)}}{=} \operatorname{ess\,sup} (\mathbb{E}[\langle \iota(X \mid Y) \mid Y\rangle_\varepsilon \mid Y] \\
&\qquad - \log(\mathcal{H}(X \mid Y)+1)) - \log e \\
&\overset{\text{(d)}}{\ge} \operatorname{ess\,sup} \mathbb{E}[\langle \iota(X \mid Y) \mid Y\rangle_\varepsilon \mid Y] \\
&\qquad - \operatorname{ess\,sup} \log(1+\mathcal{H}(X \mid Y)) - \log e,
\end{aligned} \tag{59}$$

where

- (a) and (c) follow from (52),
- (b) follows from the following equation given in [2],

$$\mathbb{E}[\lfloor \log \varsigma_Y^{-1}(X)\rfloor \mid Y] \ge \mathcal{H}(X \mid Y) - \log(\mathcal{H}(X \mid Y)+1) - \log e, \tag{60}$$

- (d) follows from (17) of Proposition 1.

This completes the proof of Lemma 2. ■

Since Hypothesis (b) in Theorem 1 holds, we know that $H(X^n \mid Y^n)$ is finite and $H(X^n \mid Y^n) = nH(X \mid Y)$ holds. From this, we can see from Lemma 2 that the following equation holds:

$$L^*_{\max,\max}(n, \varepsilon, X, Y) = \operatorname{ess\,sup} \mathbb{E}[\langle \iota(X^n \mid Y^n) \mid Y^n\rangle_\varepsilon \mid Y^n] + O(\log n) \tag{61}$$

as $n \to \infty$.

Lemma 3. *For every* $0 \le \varepsilon \le 1$*, it holds that*

$$\operatorname{ess\,sup} \mathbb{E}[\langle \iota(X^n \mid Y^n) \mid Y^n\rangle_\varepsilon \mid Y^n] = n(1-\varepsilon)\operatorname{ess\,sup} \mathcal{H}(X \mid Y) + \theta_n, \tag{62}$$

where θ_n *is given as in* (39).

Proof: From Proposition 2,

$$\begin{aligned}
&\mathbb{E}[\langle \iota(X^n \mid Y^n) \mid Y^n\rangle_\varepsilon \mid Y^n] \\
&= (1-\varepsilon)\mathcal{H}(X^n \mid Y^n) - \int_{\eta_{Y^n}}^\infty \mathbb{P}\{\iota(X^n \mid Y^n) > t \mid Y^n\}\mathrm{d}t \\
&\quad - \varepsilon(\eta_{Y^n} - \mathcal{H}(X^n \mid Y^n))
\end{aligned} \tag{63}$$

a.s., where $\eta_{Y^n} \ge 0$ and $0 \le \beta_{Y^n} < 1$ are given so that

$$\mathbb{P}\{\iota(X^n \mid Y^n) > \eta_{Y^n} \mid Y^n\} + \beta_{Y^n}\mathbb{P}\{\iota(X^n \mid Y^n) = \eta_{Y^n} \mid Y^n\} \\ = \varepsilon \qquad \text{(a.s.)}. \tag{64}$$

As shown in [4, Equation (150)], we obtain

$$\int_{\eta_{Y^n}}^\infty \mathbb{P}\{\iota(X^n \mid Y^n) > t \mid Y^n\}\mathrm{d}t \\ = \sqrt{\mathcal{V}(X^n \mid Y^n)}(f_G(1-\varepsilon) - \varepsilon\Phi^{-1}(1-\varepsilon)) - B_{Y^n} + D_{Y^n} \tag{65}$$

a.s., where

$$B_{Y^n} \triangleq \operatorname{sgn}(b_{Y^n}) \int_{\min\{0,b_{Y^n}\}}^{\max\{0,b_{Y^n}\}} \mathbb{P}\{\iota(X^n \mid Y^n) > \mathcal{H}(X^n \mid Y^n) \\ + \sqrt{\mathcal{V}(X^n \mid Y^n)}\Phi^{-1}(1-\varepsilon) + t \mid Y^n\}\mathrm{d}t, \tag{66}$$

