Asymptotic Properties on Codeword Lengths of an Optimal FV Code for General Sources

Hiroki Koga, *Member, IEEE,* and Hirosuke Yamamoto, *Senior Member, IEEE*

***Abstract*—This correspondence is concerned with asymptotic properties on the codeword length of a fixed-to-variable length code (FV code) for a general source $\{X^n\}_{n=1}^{\infty}$ with a finite or countably infinite alphabet. Suppose that for each $n \geq 1$ X^n is encoded to a binary codeword $\varphi_n(X^n)$ of length $l(\varphi_n(X^n))$. Letting ε_n denote the decoding error probability, we consider the following two criteria on FV codes: i) $\varepsilon_n = 0$ for all $n \geq 1$ and ii) $\limsup_{n\to\infty} \varepsilon_n \leq \varepsilon$ for an arbitrarily given $\varepsilon \in [0, 1)$. Under criterion i), we show that, if X^n is encoded by an arbitrary prefix-free FV code asymptotically achieving the entropy,**

$$\frac{1}{n} l\left(\varphi_n(X^n)\right) - \frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)} \to 0$$

in probability as $n \to \infty$ under a certain condition, where P_{X^n} denotes the probability distribution of X^n. Under criterion ii), we first determine the minimum rate achieved by FV codes. Next, we show that $\frac{1}{n} l\left(\varphi_n(X^n)\right)$ of an arbitrary FV code achieving the minimum rate in a certain sense has a property similar to the lossless case.

***Index Terms*—Asymptotic optimality, convergence in distribution, fixed-to-variable length coding, general source, information spectrum.**

I. Introduction

Suppose the situation that a random variable $X^n \in \mathcal{X}^n, n \geq 1$, generated from a discrete source is encoded to a binary codeword $\varphi_n(X^n)$ by using a fixed-to-variable length code (FV code) $\varphi_n : \mathcal{X}^n \to \{0,1\}^*$ satisfying the prefix condition, where $\mathcal{X}$ is a finite source alphabet and $\{0,1\}^*$ denotes the set of all binary sequences of finite length. Denote by P_{X^n} the probability distribution of X^n and $l(\varphi_n(X^n))$ the codeword length of the codeword $\varphi_n(x^n)$ for a sequence $x^n \in \mathcal{X}^n$. It is well known that, if no decoding error is permitted, the expected codeword length $E\left[l\left(\varphi_n(X^n)\right)\right]$ satisfies

$$E\left[l\left(\varphi_n(X^n)\right)\right] \geq H(X^n), \qquad \text{for each } n \geq 1 \tag{1}$$

where $H(X^n)$ denotes the entropy of X^n of base 2 and for each $n \geq 1$, the equality in (1) holds if and only if

$$l\left(\varphi_n(x^n)\right) = \log_2 \frac{1}{P_{X^n}(x^n)}, \qquad \text{for all } x^n \in \mathcal{X}^n.$$

In addition, it is also well known that there exists an asymptotically optimal FV code, such as the Huffman code or the Shannon–FanElias code, satisfying

$$\frac{1}{n} E\left[l\left(\varphi_n(X^n)\right)\right] - H(X) \to 0, \qquad \text{as } n \to \infty \tag{2}$$

provided that X^n is composed of n independent and identically distributed (i.i.d.) random variables, where $H(X)$ denotes the entropy of X to base 2. Therefore, we can easily guess from (1) and (2) that the codeword length $l\left(\varphi_n(x^n)\right)$ of the optimal FV code satisfying (2) is nearly equal to $\log_2 \frac{1}{P_{X^n}(x^n)}$.

There are a few studies [2], [9], [10] that discuss relationships between $\log_2 \frac{1}{P_{X^n}(x^n)}$, the *ideal* codeword length, and $l\left(\varphi_n(x^n)\right)$, the *actual* codeword length, for $x^n \in \mathcal{X}^n$ in detail. In fact, the Huffman code satisfies

$$l\left(\varphi_n(x^n)\right) = \log_2 \frac{1}{P_{X^n}(x^n)}, \qquad \text{for all } x^n \in \mathcal{X}^n$$

if and only if X^n is dyadic [2]. Nemetz and Simmon [10] show that

$$\frac{1}{n}\log_2 \frac{1}{P_{X^n}(x^n)} - \frac{1}{n} l\left(\varphi_n(x^n)\right) \to 0$$

uniformly in $x^n \in \mathcal{X}^n$ under the assumption that X^n is i.i.d. In addition to these results, there are FV codes that are known to achieve the entropy rate in almost sure sense for stationary ergodic sources (see, e.g., [11], [12], [16]). For such FV codes it holds that

$$\frac{1}{n} l\left(\varphi_n(x^n)\right) - \frac{1}{n}\log_2 \frac{1}{P_{X^n}(x^n)} \to 0$$

almost surely (therefore in probability) as $n \to \infty$ because of the asymptotic equipartition property (AEP).

