On Some Properties for Universal Coding of Integers and Its Generalization

Wei Yan, Yunghsiang S. Han, *Fellow, IEEE*, and Guozheng Yang

***Abstract*—In the field of lossless source coding, universal coding of integers (UCI) and generalized universal coding of integers (GUCI) are binary codes that are suitable for probability distributions without prior knowledge. UCI $\mathcal{C}$ is defined as a prefix coding in which the constant expansion factor $K_{\mathcal{C}}$ times $\max\{1, H(P)\}$ is greater than or equal to the expected codeword length, where P is the decreasing probability distribution of the source and $H(P)$ is the entropy of P. Since P is decreasing, when the set of codewords of the prefix code $\mathcal{C}$ is determined, the length of the $n+1$-th codeword of the prefix code $\mathcal{C}$ is greater than or equal to the length of the n-th codeword for any positive integer n, at which time the expected codeword length of $\mathcal{C}$ is minimized, and $\mathcal{C}$ is said to be minimal. GUCI $\mathcal{G}$ is defined as a prefix variable-to-variable length (VV) coding for which the constant expansion factor $K_{\mathcal{G}}$ times $H(P)$ is greater than or equal to the coding rate. In this paper, we prove two important theorems for UCI. First, we provide and prove the necessary and sufficient conditions for a minimal prefix code to be UCI. Second, we provide the first proof of an essential theorem for VV codes. This theorem can reveal the connection between UCI and GUCI and prove the converse part of Shannon's first theorem concerning VV codes.**

***Index Terms*—Variable-length codes, Source coding, Universal coding of integers.**

I. INTRODUCTION

A variable-length code is a source code that encodes a single source symbol into variable-length binary bits. Huffman codes [1] are variable-length codes that provide the best compression effect when the underlying probability distribution is known. However, in reality, the probability distributions of most sources are unknown and difficult to measure. Therefore, in 1975, Elias [2] considered the coding problem of universal codes. This class of universal codes is called universal coding of integers (UCI). A variable-length code is termed a *prefix code* if no codeword is a prefix of any other codeword.

Suppose that the discrete memoryless source $\mathcal{S} = (\mathcal{A}, P)$ is considered, where $\mathcal{A} \triangleq \mathbb{N} = \{1, 2, \cdots, r, \cdots\}$ is a countable alphabet and P is a probability distribution that decreases over $\mathcal{A}$ (i.e., $\sum_{r=1}^{\infty} P(r) = 1$, and $0 \leq P(r+1) \leq P(r)$ for all $r \in \mathcal{A}$). The distributions considered in this paper are all decreasing probability distributions. Let $\mathcal{C}$ denote a prefix code for the discrete memoryless source $\mathcal{S} = (\mathcal{A}, P)$, and let $L_{\mathcal{C}}(\cdot)$ be a length function such that $L_{\mathcal{C}}(r)$ denotes the length of the r-th codeword for all $r \in \mathcal{A}$. Let $H(P) \triangleq -\sum_{r=1}^{\infty} P(r) \log_2 P(r)$[1] be the entropy of P, and let $E_P(L_{\mathcal{C}}) \triangleq \sum_{r=1}^{\infty} P(r) L_{\mathcal{C}}(r)$ denote the expected codeword length for $\mathcal{C}$. Elias [2] defined that $\mathcal{C}$ is said to be *universal* if there exists a constant $K_{\mathcal{C}}$ such that

$$\frac{E_P(L_{\mathcal{C}})}{\max\{1, H(P)\}} \leq K_{\mathcal{C}} \tag{1}$$

for all P with $H(P) < \infty$. $K_{\mathcal{C}}$ is termed the *expansion factor* of $\mathcal{C}$. Elias found that when the set of codewords $\{\mathcal{C}(1), \mathcal{C}(2), \cdots, \mathcal{C}(r), \cdots\}$ of the prefix code $\mathcal{C}$ is determined, $E_P(L_{\mathcal{C}})$ is minimized when the codeword length satisfies

$$L_{\mathcal{C}}(1) \leq L_{\mathcal{C}}(2) \leq \cdots \leq L_{\mathcal{C}}(r) \leq \cdots.$$

Elias [2] defined that $\mathcal{C}$ is said to be *minimal* if it satisfies $L_{\mathcal{C}}(r+1) \geq L_{\mathcal{C}}(r)$ for all $r \in \mathcal{A}$. Therefore, it is reasonable to consider minimal prefix codes when studying UCI.

Since this pioneering work was published, many UCI variants have been constructed, and they can be broadly divided into two categories [3, 4]: (1) flag strategy UCIs (see [5–7]), (2) message length strategy UCIs (see [8–11]). Today, UCIs are applied in many applications, such as the inverted file index [12], H.265 video coding standards [13], evolving secret sharing [14], and biological sequence data compression [15, 16].

Recently, Yan *et al.* [17, 18] made it possible to construct a new class of codes that satisfies an inequality similar to

$$\frac{E_P(L_{\mathcal{C}})}{H(P)} \leq K_{\mathcal{C}}.$$

They solved the problem by introducing variable-to-variable length (VV) codes, where the VV codes encode variable-length input chunks as variable-length codewords. They [17, 18] defined generalized universal coding of integers (GUCI) as follows. A VV code $\mathcal{G}$ with a prefix property is said to be *generalized universal* if there exists a constant $K_{\mathcal{G}}$ such that

$$\frac{R_{\mathcal{G}}}{H(P)} \leq K_{\mathcal{G}} \tag{2}$$

for all P with $0 < H(P) < \infty$, where $R_{\mathcal{G}}$ denotes the coding rate of $\mathcal{G}$ and $K_{\mathcal{G}}$ is termed the *expansion factor* of $\mathcal{G}$. The coding rate and the prefix property of VV codes are defined in Section II.

[1]We define $P(r) \log_2 P(r) = 0$ when $P(r) = 0$.

This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFA1004902 and in part by the National Natural Science Foundation of China under Grant 62401599.(*Corresponding author: Guozheng Yang*)

Wei Yan and Guozheng Yang are with the College of Electronic Engineering, National University of Defense Technology, Hefei 230037, China, and also with Anhui Province Key Laboratory of Cyberspace Security Situation Awareness and Evaluation, Hefei 230037, China (email: yan.wei2023@nudt.edu.cn; yangguozheng17@nudt.edu.cn).

Yunghsiang S. Han is with Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen 518000, China (email: yunghsiangh@gmail.com).

In this paper, we prove two important theorems for UCI. First, we provide and prove the necessary and sufficient conditions for a minimal prefix code to be UCI. Second, we provide a proof of an essential theorem for VV codes. This theorem was first introduced by Nishiara *et al.* [19], but they did not present its proof. The theorem was used in their paper [19] to prove a coding theorem for VV codes. Moreover, this theorem can reveal the connection between UCI and GUCI and prove the converse part of Shannon's first theorem about VV codes [17, 18]. The main contributions of this study are summarized below.

1) We are the first to propose the necessary and sufficient conditions for a minimal prefix code to be UCI.
2) We are the first to prove an essential theorem for VV codes, which characterizes the relationship between dictionary entropy $H(\mathcal{D})$ and Shannon entropy $H(P)$ (see Theorem 5 in Section III).

In the remainder of this paper, Section II presents some background knowledge. Section III provides the two main theorems to be proven in this paper. These two theorems are subsequently proven in Section IV and Section V. Section VI summarizes this work.

II. Preliminaries

In this section, we first introduce two lemmas related to UCI and then some definitions concerning variable-to-fixed length (VF) codes and VV codes. Let $|y|$ denote the length of a string y. A dictionary $\mathcal{D}$ denotes a set of some finite-length strings; that is, $\mathcal{D} \subseteq \mathcal{A}^*$, where $\mathcal{A}^*$ denotes the set of all finite-length strings over $\mathcal{A}$. Let $\mathcal{F}$ denote the set of all infinite-length sequences.