$$D_{Y^n} \triangleq \sqrt{\mathcal{V}(X^n \mid Y^n)} \int_{\Phi^{-1}(1-\varepsilon)}^{\infty} 1 \times \left(\mathbb{P}\left\{ \frac{\iota(X^n \mid Y^n) - \mathcal{H}(X^n \mid Y^n)}{\sqrt{\mathcal{V}(X^n \mid Y^n)}} > r \,\middle|\, Y^n \right\} - (1 - \Phi(r)) \right) \mathrm{d}r, \tag{67}$$

$$b_{Y^n} \triangleq \eta_{Y^n} - \mathcal{H}(X^n \mid Y^n) - \sqrt{\mathcal{V}(X^n \mid Y^n)}\Phi^{-1}(1-\varepsilon), \tag{68}$$

$$\mathrm{sgn}(u) \triangleq \begin{cases} -1 & \text{if } u < 0, \\ 0 & \text{if } u = 0, \\ 1 & \text{if } u > 0. \end{cases} \tag{69}$$

Combining (63) and (65), we can see that

$$\begin{aligned} &\mathbb{E}[\langle \iota(X^n \mid Y^n) \mid Y^n \rangle_\varepsilon \mid Y^n] \\ &= (1-\varepsilon)\mathcal{H}(X^n \mid Y^n) - \sqrt{\mathcal{V}(X^n \mid Y^n)}(f_G(1-\varepsilon) \\ &\quad - \varepsilon\Phi^{-1}(1-\varepsilon)) + B_{Y^n} - D_{Y^n} - \varepsilon(\eta_{Y^n} - \mathcal{H}(X^n \mid Y^n)) \\ &= (1-\varepsilon)\mathcal{H}(X^n \mid Y^n) \\ &\quad - \sqrt{\mathcal{V}(X^n \mid Y^n)}(f_G(1-\varepsilon) - \varepsilon\Phi^{-1}(1-\varepsilon)) \\ &\quad + B_{Y^n} - D_{Y^n} - \varepsilon(b_Y^n + \sqrt{\mathcal{V}(X^n \mid Y^n)}\Phi^{-1}(1-\varepsilon)) \\ &= (1-\varepsilon)\mathcal{H}(X^n \mid Y^n) - \sqrt{\mathcal{V}(X^n \mid Y^n)} f_G(1-\varepsilon) \\ &\quad + B_{Y^n} - D_{Y^n} - \varepsilon b_Y^n. \end{aligned} \tag{70}$$

Taking the essential supremum in (70), we have

$$\begin{aligned} &\text{ess sup}\, \mathbb{E}[\langle \iota(X^n \mid Y^n) \mid Y^n \rangle_\varepsilon \mid Y^n] \\ &\quad \le \text{ess sup}\, \{(1-\varepsilon)\mathcal{H}(X^n \mid Y^n)\} \\ &\qquad + \text{ess sup}\, \{-\sqrt{\mathcal{V}(X^n \mid Y^n)} f_G(1-\varepsilon) \\ &\qquad\quad + B_{Y^n} - D_{Y^n} - \varepsilon b_{Y^n}\}, \end{aligned} \tag{71}$$

Here, we applied (16) of Proposition 1. Consider the first term in the right-hand side of (71). By Hypothesis (b) in Theorem 1, it follows that

$$\text{ess sup}\, \mathcal{H}(X^n \mid Y^n) = n\, \text{ess sup}\, \mathcal{H}(X \mid Y), \tag{72}$$

Hypothesis (a) and (b) in Theorem 1 imply that

$$V_{\inf}(X \mid Y) > 0 \tag{73}$$

$$T_{\sup}(X \mid Y) \triangleq \text{ess sup}\, \mathcal{T}(X \mid Y) < \infty, \tag{74}$$

respectively. From [4, Equations (171), (174), and (176)],

$$\text{ess sup}\, |b_{Y^n}| \le \text{ess sup}\, \frac{AT_{\sup}(X \mid Y)^{4/3}}{cV_{\inf}(X \mid Y)^{3/2}}, \tag{75}$$

$$\text{ess sup}\, |B_{Y^n}| \le \text{ess sup}\, \frac{AT_{\sup}(X \mid Y)^{4/3}}{cV_{\inf}(X \mid Y)^{3/2}}, \tag{76}$$