The objective of this correspondence is clarifying relationships between the two random variables $\frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)}$ and $\frac{1}{n} l\left(\varphi_n(X^n)\right)$ for a wide class of sources. We take the information-spectrum approach that originates from [4] and is described in detail in [5]. No assumption is imposed on X^n in the information-spectrum framework, which enables us to treat nonstationary and/or nonergodic sources not satisfying the asymptotic equipartition property. We can show that, for any sequence of FV codes achieving the entropy in a certain sense,

$$\frac{1}{n} l\left(\varphi_n(X^n)\right) - \frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)} \to 0$$

in probability as $n \to \infty$ for any source with a finite alphabet $|\mathcal{X}|$. The same result holds for sources with countably infinite alphabets under a certain assumption on the sources.

It is also interesting to investigate relationships between

$$\frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)} \quad \text{and} \quad \frac{1}{n} l\left(\varphi_n(X^n)\right)$$

under another criterion on the decoding error. In this correspondence, we consider a criterion under which the limit superior of the decoding error probability must be less than or equal to ε, where $\varepsilon \in [0, 1)$ is an arbitrary constant given in advance. We first obtain the minimum rate achieved by prefix-free FV codes satisfying the criterion. This result can be regarded as an extension of Han's result [5], [6] that treats the case of $\varepsilon = 0$, We next unveil a relationship between

$$\frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)} \quad \text{and} \quad \frac{1}{n} l\left(\varphi_n(X^n)\right).$$

It is shown that $\frac{1}{n} l\left(\varphi_n(X^n)\right)$ has a property similar to the lossless case for any FV code with a certain optimality.

This correspondence is organized as follows. In Section II, lossless FV codes are treated, and notation is introduced. A class of *mean-optimal* FV codes, which was first defined in [15], is introduced. The codeword length of the mean-optimal FV code is analyzed in detail. Section III is devoted to a characterization of FV codes with the decoding error probability asymptotically upper-bounded by an

Manuscript received July 14, 1999; revised August 9, 2002.

H. Koga is with the Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba-shi, Ibaraki, 305-8573, Japan (e-mail: koga@esys.tsukuba.ac.jp).

H. Yamamoto is with the School of Frontier Sciences, the University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba, 277-8561, Japan (e-mail: Hirosuke@ieee.org).

Communicated by I. Csiszár, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2005.844098

arbitrarily given $\varepsilon \in [0,1)$. We give a formula for the minimum rate achieved by such FV codes. The obtained minimum rate motivates us to introduce a new class of FV codes called the *ε-mean-optimal* FV codes. After that, we analyze the codeword length of the ε-mean-optimal FV code.

II. Lossless FV Codes

Let $\mathcal{X}$ be a finite or countably infinite source alphabet. For each $n \geq 1$, let X^n be a random variable on $\mathcal{X}^n$ subject to a probability distribution P_{X^n}. We set $\boldsymbol{X} = \{X^n\}_{n=1}^{\infty}$. Actually, $\boldsymbol{X}$ can be regarded as an infinite sequence of probability distributions P_{X^n}, $n \geq 1$, not required to satisfy the consistency condition, i.e.,

$$P_{X^n}(x^n) = \sum_{x\in\mathcal{X}} P_{X^{n+1}}(x^n x), \qquad \text{for all } x^n \in \mathcal{X}^n.$$

Such an $\boldsymbol{X}$ is called a general source [4], [5]. The class of the general sources includes nonstationary and/or nonergodic processes.