A. Two lemmas related to UCI

We begin with two lemmas related to UCI.

Lemma 1 ([20])**.** *Let P be a decreasing probability distribution, and let $H(P) < \infty$. Then,*

$$\sum_{r=1}^{\infty} P(r) \log_2 r \leq H(P) < \infty.$$

Proof. Because P is decreasing,

$$rP(r) \leq \sum_{z=1}^{r} P(z) \leq \sum_{z=1}^{\infty} P(z) = 1.$$

Thus, we have that $\log_2 r \leq -\log_2 P(r)$ for all $r \in \mathcal{A}$ and $P(r) \neq 0$. Furthermore, we obtain

$$\sum_{r=1}^{\infty} P(r) \log_2 r \leq -\sum_{r=1}^{\infty} P(r) \log_2 P(r) = H(P) < \infty.$$

□

A sufficient condition for a prefix code to be UCI is obtained using Lemma 1.

Lemma 2 ([10, 21])**.** *Let $\mathcal{C}$ be a prefix code. If there are two constants R_1 and R_2 such that $L_{\mathcal{C}}(r) \leq R_1 + R_2 \log_2 r$ for all $r \in \mathcal{A}$, then $\mathcal{C}$ is UCI.*

Proof. For any probability distribution P, we have

$$\begin{aligned} E_P(L_{\mathcal{C}}) &= \sum_{r=1}^{\infty} P(r) L_{\mathcal{C}}(r) \\ &\leq \sum_{r=1}^{\infty} \left[R_1 P(r) + R_2 P(r) \log_2 r \right] \\ &= R_1 + R_2 \sum_{r=1}^{\infty} P(r) \log_2 r \\ &\overset{(a)}{\leq} R_1 + R_2 H(P), \end{aligned}$$

where (a) is derived from Lemma 1. Thus, we obtain

$$\frac{E_P(L_{\mathcal{C}})}{\max\{1, H(P)\}} \leq \frac{R_1 + R_2 H(P)}{\max\{1, H(P)\}} \leq R_1 + R_2.$$

Therefore, $\mathcal{C}$ is UCI, and $R_1 + R_2$ is an expansion factor of $\mathcal{C}$. □

B. Variable-to-fixed length codes

VF codes refer to source codes that map variable-length input chunks to fixed-length codewords. A VF encoder comprises a *string encoder* and a *parser*. The encoding process is as follows. The parser first partitions an input sequence y into a series of concatenated chunks y^1, y^2, $\cdots$ from a dictionary $\mathcal{D}$; that is, $y^i \in \mathcal{D}$. Then, the string encoder maps each chunk $y^i \in \mathcal{D}$ to a fixed-length string. Below, we describe the two properties that a dictionary $\mathcal{D}$ may have.

Definition 1 ([22])**.**
1) *Suppose that y^i and y^j are any two different elements contained in the dictionary $\mathcal{D}$. If $y^i \in \mathcal{D}$ is not a prefix of $y^j \in \mathcal{D}$, then $\mathcal{D}$ is called proper.*
2) *Suppose that y is any infinite-length sequence. A dictionary $\mathcal{D}$ is said to be complete if y has a prefix in $\mathcal{D}$.*

For example, the dictionary $\mathcal{D} = \{0, 1, 20, 21, 22\}$ defined over $\{0, 1, 2\}$ is clearly proper. Suppose that y is an infinite-length sequence. If 0 or 1 is the first element of y, then $0 \in \mathcal{D}$ or $1 \in \mathcal{D}$ is its prefix; if 2 is the first element of y, then $20 \in \mathcal{D}$, $21 \in \mathcal{D}$ or $22 \in \mathcal{D}$ is its prefix. Thus, the dictionary $\mathcal{D}$ is complete.

C. Variable-to-variable length codes

Fixed-to-variable length (FV) codes encode fixed-length input chunks as variable-length codewords. A VV code is a concatenation of VF and FV codes. The encoding process for a VV code is as follows.

First, the VF encoder divides the input sequence into chunks with various lengths, and each chunk is mapped to a fixed-length string. Second, the FV encoder then encodes these constant-length strings as variable-length codewords.

Suppose that $d = d_1 d_2 \cdots d_n$ is a finite-length sequence and that $y = y_1 y_2 \cdots$ is an infinite-length sequence, where $d_i \in \mathcal{A}$

and $y_j \in \mathcal{A}$. In this paper, source $\mathcal{S} = (\mathcal{A}, P)$ is a discrete memoryless source; then, we obtain that

$$P(d) = \prod_{i=1}^{n} P(d_i),$$
$$P(y) = \prod_{j=1}^{\infty} P(y_j).$$

The code rate and almost surely complete (ASC) dictionary for VV codes are defined as follows.

Definition 2 ([19]). 1) *A dictionary $\mathcal{D}$ is said to be ASC if the probability that $\mathcal{D}$ has the prefix of an infinite-length sequence is* 1*; that is,*

$$\sum_{y \in \mathcal{B}} P(y) = 1,$$

where

$$\mathcal{B} \triangleq \{y \in \mathcal{F} \mid \exists d \in \mathcal{D}, \text{such that } d \text{ is a prefix of } y\}.$$

2) *Suppose that $\mathcal{G}$ is a VV code with a VV encoder ξ, and an ASC and proper dictionary $\mathcal{D}$. Then, the coding rate of $\mathcal{G}$ is*

$$R_{\mathcal{G}} \triangleq \frac{\sum_{d \in \mathcal{D}} P(d)|\xi(d)|}{\sum_{d \in \mathcal{D}} P(d)|d|}.$$

When the dictionary $\mathcal{D}$ of the VV code $\mathcal{G}$ is equal to the alphabet $\mathcal{A}$, i.e., the VV code $\mathcal{G}$ is a variable-length code, we obtain

$$R_{\mathcal{G}} = \frac{\sum_{r \in \mathcal{A}} P(r)|\xi(r)|}{\sum_{r \in \mathcal{A}} P(r)} = E_P(L_{\mathcal{G}}).$$

An example of a dictionary $\mathcal{D}$ over $\{0, 1\}$ that is proper and ASC is $\mathcal{D} = \{0, 10, 110, 1110, \cdots\}$. Note that the dictionary $\mathcal{D}$ is not complete since the infinite sequence containing all ones has no prefix in the dictionary $\mathcal{D}$.

The VV code $\mathcal{G}$ with a VV encoder ξ and a dictionary $\mathcal{D}$ possesses the prefix property, which means that $\xi(z)$ is not a prefix of $\xi(y)$ for all $y \neq z \in \mathcal{D}$.

D. Three related theorems

First, two theorems related to Theorem 5 in Section III are presented.

Theorem 1 (Lemma 5 in [22])**.** *Let $\mathcal{S} = (\mathcal{J}_c, P)$ be a discrete memoryless source with a finite alphabet $\mathcal{J}_c$ of size c. Let*

$$\overline{l(\mathcal{D})} \triangleq \sum_{d \in \mathcal{D}} P(d)|d|,$$
$$H(\mathcal{D}) \triangleq -\sum_{d \in \mathcal{D}} P(d) \log_2 P(d).$$

Suppose that $\mathcal{C}$ is a VF code with a complete and proper dictionary $\mathcal{D}$; then,

$$H(\mathcal{D}) = \overline{l(\mathcal{D})} H(P).$$

Theorem 2 (Theorem 1 in [23])**.** *Let T denote any stopping time for a sequence of independent identically distributed random variables $Y_1, Y_2, \cdots$. Disregarding the cases $ET = \infty$ and $H(Y_1) = 0$; then,*

$$H(Y^T) = (ET)H(Y_1) + H(T|Y^{\infty}).$$

Theorem 1 is close in form to the formulation of Theorem 5 (which will be proven in this paper), wherea Theorem 2 still seems quite different. In Section III, we explain how the above two theorems are related to Theorem 5.