$$D_{Y^n} \le \frac{3AT_{\sup}(X \mid Y)}{V_{\inf}(X \mid Y)}. \tag{77}$$

Considering (71) again, we can see that

$$\begin{aligned} &\text{ess sup}\, \mathbb{E}[\langle \iota(X^n \mid Y^n) \mid Y^n \rangle_\varepsilon \mid Y^n] \\ &\overset{\text{(a)}}{\le} \text{ess sup}\, \{n(1-\varepsilon)\mathcal{H}(X \mid Y)\} \\ &\quad + \text{ess sup}\, \{-\sqrt{\mathcal{V}(X^n \mid Y^n)} f_G(1-\varepsilon)\} \\ &\quad + \text{ess sup}\, \{B_{Y^n} - D_{Y^n} - \varepsilon b_{Y^n}\} \\ &\overset{\text{(b)}}{\le} \text{ess sup}\, \{n(1-\varepsilon)\mathcal{H}(X \mid Y)\} \\ &\quad - \text{ess inf}\, \{\sqrt{\mathcal{V}(X^n \mid Y^n)} f_G(1-\varepsilon)\} + O(1) \end{aligned} \tag{78}$$

where

- (a) follows from (72) and (16) of Proposition 1,
- (b) follows from (75), (76) and (77).

Noting (73), we see that

$$\sqrt{nV_{\inf}(X \mid Y)} \le \sqrt{V_{\inf}(X^n \mid Y^n)} \tag{79}$$

Combining (78) and (79), we obtain

$$\begin{aligned} &\text{ess sup}\, \mathbb{E}[\langle \iota(X^n \mid Y^n) \mid Y^n \rangle_\varepsilon \mid Y^n] \\ &\quad \le \text{ess sup}\, \{n(1-\varepsilon)\mathcal{H}(X \mid Y)\} && (80) \\ &\qquad - \sqrt{nV_{\inf}(X \mid Y)} f_{\mathrm{G}}(\varepsilon) + O(1). && (81) \end{aligned}$$

Similarly, we consider the lower bound as follows:

$$\begin{aligned} &\text{ess sup}\, \mathbb{E}[\langle \iota(X^n \mid Y^n) \mid Y^n \rangle_\varepsilon \mid Y^n] \\ &\quad \ge \text{ess sup}\, \{n(1-\varepsilon)\mathcal{H}(X \mid Y)\} \\ &\quad - \text{ess sup}\, \{\sqrt{\mathcal{V}(X^n \mid Y^n)} f_G(1-\varepsilon)\} + O(1). \end{aligned} \tag{82}$$

Similar to (79), one has

$$\sqrt{V_{\sup}(X^n \mid Y^n)} \le \sqrt{nV_{\sup}(X \mid Y)}. \tag{83}$$

Combining (82) and (83),

$$\begin{aligned} &\text{ess sup}\, \mathbb{E}[\langle \iota(X^n \mid Y^n) \mid Y^n \rangle_\varepsilon \mid Y^n] \\ &\quad \ge \text{ess sup}\, \{n(1-\varepsilon)\mathcal{H}(X \mid Y)\} && (84) \\ &\qquad - \sqrt{nV_{\sup}(X \mid Y)} f_{\mathrm{G}}(\varepsilon) + O(1). && (85) \end{aligned}$$

Finally, by (81) and (85), we obtain

$$\begin{aligned} &\text{ess sup}\, \mathbb{E}[\langle \iota(X^n \mid Y^n) \mid Y^n \rangle_\varepsilon \mid Y^n] \\ &\quad = n(1-\varepsilon)\text{ess sup}\, \mathcal{H}(X \mid Y) + \tilde{\theta}_n, \end{aligned} \tag{86}$$

where the reminder term $\tilde{\theta}_n$ satisfies that

$$\begin{aligned} &-\sqrt{nV_{\sup}(X \mid Y)} f_{\mathrm{G}}(\varepsilon) + O(1) \\ &\qquad \le \tilde{\theta}_n \le -\sqrt{nV_{\inf}(X \mid Y)} f_{\mathrm{G}}(\varepsilon) + O(1) \end{aligned} \tag{87}$$

as $n \to \infty$. This completes the proof of Lemma 3. ■

Theorem1 is now proved from Lemmas 1, 2, and 3.