Second, the converse part of Shannon's first theorem about VV codes is given below.

Theorem 3 ([17, 18])**.** *Consider a discrete memoryless source $S = (\mathcal{A}, P)$, where $\mathcal{A}$ is a countable alphabet and $H(P) < \infty$. If a VV code $\mathcal{G}$ possesses the prefix property, then $R_{\mathcal{G}} \geq H(P)$.*

III. Results

We provide the main conclusions to be proven in this paper. First, we are the first to propose the necessary and sufficient conditions for a minimal prefix code to be UCI. These conditions are shown in Theorem 4.

Theorem 4. *Suppose that $\mathcal{C}$ is a minimal prefix code; then, the following statements are equivalent:*

(a) *$\mathcal{C}$ is UCI.*
(b) *There are two constants R_1 and R_2 such that $L_{\mathcal{C}}(r) \leq R_1 + R_2\lfloor \log_2 r \rfloor$ for all $r \in \mathcal{A}$.*
(c) *There are two constants R_1 and R_2 such that $L_{\mathcal{C}}(r) \leq R_1 + R_2 \log_2 r$ for all $r \in \mathcal{A}$.*

Remark 1 in Section IV tells us that when a prefix code is not minimal, the necessary and sufficient conditions do not hold. Specifically, when $\mathcal{C}$ is a non-minimal prefix code, (a) in Theorem 4 is a necessary but not sufficient condition for (b) or (c) of Theorem 4; that is, when $\mathcal{C}$ is a non-minimal prefix code, (b) or (c) of Theorem 4 leads to (a) of Theorem 4, but (a) does not lead to (b) or (c). However, any non-minimal prefix code $\mathcal{C}_1$ can be obtained by adjusting the order of the corresponding codewords to obtain a minimal prefix code $\mathcal{C}_2$ with the same set of codewords. Therefore, it is possible to change the non-minimal prefix code $\mathcal{C}_1$ to the minimal prefix code $\mathcal{C}_2$, after which the necessary and sufficient conditions can be applied.

Next, we present the following Theorem 5. This theorem reveals the connection between UCI and GUCI (see [17], Theorems 2 and 6) and proves the converse part of Shannon's first theorem about VV codes (see [17], Theorem 1).

Theorem 5 (Lemma 1 in [19])**.** *Let $\mathcal{S} = (\mathcal{A}, P)$ be a discrete memoryless source with a countable alphabet $\mathcal{A}$ and entropy $H(P) < \infty$. Suppose that $\mathcal{G}$ is a VV code with an ASC and proper dictionary $\mathcal{D}$; then,*

$$H(\mathcal{D}) = \overline{l(\mathcal{D})} H(P), \tag{3}$$

where $\overline{l(\mathcal{D})} = \sum_{d \in \mathcal{D}} P(d)|d|$ is the average length of an element in $\mathcal{D}$ and $H(\mathcal{D}) = -\sum_{d \in \mathcal{D}} P(d) \log_2 P(d)$ is the entropy of $\mathcal{D}$.

Theorem 5 was first introduced by Nishiara *et al.* (Lemma 1 in [19]), but they did not provide a corresponding proof. The proof regarding the proper and complete dictionary and the finite alphabet can be found in [22]. Nishiara *et al.* [19] claimed that [23] provided a proof of the proper and ASC dictionary and the finite alphabet version of Theorem 5. However, Ekroot *et al.* [23] studied the entropy of randomly stopped sequences, and some differences exist between the forms of Theorem 5 and Theorem 2. In addition, a similar theorem called *conservation of entropy* [24] was used for memory sources. Unlike previous work, the complete proof presented in Section V is the first to proof of the infinite countable alphabet version of Theorem 5. Note that when the alphabet is infinitely countable, the proof of Theorem 5 is nontrivial and cannot be obtained directly from the conclusions of the relevant references [19, 22, 23]. In addition, Shannon [25] proved Shannon's first theorem about fixed-to-fixed length codes, and the famous textbook on information theory [26, 27] provided Shannon's first theorem about FV codes. The proof of the converse part of Shannon's first theorem about VV codes [17] (i.e., Theorem 3) is necessary to use Theorem 5. Therefore, a complete proof of Theorem 5 is valuable.

IV. Proof of Theorem 4

We provide an auxiliary lemma before proving Theorem 4.

Lemma 3. *Let $\mathcal{C}$ be a minimal UCI and let K be an expansion factor of $\mathcal{C}$. Then, we have that*

$$L_{\mathcal{C}}(r) \leq K + 5K\lfloor \log_2 r \rfloor$$

for all $r \in \mathcal{A}$.

Proof. This lemma is proven by contradiction. Suppose that there is an integer $z_0 \in \mathcal{A}$ such that

$$L_{\mathcal{C}}(z_0) \geq K + 5K\lfloor \log_2 z_0 \rfloor + 1.$$

Let $T \triangleq \lfloor \log_2 z_0 \rfloor$; then, $2^T \leq z_0 \leq 2^{T+1} - 1$. Consider the following two cases.

(1) When $T = 0$, then $z_0 = 1$ and $L_{\mathcal{C}}(1) \geq 1 + K$. We consider the probability distribution $P_1 = (0.5, 0.5)$; then,

$$K \geq \frac{E_{P_1}(L_{\mathcal{C}})}{\max\{1, H(P_1)\}} \geq \frac{1+K}{1} = 1 + K.$$

This is a contradiction. That is, as long as there is a decreasing probability distribution P_1 that fails to satisfy Equation (1), $\mathcal{C}$ is not a UCI.

(2) When $T \geq 1$, we consider the probability distribution P_2:

$$P_2(r) = \begin{cases} \frac{1}{2^{5T}}, & \text{if } r = 1, 2, \cdots, 2^{5T}, \\ 0, & \text{otherwise.} \end{cases}$$

Then, the entropy of P_2 is $H(P_2) = \log_2(2^{5T}) = 5T$. Furthermore, we obtain

$$\begin{aligned} E_{P_2}(L_{\mathcal{C}}) &= \sum_{r=1}^{z_0-1} P_2(r)L_{\mathcal{C}}(r) + \sum_{r=z_0}^{2^{5T}} P_2(r)L_{\mathcal{C}}(r) \\ &> \frac{1}{2^{5T}} \sum_{r=z_0}^{2^{5T}} L_{\mathcal{C}}(r) \\ &\geq \frac{1}{2^{5T}}\left(2^{5T} - z_0 + 1\right)L_{\mathcal{C}}(z_0) \\ &\geq \frac{1}{2^{5T}}\left(2^{5T} - 2^{T+1} + 2\right)(K + 5KT + 1) \\ &> \frac{1}{2^{5T}}\left(2^{5T} - 2^{T+1}\right)(K + 5KT + 1) \\ &= 5KT + \frac{K(2^{4T-1} - 5T - 1) + 2^{4T-1} - 1}{2^{4T-1}} \\ &\overset{(a)}{>} 5KT, \end{aligned}$$

where (a) is due to the fact that $2^{4T-1} - 5T - 1 > 0$ when $T \geq 1$. Thus, we have

$$K \geq \frac{E_{P_2}(L_{\mathcal{C}})}{\max\{1, H(P_2)\}} > \frac{5KT}{5T} = K.$$

This is a contradiction. That is, the decreasing probability distribution P_2 cannot satisfy Equation (1).

The proof is complete. □

Next, Theorem 4 is proven as follows.

Proof. We show that (a), (b) and (c) are equivalent by proving that (a)⇒(b), (b)⇒(c) and (c)⇒(a).

(a)⇒(b): Since $\mathcal{C}$ is UCI, there exists a constant K that is an expansion factor of $\mathcal{C}$. Let $R_1 \triangleq K$ and $R_2 \triangleq 5K$; from Lemma 3, we obtain that $L_{\mathcal{C}}(r) \leq R_1 + R_2\lfloor \log_2 r \rfloor$ for all $r \in \mathcal{A}$.

(b)⇒(c): Because $L_{\mathcal{C}}(r) \leq R_1 + R_2\lfloor \log_2 r \rfloor \leq R_1 + R_2 \log_2 r$, the implication clearly holds.

(c)⇒(a): This holds according to Lemma 2.

□

Remark 1. *When $\mathcal{C}$ is not minimal, (a) of Theorem 4 is a necessary but not sufficient condition for (b) or (c) of Theorem 4. First, it follows from the content of Lemma 2 and the proof of Theorem 4 that (b) or (c) leads to (a). Second, we show that neither (b) nor (c) can be derived from (a). That is, we prove that there exists a UCI $\mathcal{C}_0$ with an integer $n_0 \in \mathcal{A}$ for any two constants R_1 and R_2 such that $L_{\mathcal{C}_0}(n_0) > R_1 + R_2 \log_2 n_0 \geq R_1 + R_2\lfloor \log_2 n_0 \rfloor$. Suppose that $\mathcal{C}$ is any UCI and that K is an expansion factor of $\mathcal{C}$; i.e., $K \max\{1, H(P)\} \geq E_P(L_{\mathcal{C}})$ for all P with $H(P) < \infty$. The prefix code $\mathcal{C}_0$ is constructed as follows.*

$$\mathcal{C}_0(r) = \begin{cases} \mathcal{C}(r)\underbrace{00\cdots0}_{3^{2z}+3^{z}}, & \textit{if } r = 2^{3^z} \textit{ and } z \geq 2, \\ \mathcal{C}(r), & \textit{otherwise.} \end{cases}$$

First, we prove that $\mathcal{C}_0$ is a UCI. Because $P(2^{3^z}) \le \frac{1}{2^{3^z}}$, we obtain

$$\begin{aligned}\sum_{z=2}^{\infty} P(2^{3^z})(3^{2z}+3^z) &\le \sum_{z=2}^{\infty} \frac{1}{2^{3^z}}(3^{2z}+3^z)\\ &< \sum_{z=2}^{\infty} \frac{1}{2^{3^z}}(3^z+1)^2\\ &< \sum_{z=2}^{\infty} \frac{1}{2^{3^z}}(4^z)^2\\ &= \sum_{z=2}^{\infty} 2^{4z-3^z}\\ &\overset{(a)}{\le} \sum_{z=2}^{\infty} 2^{-z+1} = 1,\end{aligned}$$

where (a) is derived from the fact that $4z - 3^z \le -z+1$ for an integer $z \ge 2$. Furthermore, we obtain that

$$\begin{aligned}E_P(L_{\mathcal{C}_0}) &= \sum_{z=2}^{\infty} P(2^{3^z})(3^{2z}+3^z) + E_P(L_{\mathcal{C}})\\ &< 1 + E_P(L_{\mathcal{C}})\\ &\le 1 + K\max\{1, H(P)\}\\ &\le (K+1)\max\{1, H(P)\}\end{aligned}$$

for all P with $H(P) < \infty$. Therefore, $\mathcal{C}_0$ is a UCI.

Second, we prove that for any two constants R_1 and R_2, there exists an integer n_0 such that $L_{\mathcal{C}_0}(n_0) > R_1 + R_2\log_2 n_0 \ge R_1 + R_2\lfloor\log_2 n_0\rfloor$. For any fixed R_1 and R_2, there must be an integer $k_0 \ge 2$ that satisfies $3^{k_0} > \max\{R_1, R_2\}$. We define the integer n_0 as $2^{3^{k_0}}$; then,

$$\begin{aligned}L_{\mathcal{C}_0}(n_0) &= L_{\mathcal{C}}(n_0) + 3^{2k_0} + 3^{k_0}\\ &> 3^{k_0}\log_2 n_0 + 3^{k_0}\\ &> R_1 + R_2\log_2 n_0\\ &\ge R_1 + R_2\lfloor\log_2 n_0\rfloor.\end{aligned}$$

V. Proof of Theorem 5

The proof of Theorem 5 in this section draws on the proof techniques employed in [22, 23]. For a better understanding by the reader, we provide a complete proof. Let us first present some definitions.

Definition 3. *Suppose that $\mathcal{D}$ is a dictionary.*

1) *Let $\mathcal{A}^z$ denote all strings of length z over the alphabet $\mathcal{A}$. For any integer $z \in \mathbb{N}$, the three corresponding dictionaries are defined as follows:*

$$\begin{aligned}T_z &\triangleq \{\zeta \in \mathcal{A}^z \,\big|\, \text{any string } \eta \in \mathcal{D} \text{ is not a prefix of } \zeta\},\\ \mathcal{D}_z^{\perp} &\triangleq \{\zeta \in \mathcal{D} \,\big|\, |\zeta| = z\} \cup T_z,\\ \mathcal{D}_z &\triangleq \{\zeta \in \mathcal{D} \,\big|\, |\zeta| < z\} \cup \mathcal{D}_z^{\perp}.\end{aligned}$$

In particular, $T_1 = \{\zeta \in \mathcal{A} \,|\, \zeta \notin \mathcal{D}\}$ and $\mathcal{D}_1^{\perp} = \mathcal{D}_1 = \mathcal{A}$.

2) *For every string η, let*

$$(\mathcal{D}, \eta) \triangleq \{\zeta \in \mathcal{D} \,\big|\, \eta \text{ is a prefix of } \zeta\}.^2$$

In particular, when $\mathcal{D}$ is proper and $\eta \in \mathcal{D}$, we have that $(\mathcal{D}, \eta) = \{\eta\}$.

3) *For every $\eta \in \mathcal{D}$, let $\mathcal{D}[\eta]$ denote a dictionary as follows:*

$$\mathcal{D}[\eta] \triangleq (\mathcal{D} \setminus \{\eta\}) \cup \eta\mathcal{A},$$

where $\eta\mathcal{A} \triangleq \{\eta\zeta \mid \zeta \in \mathcal{A}\}$. The dictionary $\mathcal{D}[\eta]$ is said to be an extension of dictionary $\mathcal{D}$, and η is termed the extending string from $\mathcal{D}$ to $\mathcal{D}[\eta]$.

When $\mathcal{D}$ is proper and complete, $\mathcal{D}[\eta]$ is also proper and complete. Before Theorem 5 is proven, an auxiliary lemma is introduced.

Lemma 4. (1) *If the dictionary $\mathcal{D}$ is proper, then $\mathcal{D}_z$ is proper and complete.*

(2) *T_z is the set of extending strings from $\mathcal{D}_z$ to $\mathcal{D}_{z+1}$.*

(3) *If the dictionary $\mathcal{D}$ is proper and ASC, then*

$$P(\eta) = \sum_{\zeta \in (\mathcal{D}, \eta)} P(\zeta)$$

for any given string η, where η has no prefix with a length less than $|\eta|$ in $\mathcal{D}$.

(4) *If the dictionary $\mathcal{D}$ is proper, then*

$$\sum_{\eta \in \mathcal{D}_z^{\perp}} (\mathcal{D}, \eta) = \{\zeta \in \mathcal{D} \,\big|\, |\zeta| \ge z\}$$

for all $z \in \mathbb{N}$.

Proof. (1) First, we prove that the dictionary $\mathcal{D}_z$ is proper. The dictionary $\mathcal{D}_z$ can be expressed as follows:

$$\mathcal{D}_z \triangleq \{\zeta \in \mathcal{D} \,\big|\, |\zeta| \le z\} \cup T_z.$$

The following two cases are considered for any string $\zeta_i \in \mathcal{D}_z$.

a) Assume that $\zeta_i \in \{\zeta \in \mathcal{D} \,|\, |\zeta| \le z\}$. Since $\mathcal{D}$ is proper, ζ_i is not a prefix of $\zeta_j \in \{\zeta \in \mathcal{D} \,|\, |\zeta| \le z\} \setminus \{\zeta_i\}$. According to the definition of T_z, ζ_i is not a prefix of η for all $\eta \in T_z$.

b) Assume that $\zeta_i \in T_z$. Because $|\zeta_i| = z$, ζ_i is not a prefix of ζ_j for all $\zeta_j \in \mathcal{D}_z \setminus \{\zeta_i\}$.