4. Concluding Remarks

We discussed the variable-length data compression for $(L, \varepsilon)_{\max,\max}$-code defined in Definition 6. To investigate the fundamental limit $L^*_{\max,\max}$ of that code in the regime of nonvanishing error probabilities, we derived one-shot bounds formulated by the conditional cutoff operation and its asymptotic expansion up to the remainder term $\Theta(\sqrt{n})$. Since $L^*_{\max,\max}$ can be a converse part of Slepian–Wolf coding, we may immediately obtain

$$L^*_{\max,\max} \le L^*_{\mathrm{SW}}, \tag{88}$$

where L^*_{SW} is the fundamental limit of Slepian–Wolf coding under similar conditions as $L^*_{\max,\max}$. Achievability part of L^*_{SW} remains an open problem. While the exact first-order asymptotics of $L^*_{\max,\max}$ was characterized in a single-letter form in Theorem 1, deriving the exact coefficient of the remainder term $\Theta(\sqrt{n})$ remains a future work of second-order asymptotics. In addition, any analysis of $(L,\varepsilon)_{\max,\mathrm{avg}}$-code (see Table 1) is also left for the future work.

References

[1] A.D. Wyner, "An upper bound on the entropy series," Inf. Control, vol.20, no.2, pp.176–181, March 1972.

[2] N. Alon and A. Orlitsky, "A lower bound on the expected length of one-to-one codes," IEEE Trans. Inf. Theory, vol.40, no.5, pp.1670–1672, Sept. 1994.

[3] V. Kostina, Y. Polyanskiy, and S. Verdú, "Variable-length compression allowing errors," IEEE Trans. Inf. Theory, vol.61, no.8, pp.4316–4330, Aug. 2015.

[4] Y. Sakai and V.Y.F. Tan, "Variable-length source dispersions differ under maximum and average error criteria," IEEE Trans. Inf. Theory, vol.66, no.12, pp.7565–7587, Dec. 2020.

[5] Y. Sakai and V.Y.F. Tan, "On smooth Rényi entropies: A novel information measure, one-shot coding theorems, and asymptotic expansions," IEEE Trans. Inf. Theory, vol.68, no.3, pp.1496–1531, March 2022.

[6] Y. Sakai, R.C. Yavas, and V.Y.F. Tan, "Third-order asymptotics of variable-length compression allowing errors," IEEE Trans. Inf. Theory, vol.67, no.12, pp.7708–7722, Dec. 2021.

[7] D. Slepian and J. Wolf, "Noiseless coding of correlated information source," IEEE Trans. Inf. Theory, vol.IT-19, no.4, pp.471–480, July 1973.

[8] D.-K. He, L.A. Lastras-Montaro, E.-H. Yang, A. Jagmohan, and J. Chen, "On the redundancy of Slepian–Wolf coding," IEEE Trans. Inf. Theory, vol.55, no.12, pp.5607–5627, Dec. 2009.

[9] A. El Gamel and Y.-H. Kim, Network Information Theory, Cambridge University Press, 2011.

Appendix A: Proof of Proposition 1

Let ess sup $X = x$ and ess sup $Y = y$. For an arbitrary small $\delta > 0$, we see that

$$\mathbb{P}\left\{X > x + \frac{\delta}{2}\right\} = \mathbb{P}\left\{Y > y + \frac{\delta}{2}\right\} = 0. \tag{A·1}$$

We know that the following inequality holds

$$\begin{aligned}\mathbb{P}\{X+Y > \alpha+\beta\} &\le \mathbb{P}(\{X > \alpha\} \cup \{Y > \beta\}) &\text{(A·2)}\\ &\le \mathbb{P}\{X > \alpha\} + \mathbb{P}\{Y > \beta\}. &\text{(A·3)}\end{aligned}$$

Letting $\alpha = x + \delta/2$ and $\beta = x + \delta/2$, we get

$$\begin{aligned}\mathbb{P}\{X+Y > x+y+\delta\} &\le \mathbb{P}\left\{X > x + \frac{\delta}{2}\right\} + \mathbb{P}\left\{Y > y + \frac{\delta}{2}\right\}\\ &= 0, &\text{(A·4)}\end{aligned}$$

where (A·4) is satisfied by the definition of ess sup, i.e.,

$$\begin{aligned}\text{ess sup}\,(X+Y) &= \inf\{\gamma \mid \mathbb{P}\{X+Y>\gamma\} = 0\}\\ &= \inf S, &\text{(A·5)}\end{aligned}$$

where

$$\gamma \in S \Leftrightarrow \mathbb{P}\{X+Y > \gamma\} = 0. \tag{A·6}$$

From (A·6),

$$x + y + \delta = \text{ess sup}\, X + \text{ess sup}\, Y + \delta \in S. \tag{A·7}$$

Hence,

$$\text{ess sup}\,(X+Y) \le \text{ess sup}\, X + \text{ess sup}\, Y + \delta. \tag{A·8}$$

Let S' be

$$S' = \{\text{ess sup}\, X + \text{ess sup}\, Y + \delta \mid \delta > 0\}. \tag{A·9}$$

It follows from (A·8) that

$$\begin{aligned}a \in S' &\Rightarrow a \ge \text{ess sup}\,(X+Y), &\text{(A·10)}\\ \inf S' &= \text{ess sup}\, X + \text{ess sup}\, Y. &\text{(A·11)}\end{aligned}$$

Since ess sup $(X+Y)$ is a lower bound of S', we have

$$\inf S' \ge \text{ess sup}\,(X+Y). \tag{A·12}$$

From (A·11) and (A·12)

$$\text{ess sup}\, X + \text{ess sup}\, Y \ge \text{ess sup}\,(X+Y). \tag{A·13}$$

Also, from (16) we see that

$$\text{ess sup}\,(X-Y) \ge \text{ess sup}\, X - \text{ess sup}\, Y, \tag{A·14}$$

proving Proposition 1. ■

Appendix B: Proof of Proposition 3

Let $\epsilon^*(Z,W)$ a $[0,1]$-valued r.v. given as

$$\epsilon^*(Z,W) = \begin{cases} 0 & \text{if } Z < \eta_W,\\ 1 - B_W & \text{if } Z = \eta_W,\\ 1 & \text{if } Z > \eta_W \end{cases} \tag{A·15}$$

where B_W is a Bernoulli r.v. Combining (22) and (A·15), we can see that

$$\begin{aligned}&\mathbb{E}[\epsilon^*(Z,W) \mid W]\\ &\overset{\text{(a)}}{=} \mathbb{E}[(1-B_W)\cdot \mathbf{1}_{\{Z=\eta_W\}} + \mathbf{1}_{\{Z>\eta_W\}} \mid W]\\ &\overset{\text{(b)}}{=} \mathbb{E}[(1-B_W)\cdot \mathbf{1}_{\{Z=\eta_W\}} \mid W] + \mathbb{E}[\mathbf{1}_{\{Z>\eta_W\}} \mid W]\\ &\overset{\text{(c)}}{=} \mathbb{E}[1-B_W \mid W]\cdot \mathbb{E}[\mathbf{1}_{\{Z=\eta_W\}} \mid W] + \mathbb{E}[\mathbf{1}_{\{Z>\eta_W\}} \mid W]\\ &\overset{\text{(d)}}{=} \beta_W\, \mathbb{E}[\mathbf{1}_{\{Z=\eta_W\}} \mid W] + \mathbb{E}[\mathbf{1}_{\{Z>\eta_W\}} \mid W]\\ &\overset{\text{(e)}}{=} \beta_W\, \mathbb{P}\{Z=\eta_W \mid W\} + \mathbb{P}\{Z>\eta_W \mid W\}\end{aligned}$$

$$\overset{\text{(f)}}{=} \varepsilon \qquad \text{(a.s.).} \tag{A$\cdot$16}$$

where

- (a) follows from (A· 15),
- (b) follows from the linearity of conditional expectation,
- (c) follows from the conditional independence $B_W \perp Z \mid W$,
- (d) follows from (21),
- (e) follows by the definition of conditional probability,
- (f) follows from (22).

In addition, we can see that the following equation of three cases. Firstly, if $0 \leq t < \eta_W$,