Therefore, $\mathcal{D}_z$ is proper.

Second, we prove that $\mathcal{D}_z$ is complete. For any infinite sequence, we consider its first z-bit string η. Assume that an $\alpha \in \mathcal{D}$ exists such that α is a prefix of η. Then, $\alpha \in \{\zeta \in \mathcal{D} \,|\, |\zeta| \le z\} \subseteq \mathcal{D}_z$ is the prefix of the infinite sequence. Assume that any string $\zeta \in \mathcal{D}$ is not a prefix of η. Then, $\eta \in T_z \subseteq \mathcal{D}_z$ is the prefix of the infinite sequence. This part of the proof is complete.

(2) According to the definition of $\mathcal{D}_z$, the set of extending strings from $\mathcal{D}_z$ to $\mathcal{D}_{z+1}$ is essentially composed of elements with lengths of z in $\mathcal{D}_z$ that do not belong to $\mathcal{D}$. This is because T_z consists of elements with lengths

[2] In our setup, η is the prefix of η. For example, the string dcba has the prefixes d, dc, dcb, and $dcba$, while its proper prefixes are d, dc, and dcb. Therefore, $\eta \in (\mathcal{D}, \eta) \neq \emptyset$ when $\eta \in \mathcal{D}$.

of z in $\mathcal{D}_z$ that do not belong to $\mathcal{D}$. Therefore, T_z is the set of extending strings from $\mathcal{D}_z$ to $\mathcal{D}_{z+1}$.

(3) For any finite sequence ξ, let $\mathcal{I}_\xi$ denote the set consisting of all infinite sequences beginning with ξ, and let $\mathcal{H}_\xi$ denote the set consisting of all infinite sequences beginning with ξ that possess probabilities greater than 0. That is,

$$\mathcal{I}_\xi = \{\delta = \xi y_1 y_2 \cdots \in \mathcal{F} \,|\, y_i \in \mathcal{A}\},$$
$$\mathcal{H}_\xi = \{\delta = \xi y_1 y_2 \cdots \in \mathcal{F} \,|\, y_i \in \mathcal{A}, P(\delta) > 0\}.$$

First, the sum of the probabilities of the elements contained in $\mathcal{H}_\xi$ is

$$\begin{aligned}\sum_{\delta \in \mathcal{H}_\xi} P(\delta) = \sum_{\delta \in \mathcal{I}_\xi} P(\delta) & \\ &= \sum_{\delta = \xi y_1 y_2 \cdots \in \mathcal{I}_\xi} \big(P(\xi) \prod_{j=1}^{\infty} P(y_j)\big) \\ &= P(\xi) \sum_{y_j \in \mathcal{A}, j \in \mathbb{N}} \prod_{j=1}^{\infty} P(y_j) \\ &= P(\xi) \prod_{j=1}^{\infty} \big(\sum_{y_j \in \mathcal{A}} P(y_j)\big) \\ &= P(\xi).\end{aligned} \tag{4}$$

Second, the following proves that

$$\mathcal{H}_\eta = \bigsqcup_{\zeta \in (\mathcal{D}, \eta)} \mathcal{H}_\zeta, \tag{5}$$

where $\bigsqcup$ denotes the disjoint union of sets. This process is proven in the following three steps.

a) $\bigcup_{\zeta \in (\mathcal{D},\eta)} \mathcal{H}_\zeta \subseteq \mathcal{H}_\eta$: For every $\delta \in \bigcup_{\zeta \in (\mathcal{D},\eta)} \mathcal{H}_\zeta$, it follows from the definitions of $(\mathcal{D}, \eta)$ and $\mathcal{H}_\zeta$ that δ is an infinite sequence beginning with η; that is, $\delta \in \mathcal{H}_\eta$.

b) $\mathcal{H}_\eta \subseteq \bigcup_{\zeta \in (\mathcal{D},\eta)} \mathcal{H}_\zeta$: For every $\delta \in \mathcal{H}_\eta$, since $\mathcal{D}$ is proper and ASC, the infinite sequence δ has a unique prefix $\zeta \in \mathcal{D}$. Suppose that $|\zeta| < |\eta|$; then, $\zeta \in \mathcal{D}$ is a prefix of η because $\delta \in \mathcal{H}_\eta$. This contradicts the fact that η has no prefix with a length less than $|\eta|$ in $\mathcal{D}$. Thus, $|\zeta| \geq |\eta|$ and η is a prefix of $\zeta \in \mathcal{D}$; that is, $\zeta \in (\mathcal{D}, \eta)$. Hence, $\delta \in \bigcup_{\zeta \in (\mathcal{D},\eta)} \mathcal{H}_\zeta$.

c) $\mathcal{H}_{\zeta_1} \bigcap \mathcal{H}_{\zeta_2} = \emptyset$ for every $\zeta_1, \zeta_2 \in (\mathcal{D}, \eta)$ and $\zeta_1 \neq \zeta_2$: Otherwise, assume that $\delta \in \mathcal{H}_{\zeta_1} \bigcap \mathcal{H}_{\zeta_2} \neq \emptyset$; then, the infinite sequence δ has the prefixes ζ_1 and ζ_2. Without loss of generality, assume that $|\zeta_1| < |\zeta_2|$; then, ζ_1 is a prefix of ζ_2. Since $\zeta_1, \zeta_2 \in (\mathcal{D}, \eta)$, we have that $\zeta_1, \zeta_2 \in \mathcal{D}$. This contradicts the fact that the dictionary $\mathcal{D}$ is proper.

Finally, we obtain

$$\begin{aligned}P(\eta) &\overset{(a)}{=} \sum_{\delta \in \mathcal{H}_\eta} P(\delta) \\ &\overset{(b)}{=} \sum_{\delta \in \bigsqcup_{\zeta \in (\mathcal{D},\eta)} \mathcal{H}_\zeta} P(\delta) \\ &= \sum_{\zeta \in (\mathcal{D},\eta)} \sum_{\delta \in \mathcal{H}_\zeta} P(\delta) \\ &\overset{(c)}{=} \sum_{\zeta \in (\mathcal{D},\eta)} P(\zeta)\end{aligned}$$

where (a) and (c) are derived from Equation (4), and (b) stems from Equation (5).

(4) First, because $\eta \in \mathcal{D}_z^\perp$, we have $|\eta| = z$. Thus, $|\zeta| \geq z$ for every $\zeta \in (\mathcal{D}, \eta)$. Therefore, we obtain

$$\sum_{\eta \in \mathcal{D}_z^\perp} (\mathcal{D}, \eta) \subseteq \{\zeta \in \mathcal{D} \,\big|\, |\zeta| \geq z\}.$$

Second, owing to Lemma 4(1), $\mathcal{D}_z$ is proper and complete. Therefore, for every $\delta \in \{\zeta \in \mathcal{D} \,\big|\, |\zeta| \geq z\}$, δ has a unique prefix $\eta \in \mathcal{D}_z$. Since δ belongs to $\mathcal{D}$ and $\mathcal{D}$ is proper, $\eta \notin \{\zeta \in \mathcal{D} \,|\, |\zeta| < z\}$; that is, $\eta \in \mathcal{D}_z^\perp$. Thus, we obtain

$$\delta \in \sum_{\eta \in \mathcal{D}_z^\perp} (\mathcal{D}, \eta).$$

Therefore, we obtain

$$\sum_{\eta \in \mathcal{D}_z^\perp} (\mathcal{D}, \eta) \supseteq \{\zeta \in \mathcal{D} \,\big|\, |\zeta| \geq z\}.$$

□

Now, we begin the proof of Theorem 5.