$$\begin{aligned}
&\mathbb{P}\{(1-\epsilon^*(Z,W))Z > t \mid W\} \\
&\overset{\text{(a)}}{=} \mathbb{E}[\mathbf{1}_{\{(1-\epsilon^*(Z,W))Z>t\}} \cdot (\mathbf{1}_{\{Z<\eta_W\}} \\
&\qquad + \mathbf{1}_{\{Z=\eta_W\}} + \mathbf{1}_{\{Z>\eta_W\}}) \mid W] \\
&\overset{\text{(b)}}{=} \mathbb{E}[\mathbf{1}_{\{(1-\epsilon^*(Z,W))Z>t\}} \cdot \mathbf{1}_{\{Z<\eta_W\}} \\
&\qquad + \mathbf{1}_{\{(1-\epsilon^*(Z,W))Z>t\}} \cdot \mathbf{1}_{\{Z=\eta_W\}} + 0 \mid W] \\
&\overset{\text{(c)}}{=} \mathbb{E}[\mathbf{1}_{\{Z>t\}} \cdot \mathbf{1}_{\{Z<\eta_W\}} + \mathbf{1}_{\{(1-(1-B_W))Z>t\}} \cdot \mathbf{1}_{\{Z=\eta_W\}} \mid W] \\
&\overset{\text{(d)}}{=} \mathbb{E}[\mathbf{1}_{\{\eta_W>Z>t\}} + \mathbf{1}_{\{B_W=1\}} \cdot \mathbf{1}_{\{Z=\eta_W\}} \mid W] \\
&\overset{\text{(e)}}{=} \mathbb{E}[\mathbf{1}_{\{\eta_W>Z>t\}}] + \mathbb{E}[\mathbf{1}_{\{B_W=1\}} \mid W] \cdot \mathbb{E}[\mathbf{1}_{\{Z=\eta_W\}} \mid W] \\
&\overset{\text{(f)}}{=} \mathbb{P}\{\eta_W > Z > t \mid W\} + \mathbb{P}\{B_W = 1 \mid W\} \cdot \mathbb{P}\{Z = \eta_W \mid W\} \\
&\overset{\text{(g)}}{=} \mathbb{P}\{\eta_W > Z > t \mid W\} + (1-\beta_W) \cdot \mathbb{P}\{Z = \eta_W \mid W\} \\
&= \mathbb{P}\{Z > t \mid W\} - \mathbb{P}\{Z > \eta_W \mid W\} - \beta_W \mathbb{P}\{Z = \eta_W \mid W\} \\
&\overset{\text{(h)}}{=} \mathbb{P}\{Z > t \mid W\} - \varepsilon.
\end{aligned} \tag{A$\cdot$17}$$

where

- (a) follows the defining function divided in range of r.v. Z,
- (b) follows from the conditions $0 \leq t < \eta_W$,
- (c) follows from (A· 15),
- (d) follows from the following equation

  $$\mathbf{1}_{\{B_W\eta_W>t\}} \cdot \mathbf{1}_{\{B_W=1\}} = 1 - \mathbf{1}_{\{B_W=0\}}, \tag{A$\cdot$18}$$

  since $0 \leq t < \eta_W$.
- (e) follows from the linearity of the conditional expectation and the conditional independence $B_W \perp Z \mid W$,
- (f) follows from the definition of conditional probability,
- (g) follows from (21),
- (h) follows from (22).

Secondly, if $t < 0$,

$$\begin{aligned}
&\mathbb{P}\{(1-\epsilon^*(Z,W))Z > t \mid W\} \\
&= \mathbb{E}[\mathbf{1}_{\{(1-\epsilon^*(Z,W))Z>t\}} \cdot (\mathbf{1}_{\{Z<\eta_W\}} \\
&\qquad + \mathbf{1}_{\{Z=\eta_W\}} + \mathbf{1}_{\{Z>\eta_W\}}) \mid W] \\
&= \mathbb{E}[\mathbf{1}_{\{Z>t\}} \cdot \mathbf{1}_{\{Z<\eta_W\}} + \mathbf{1}_{\{B_W Z>t\}} \cdot \mathbf{1}_{\{Z=\eta_W\}} \\
&\qquad + \mathbf{1}_{\{t<0\}} \cdot \mathbf{1}_{\{Z>\eta_W\}} \mid W] \\
&= \mathbb{P}\{\eta_W > Z > t \mid W\} + \mathbb{P}\{Z = \eta_W \mid W\} + \mathbb{P}\{Z > \eta_W \mid W\} \\
&= \mathbb{P}\{Z > t \mid W\}.
\end{aligned} \tag{A$\cdot$19}$$

Thirdly, if $t \geq \eta_W$,

$$\begin{aligned}
\mathbb{P}\{(1-\epsilon^*(Z,W))Z > t \mid W\} &= \mathbb{P}\{t < 0 \mid W\} \\
&= 0.
\end{aligned} \tag{A$\cdot$20}$$