Proof. The proof is divided into three parts. First, we prove that if a dictionary satisfies Equation (3), then the dictionary obtained after applying finite extensions also satisfies Equation (3). Next, the following equation is proven:

$$H(\mathcal{D}_n) = \overline{l(\mathcal{D}_n)} H(P) \tag{6}$$

for all $n \in \mathbb{N}$. Finally, the proof for Equation (3) is presented.

1) Suppose that $\mathcal{S}$ is a dictionary. We need to prove that when $\mathcal{S}$ satisfies Equation (3), the $\mathcal{S}[\eta]$ acquired after one extension also satisfies Equation (3). Note that

$$\begin{aligned}\overline{l(\mathcal{S}[\eta])} &= \sum_{\zeta \in \mathcal{S} \setminus \{\eta\}} P(\zeta)|\zeta| + \sum_{\zeta \in \eta\mathcal{A}} P(\zeta)|\zeta| \\ &= \sum_{\zeta \in \mathcal{S}} P(\zeta)|\zeta| - P(\eta)|\eta| + P(\eta)(|\eta| + 1) \\ &= \overline{l(\mathcal{S})} + P(\eta),\end{aligned}$$

and

$$\begin{aligned}H(\mathcal{S}[\eta]) &= -\sum_{\zeta\in\mathcal{S}\setminus\{\eta\}}P(\zeta)\log_2 P(\zeta) - \sum_{\zeta\in\eta\mathcal{A}}P(\zeta)\log_2 P(\zeta)\\ &= -\sum_{\zeta\in\mathcal{S}}P(\zeta)\log_2 P(\zeta) + P(\eta)\log_2 P(\eta)\\ &\quad -\sum_{\zeta\in\mathcal{A}}P(\eta)P(\zeta)\log_2 P(\eta)P(\zeta)\\ &= H(\mathcal{S}) + P(\eta)\log_2 P(\eta) - P(\eta)\log_2 P(\eta)\\ &\quad - P(\eta)\sum_{\zeta\in\mathcal{A}}P(\zeta)\log_2 P(\zeta)\\ &= \overline{l(\mathcal{S})}H(P) + P(\eta)H(P)\\ &= \overline{l(\mathcal{S}[\eta])}H(P).\end{aligned}$$

We have proven that for one extension, the extended dictionary also satisfies Equation (3). With a similar process, we can prove that for any finite number of extensions, the corresponding extended dictionary satisfies Equation (3). The first part of the proof is complete.

2) We prove Equation (6) via mathematical induction. When $n = 1$, we have $\mathcal{D}_1 = \mathcal{A}$ and

$$\overline{l(\mathcal{D}_1)}H(P) = 1 \times H(P) = H(\mathcal{D}_1).$$

Suppose that Equation (6) holds when $n = r$. Now, we consider the extension process from $\mathcal{D}_r$ to $\mathcal{D}_{r+1}$. If $|T_r| < \infty$, then $\mathcal{D}_{r+1}$ is obtained by $\mathcal{D}_r$ after a finite number of extensions. We obtain $H(\mathcal{D}_{r+1}) = \overline{l(\mathcal{D}_{r+1})}H(P)$ because of the first part of the proof. If $|T_r| = \infty$, because $\mathcal{A}$ is countable and the lengths of the elements contained in T_r are all r, T_r is also countable. Therefore, it can be assumed that $T_r \triangleq \{\alpha_i\}_{i=1}^{\infty}$, the extension process from $\mathcal{D}_r$ to $\mathcal{D}_{r+1}$ is as follows:

$$\begin{aligned}\mathcal{D}_{r+1,1} &\triangleq (\mathcal{D}_r \setminus \{\alpha_1\}) \cup \alpha_1\mathcal{A},\\ \mathcal{D}_{r+1,2} &\triangleq (\mathcal{D}_{r+1,1} \setminus \{\alpha_2\}) \cup \alpha_2\mathcal{A}\\ &= (\mathcal{D}_r \setminus \{\alpha_i\}_{i=1}^{2}) \cup \{\alpha_i\mathcal{A}\}_{i=1}^{2},\\ &\vdots\\ \mathcal{D}_{r+1,z} &\triangleq (\mathcal{D}_{r+1,z-1} \setminus \{\alpha_z\}) \cup \alpha_z\mathcal{A}\\ &= (\mathcal{D}_r \setminus \{\alpha_i\}_{i=1}^{z}) \cup \{\alpha_i\mathcal{A}\}_{i=1}^{z},\\ &\vdots\end{aligned}$$

Then, we have the following three equations.

(i) $H(\mathcal{D}_{r+1,z}) = \overline{l(\mathcal{D}_{r+1,z})}H(P)$ for all $z \in \mathbb{N}$.

(ii) $\lim_{z\to+\infty} \mathcal{D}_{r+1,z} = \mathcal{D}_{r+1}$.

(iii) $\mathcal{D}_{r+1} = (\mathcal{D}_r \setminus \{\alpha_i\}_{i=1}^{\infty}) \cup \{\alpha_i\mathcal{A}\}_{i=1}^{\infty}$.

Next, we prove the following two equations:

$$\begin{aligned}\lim_{z\to+\infty} \overline{l(\mathcal{D}_{r+1,z})} &= \overline{l(\mathcal{D}_{r+1})},\\ \lim_{z\to+\infty} H(\mathcal{D}_{r+1,z}) &= H(\mathcal{D}_{r+1}).\end{aligned} \tag{7}$$

First, we have

$$\begin{aligned}\overline{l(\mathcal{D}_{r+1})} &= \sum_{\eta\in\mathcal{D}_{r+1}} P(\eta)|\eta|\\ &\geq \sum_{\eta\in\mathcal{D}_{r+1,z}} P(\eta)|\eta|\\ &= \overline{l(\mathcal{D}_{r+1,z})}\\ &\geq \sum_{\eta\in(\mathcal{D}_r\setminus\{\alpha_i\}_{i=1}^{\infty})\cup\{\alpha_i\mathcal{A}\}_{i=1}^{z}} P(\eta)|\eta|.\end{aligned}$$

Taking the limit $z \to \infty$, we obtain

$$\begin{aligned}\overline{l(\mathcal{D}_{r+1})} &\geq \lim_{z\to+\infty} \overline{l(\mathcal{D}_{r+1,z})}\\ &\geq \lim_{z\to+\infty} \sum_{\eta\in(\mathcal{D}_r\setminus\{\alpha_i\}_{i=1}^{\infty})\cup\{\alpha_i\mathcal{A}\}_{i=1}^{z}} P(\eta)|\eta|\\ &= \overline{l(\mathcal{D}_{r+1})}.\end{aligned}$$

Then, we have

$$\begin{aligned}H(\mathcal{D}_{r+1}) &= -\sum_{\eta\in(\mathcal{D}_r\setminus\{\alpha_i\}_{i=1}^{\infty})\cup\{\alpha_i\mathcal{A}\}_{i=1}^{z}} P(\eta)\log_2 P(\eta)\\ &\quad -\sum_{\eta\in\{\alpha_i\mathcal{A}\}_{i=z+1}^{\infty}} P(\eta)\log_2 P(\eta)\\ &\overset{(a)}{\geq} -\sum_{\eta\in(\mathcal{D}_r\setminus\{\alpha_i\}_{i=1}^{\infty})\cup\{\alpha_i\mathcal{A}\}_{i=1}^{z}} P(\eta)\log_2 P(\eta)\\ &\quad -\sum_{\eta\in\{\alpha_i\}_{i=z+1}^{\infty}} P(\eta)\log_2 P(\eta)\\ &= -\sum_{\eta\in(\mathcal{D}_r\setminus\{\alpha_i\}_{i=1}^{z})\cup\{\alpha_i\mathcal{A}\}_{i=1}^{z}} P(\eta)\log_2 P(\eta)\\ &= H(\mathcal{D}_{r+1,z})\\ &\geq -\sum_{\eta\in(\mathcal{D}_r\setminus\{\alpha_i\}_{i=1}^{\infty})\cup\{\alpha_i\mathcal{A}\}_{i=1}^{z}} P(\eta)\log_2 P(\eta),\end{aligned}$$

where (a) is due to the fact that

$$-\sum_{\eta\in\alpha_i\mathcal{A}} P(\eta)\log_2 P(\eta) \geq -P(\alpha_i)\log_2 P(\alpha_i)$$

for all $i \in \mathbb{N}$. Taking the limit $z \to \infty$, we obtain

$$\begin{aligned}H(\mathcal{D}_{r+1}) &\geq \lim_{z\to+\infty} H(\mathcal{D}_{r+1,z})\\ &\geq \lim_{z\to+\infty} -\sum_{\eta\in(\mathcal{D}_r\setminus\{\alpha_i\}_{i=1}^{\infty})\cup\{\alpha_i\mathcal{A}\}_{i=1}^{z}} P(\eta)\log_2 P(\eta)\\ &= H(\mathcal{D}_{r+1}).\end{aligned}$$

Equation (7) is proven. From the perspective of mathematical analysis, Equation (7) essentially involves considering whether the function and the limit can be exchanged. For example, $\lim_{z\to+\infty} H(\mathcal{D}_{r+1,z}) = H(\lim_{z\to+\infty} \mathcal{D}_{r+1,z})$. Finally, from Equation (7), we have

$$\begin{aligned}H(\mathcal{D}_{r+1}) &= \lim_{z\to+\infty} H(\mathcal{D}_{r+1,z})\\ &= \lim_{z\to+\infty} \overline{l(\mathcal{D}_{r+1,z})}H(P)\\ &= \overline{l(\mathcal{D}_{r+1})}H(P).\end{aligned}$$

The second part of the proof is complete.

3) We prove the following two equations similar to Equation (7).

$$\lim_{r\to+\infty} \overline{l(\mathcal{D}_r)} = \overline{l(\mathcal{D})}. \qquad \lim_{r\to+\infty} H(\mathcal{D}_r) = H(\mathcal{D}). \tag{8}$$

First, according to Lemma 4(3), we obtain that

$$\sum_{\eta\in(\mathcal{D},\zeta)} P(\eta)|\eta| \geq |\zeta| \sum_{\eta\in(\mathcal{D},\zeta)} P(\eta) = P(\zeta)|\zeta|$$

for any given $\zeta \in \mathcal{D}_r^{\perp}$. Furthermore, from Lemma 4(4), we have that

$$\sum_{\eta\in\mathcal{D},|\eta|\geq r} P(\eta)|\eta| = \sum_{\zeta\in\mathcal{D}_r^{\perp}} \sum_{\eta\in(\mathcal{D},\zeta)} P(\eta)|\eta| \geq \sum_{\zeta\in\mathcal{D}_r^{\perp}} P(\zeta)|\zeta|.$$

Therefore, we obtain

$$\begin{aligned}\overline{l(\mathcal{D})} &= \sum_{\eta\in\mathcal{D},|\eta|<r} P(\eta)|\eta| + \sum_{\eta\in\mathcal{D},|\eta|\geq r} P(\eta)|\eta| \\ &\geq \sum_{\eta\in\mathcal{D},|\eta|<r} P(\eta)|\eta| + \sum_{\zeta\in\mathcal{D}_r^{\perp}} P(\zeta)|\zeta| \\ &= \overline{l(\mathcal{D}_r)} \\ &\geq \sum_{\eta\in\mathcal{D},|\eta|<r} P(\eta)|\eta|.\end{aligned}$$

Taking the limit $r \to \infty$, we obtain

$$\begin{aligned}\overline{l(\mathcal{D})} &\geq \lim_{r\to+\infty} \overline{l(\mathcal{D}_r)} \\ &\geq \lim_{r\to+\infty} \sum_{\eta\in\mathcal{D},|\eta|<r} P(\eta)|\eta| \\ &= \overline{l(\mathcal{D})}.\end{aligned}$$

Then, according to Lemma 4(3), we obtain that

$$\begin{aligned}-\sum_{\eta\in(\mathcal{D},\zeta)} P(\eta)\log_2 P(\eta) &\geq -\log_2 P(\zeta) \sum_{\eta\in(\mathcal{D},\zeta)} P(\eta) \\ &= -P(\zeta)\log_2 P(\zeta)\end{aligned}$$

for any given $\zeta \in \mathcal{D}_r^{\perp}$. Furthermore, from Lemma 4(4), we have

$$\begin{aligned}-\sum_{\eta\in\mathcal{D},|\eta|\geq r} P(\eta)\log_2 P(\eta) &= -\sum_{\zeta\in\mathcal{D}_r^{\perp}} \sum_{\eta\in(\mathcal{D},\zeta)} P(\eta)\log_2 P(\eta) \\ &\geq -\sum_{\zeta\in\mathcal{D}_r^{\perp}} P(\zeta)\log_2 P(\zeta).\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}H(\mathcal{D}) &= -\sum_{\eta\in\mathcal{D},|\eta|<r} P(\eta)\log_2 P(\eta) \\ &\quad -\sum_{\eta\in\mathcal{D},|\eta|\geq r} P(\eta)\log_2 P(\eta) \\ &\geq -\sum_{\eta\in\mathcal{D},|\eta|<r} P(\eta)\log_2 P(\eta) \\ &\quad -\sum_{\zeta\in\mathcal{D}_r^{\perp}} P(\zeta)\log_2 P(\zeta) \\ &= H(\mathcal{D}_r) \\ &\geq -\sum_{\eta\in\mathcal{D},|\eta|<r} P(\eta)\log_2 P(\eta),\end{aligned}$$

Taking the limit $r \to \infty$, we obtain

$$\begin{aligned}H(\mathcal{D}) &\geq \lim_{r\to+\infty} H(\mathcal{D}_r) \\ &\geq \lim_{r\to+\infty} -\sum_{\eta\in\mathcal{D},|\eta|<r} P(\eta)\log_2 P(\eta) \\ &= H(\mathcal{D}).\end{aligned}$$

Equation (8) is proven. Finally, from Equation (8), we have

$$\begin{aligned}H(\mathcal{D}) &= \lim_{r\to+\infty} H(\mathcal{D}_r) \\ &= \lim_{r\to+\infty} \overline{l(\mathcal{D}_r)} H(P) \\ &= \overline{l(\mathcal{D})} H(P).\end{aligned}$$

The proof is complete. □

Remark 2. *Theorem 1 [22] is the proper and complete dictionary version of Theorem 5 with a finite alphabet. The relevant proof of [22] can be equated to the first part of the proof; namely, if the given dictionary satisfies Equation* (3)*, then the dictionary obtained after performing finite extensions also satisfies Equation* (3)*. The reason for not using Theorem 1 directly in the proof is that Theorem 1 requires the dictionary $\mathcal{D}$ to be complete and the associated alphabet to be finite.*

When studying the entropy of randomly stopped sequences, Ekroot et al. *[23] presented conclusions related to Theorem 5, that is, Theorem 2. It can be assumed that [23] proved the proper and ASC dictionary version of Theorem 5 with a finite alphabet. Owing to the formal differences that are present in Theorem 2, this theorem cannot be directly used in the proof of Theorem 5. Moreover, we use Lemma 4 (which is proven in this paper) for the relevant part of the proof, which distinguishes it from [23].*

VI. Conclusion

UCI has been studied for almost half a century, but questions that are worth exploring remain. In this paper, we are the first to propose the necessary and sufficient conditions for a minimal prefix code $\mathcal{C}$ to be UCI, as stated below.