To summarize (A· 17), (A· 19) and (A· 20)

$$\begin{aligned}
&\mathbb{P}\{(1-\epsilon^*(Z,W))Z > t \mid W\} \\
&= \begin{cases} \mathbb{P}\{Z > t \mid W\} & \text{if } t < 0, \\ \mathbb{P}\{Z > t \mid W\} - \varepsilon & \text{if } 0 \leq t < \eta_W, \\ 0 & \text{if } t \geq \eta_W \end{cases}
\end{aligned} \tag{A$\cdot$21}$$

a.s. The two random variables $\langle Z \mid W \rangle_\varepsilon$ and $(1-\epsilon^*(Z,W))Z$ are equal in distribution, which implies that

$$\mathbb{E}[\langle Z \mid W \rangle_\varepsilon \mid W] = \mathbb{E}[(1-\epsilon^*(Z,W))Z \mid W]. \tag{A$\cdot$22}$$

Consider an arbitrary measureable map $\epsilon : [0,\infty) \times \mathcal{W} \to [0,1]$ satisfying

$$\mathbb{E}[\epsilon(Z,W) \mid W] \leq \varepsilon \qquad \text{(a.s.).} \tag{A$\cdot$23}$$

Then, we have

$$\begin{aligned}
&\mathbb{E}[(\epsilon(Z,W) - \epsilon^*(Z,W))Z \mid W] \\
&\overset{\text{(a)}}{=} \mathbb{E}[\epsilon(Z,W)Z\mathbf{1}_{\{Z<\eta_W\}} \mid W] \\
&\qquad + \mathbb{E}[(\epsilon(Z,W) - \beta_W)Z\mathbf{1}_{\{Z=\eta_W\}} \mid W] \\
&\qquad + \mathbb{E}[(\epsilon(Z,W) - 1)Z\mathbf{1}_{\{Z>\eta_W\}} \mid W] \\
&\overset{\text{(b)}}{\leq} \eta_W(\mathbb{E}[\epsilon(Z,W)\mathbf{1}_{\{Z<\eta_W\}} \mid W] \\
&\qquad + \mathbb{E}[(\epsilon(Z,W) - \beta_W)\mathbf{1}_{\{Z=\eta_W\}} \mid W] \\
&\qquad + \mathbb{E}[(\epsilon(Z,W) - 1)\mathbf{1}_{\{Z>\eta_W\}} \mid W]) \\
&\overset{\text{(c)}}{=} \eta_W \mathbb{E}[\epsilon(Z,W) - \epsilon^*(Z,W) \mid W] \\
&\leq 0 \qquad \text{(a.s.).}
\end{aligned} \tag{A$\cdot$24}$$

where

- (a) follows from the linearity of expectation,
- (b) follows from the fact that $0 \leq \epsilon(Z,W) \leq 1$, and
- (c) follows from the definition of ϵ^*.

Furthermore, we know that

$$\mathbb{E}[(1-\epsilon^*(Z,W))Z \mid W] \leq \mathbb{E}[(1-\epsilon(Z,W))Z \mid W]. \tag{A$\cdot$25}$$

Combining (A· 16), (A· 22), (A· 23) and (A· 25) we obtain

$$\begin{aligned}
&\mathbb{E}[\langle Z \mid W \rangle_\varepsilon \mid W] \\
&= \min_{\epsilon : \mathbb{E}[\epsilon(Z,W)\mid W] \leq \varepsilon \text{(a.s.)}} \mathbb{E}[(1-\epsilon(Z,W))Z \mid W],
\end{aligned} \tag{A$\cdot$26}$$

completing the proof of Proposition 3. ■

Sho Higuchi was born in Japan in 1999. He received the B.E. degree in Electronical Engineering and Computer Science from University of Hyogo, in 2022. He is currently a second-year master's student with the Department of Electronics and Computer Science in University of Hyogo.

Yuta Sakai was born in Japan in 1992. He received the B.E. and M.E. degrees from the Department of Information Science, University of Fukui, in 2014 and 2016, respectively, and the Ph.D. degree from the Advanced Interdisciplinary Science and Technology, University of Fukui, in 2018. He is currently an Assistant Professor with the Faculty of Materials for Energy, Shimane University. He was formerly an Assistant Professor with the Department of Electronics and Computer Science, University of Hyogo, from 2021 to 2023, and a Research Fellow with the Department of Electrical and Computer Engineering, National University of Singapore (NUS), from 2018 to 2021.