1) There are two constants R_1 and R_2 such that $L_{\mathcal{C}}(r) \leq R_1 + R_2\lfloor\log_2 r\rfloor$ for all $r \in \mathcal{A}$.
2) There are two constants R_1 and R_2 such that $L_{\mathcal{C}}(r) \leq R_1 + R_2\log_2 r$ for all $r \in \mathcal{A}$.

In addition, this paper is the first to prove Theorem 5, which characterizes the relationship between dictionary entropy $H(\mathcal{D})$ and Shannon entropy $H(P)$. This theorem leads to a series of useful conclusions [17–19]. It can also be used to prove a coding theorem for VV codes. Moreover, this theorem can reveal the connection between UCI and GUCI and prove the converse part of Shannon's first theorem about VV codes.

References

[1] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, pp. 1098–1101, Sep. 1952.

[2] P. Elias, "Universal codeword sets and representations of the integers," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 194–203, Mar. 1975.

[3] R. M. Capocelli, "Flag encodings related to the zeckendorf representation of integers," in *Sequences, Combinatorics, Compression, Security, and Transmission*. New York, NY, USA: Springer-Verlag, 1990, pp. 449–466.

[4] B. T. Ávila and R. M. C. de Souza, "Meta-Fibonacci codes: Efficient universal coding of natural numbers," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2357–2375, Apr. 2017.

[5] A. Apostolico and A. S. Fraenkel, "Robust transmission of unbounded strings using Fibonacci representations," *IEEE Trans. Inf. Theory*, vol. 33, no. 2, pp. 238–245, Mar. 1987.

[6] M. Wang, "Almost asymptotically optimal flag encoding of the integers," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 324–326, Mar. 1988.

[7] H. Yamamoto and H. Ochi, "A new asymptotically optimal code for the positive integers," *IEEE Trans. Inf. Theory*, vol. 37, no. 5, pp. 1420–1429, Sep. 1991.

[8] Q. F. Stout, "Improved prefix encodings of the natural numbers (corresp.)," *IEEE Trans. Inf. Theory*, vol. 26, no. 5, pp. 607–609, Sep. 1980.

[9] H. Yamamoto, "A new recursive universal code of the positive integers," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 717–723, Mar. 2000.

[10] W. Yan and S.-J. Lin, "On the minimum of the expansion factor for universal coding of integers," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7309–7319, Nov. 2021.

[11] W. Yan and S.-J. Lin, "A tighter upper bound of the expansion factor for universal coding of integers and its code constructions," *IEEE Trans. Commun.*, vol. 70, no. 7, pp. 4429–4438, Jul. 2022.

[12] J. Zobel and A. Moffat, "Inverted files for text search engines," *ACM Comput. Surv.*, vol. 38, no. 2, pp. 1–56, Jul. 2006.

[13] V. Sze, M. Budagavi, and G. J. Sullivan, *High Efficiency Video Coding (HEVC), Algorithms and Architectures*. Springer, 2014.

[14] W. Yan, S.-J. Lin, and Y. S. Han, "A new metric and the construction for evolving 2-threshold secret sharing schemes based on prefix coding of integers," *IEEE Trans. Commun.*, vol. 71, no. 5, pp. 2906–2915, May 2023.

[15] K. Daily, P. Rigor, S. Christley, X. Xie, and P. Baldi, "Data structures and compression algorithms for high-throughput sequencing technologies," *BMC Bioinform.*, vol. 11, p. 514, Oct. 2010.

[16] J. J. Selva and X. Chen, "SRComp: Short read sequence compression using burstsort and Elias omega coding," *PLOS ONE*, vol. 8, no. 12, pp. 1–7, 12 Dec. 2013.

[17] W. Yan and Y. S. Han, "Generalized universal coding of integers," *IEEE Trans. Commun.*, vol. 72, no. 8, pp. 4538–4550, Aug. 2024.

[18] W. Yan and S.-J. Lin, "Generalized universal coding of integers," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Oct. 2021, pp. 1–6.

[19] M. Nishiara and H. Morita, "Almost surely complete parsing and variable-to-variable length coding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2000, p. 347.

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

[21] K. Lakshmanan, "On universal codeword sets," *IEEE Trans. Inf. Theory*, vol. 27, no. 5, pp. 659–662, Sep. 1981.

[22] F. Jelinek and K. Schneider, "On variable-length-to-block coding," *IEEE Trans. Inf. Theory*, vol. 18, no. 6, pp. 765–774, Nov. 1972.

[23] L. Ekroot and T. M. Cover, "The entropy of a randomly stopped sequence," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1641–1644, Nov. 1991.

[24] S. A. Savari, "Variable-to-fixed length codes and the conservation of entropy," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1612–1620, Jul. 1999.

[25] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

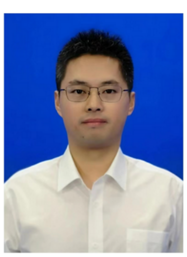
[26] T. M. Cover and J. A. Thomas, *Elements of information theory, 2nd ed.* NY, USA: Wiley, 2006.

[27] F. Alajaji and P. N. Chen, *An Introduction to Single-User Information Theory*. Singapore: Springer, 2018.

Wei Yan received the B.Sc. degree in mathematics and applied mathematics and the Ph.D. degree in Cyberspace Security from the University of Science and Technology of China (USTC), Hefei, China, in 2017 and 2022, respectively. From 2022 to 2023, he was a Researcher with the Theory Laboratory, 2012 Labs, Huawei Technologies Company Ltd. He is currently a University-appointed Associate Professor with the National University of Defense Technology (NUDT), Hefei, China. His research interests include information theory, source coding, cryptography, and information security.

Yunghsiang S. Han (Fellow, IEEE) was born in Taipei, Taiwan, in 1962. He received B.Sc. and M.Sc. degrees in electrical engineering from the National Tsing Hua University, Hsinchu, Taiwan, in 1984 and 1986, respectively, and a Ph.D. from the School of Computer and Information Science, Syracuse University, Syracuse, NY, in 1993. From 1986 to 1988, he was a lecturer at Ming-Hsin Engineering College, Hsinchu, Taiwan. He was a teaching assistant from 1989 to 1992 and a research associate in the School of Computer and Information Science at Syracuse University from 1992 to 1993. From 1993 to 1997, he was an Associate Professor in the Department of Electronic Engineering at Hua Fan College of Humanities and Technology, Taipei Hsien, Taiwan. He was with the Department of Computer Science and Information Engineering at National Chi Nan University, Nantou, Taiwan from 1997 to 2004. He was promoted to Professor in 1998. He was a visiting scholar in the Department of Electrical Engineering at the University of Hawaii at Manoa, HI from June to October 2001, the SUPRIA visiting research scholar in the Department of Electrical Engineering and Computer Science and CASE center at Syracuse University, NY from September 2002 to January 2004 and July 2012 to June 2013, and the visiting scholar in the Department of Electrical and Computer Engineering at the University of Texas at Austin, TX from August 2008 to June 2009. He was with the Graduate Institute of Communication Engineering at National Taipei University, Taipei, Taiwan from August 2004 to July 2010. From August 2010 to January 2017, he was Chair Professor with the Department of Electrical Engineering at the National Taiwan University of Science and Technology. From February 2017 to February 2021, he was with the School of Electrical Engineering & Intelligentization at Dongguan University of Technology, China. Now he is with the Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China. He is also a Chair Professor at National Taipei University since February 2015. His research interests are in error-control coding, wireless networks, and security.

Dr. Han was a winner of the 1994 Syracuse University Doctoral Prize and a Fellow of IEEE. One of his papers won the prestigious 2013 ACM CCS Test-of-Time Award in cybersecurity.



Guozheng Yang received the B.E. degree in computer application, the M.Sc. degree in communication, and the Ph.D. degree in signal and information processing from Hefei Electronic Engineering Institute, Hefei, China, in 2003, 2006, and 2009, respectively. He is currently a Professor with the National University of Defense Technology, Hefei, China. His research interests are computer application and information processing.