For each $n \geq 1$, we define an encoder φ_n as a surjective mapping from $\mathcal{X}^n$ to $\mathcal{C}_n$, where $\mathcal{C}_n \subseteq \{0,1\}^*$ is a set of codewords and $\{0,1\}^*$ means the set of all binary sequences of finite length. The length of the codeword $\varphi_n(x^n)$ for an $x^n \in \mathcal{X}^n$ is denoted by $l(\varphi_n(x^n))$. Let $E[l(\varphi_n(X^n))]$ denote the expected codeword length defined by

$$E\left[l(\varphi_n(X^n)\right] = \sum_{x^n\in\mathcal{X}^n} P_{X^n}(x^n)\, l(\varphi_n(x^n)). \tag{3}$$

We define a decoder as a mapping $\psi_n : \mathcal{C}_n \to \mathcal{X}^n$. The decoding error probability is denoted by ε_n, where

$$\varepsilon_n = \Pr\left\{\psi_n\left(\varphi_n(X^n)\right) \neq X^n\right\}. \tag{4}$$

In this section only infinite sequences of FV codes, satisfying $\varepsilon_n = 0$ for all $n \geq 1$, are of interest. Hereafter, an infinite sequence of FV codes is simply called an FV code when there is no confusion.

Kieffer [8] and Han [4], [5] characterize the infimum achievable FV coding rate for a general source $\boldsymbol{X}$. Precisely, Kieffer's result is restricted to the case where $|\mathcal{X}|$ is finite and $\boldsymbol{X}$ satisfies the consistency condition.

Definition 1: A rate R is called achievable if there exists a prefix-free FV code $\{(\varphi_n,\psi_n)\}_{n=1}^{\infty}$ satisfying

$$\limsup_{n\to\infty} \frac{1}{n} E\left[l\left(\varphi_n(X^n)\right)\right] \leq R$$

and $\varepsilon_n = 0$ for all $n \geq 1$. In particular, the infimum of the achievable rate is called the infimum achievable FV coding rate and is denoted by $R(\boldsymbol{X})$.

Theorem 1 ([4], [8]):

$$R(\boldsymbol{X}) = H(\boldsymbol{X}) \stackrel{\text{def}}{=} \limsup_{n\to\infty} \frac{1}{N} H(X^n)$$

where $H(X^n)$ denotes the entropy of X^n defined by

$$H(X^n) = \sum_{x^n\in\mathcal{X}^n} P_{X^n}(x^n) \log_2 \frac{1}{P_{X^n}(x^n)}.$$

Notice that $\frac{1}{n}H(X^n)$, $n \geq 1$, may not have the limit since the stationarity of $\boldsymbol{X}$ is not assumed.

In this correspondence, we are interested in asymptotic behaviors of two random variables

$$\frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)} \quad \text{and} \quad \frac{1}{n} l\left(\varphi_n(X^n)\right).$$

While Theorem 1 tells us that for any $\gamma > 0$ there exists an optimal FV code satisfying $\frac{1}{n}E[l(\varphi_n(X^n))] \leq H(\boldsymbol{X}) + \gamma$ for all sufficiently large n, we can know little about asymptotic behavior of these two random variables. In order to develop an interesting relationship between the two random variables, we introduce another stronger notion of optimality on FV codes.

Definition 2: An FV code $\{(\varphi_n,\psi_n)\}_{n=1}^{\infty}$ is called *mean-optimal* if it satisfies all of

(L1) $\sum_{x^n\in\mathcal{X}^n} 2^{-l(\varphi_n(x^n))} \leq 1$, for all $n \geq 1$;

(L2) $\limsup_{n\to\infty} \left[\frac{1}{n}E[l(\varphi_n(X^n))] - \frac{1}{n}H(X^n)\right] \leq 0$;

(L3) $\varepsilon_n = 0$, for all $n \geq 1$.

The class of FV codes given in Definition 2 was first introduced by Visweswariah, Kulkarni, and Verdú [15], though (L1) is not clearly written in [15]. Since for all $n \geq 1$, $E\left[l(\varphi_n(X^n))\right] \geq H(X^n)$ holds for all $\{(\varphi_n,\psi_n)\}_{n=1}^{\infty}$ satisfying (L1), the limit superior in (L2) can be replaced with the limit. However, we use the limit superior instead of the limit in order to facilitate comparison of the two classes of optimal FV codes given in Definitions 2 and 4.

It is easy to check that the Huffman code [7] is mean-optimal if for each $n \geq 1$, the Huffman algorithm is applied to a general source $\boldsymbol{X} = \{X^n\}_{n=1}^{\infty}$ with a finite alphabet $\mathcal{X}$. In addition, the Shannon–Fano–Elias code (see, e.g., [2]) is also mean optimal for $\boldsymbol{X}$ with a finite or a countably infinite alphabet $\mathcal{X}$.

For an FV code $\{(\varphi_n,\psi_n)\}_{n=1}^{\infty}$ and an arbitrary $\delta > 0$, define $\mathcal{W}_n(\delta)$ by

$$\mathcal{W}_n(\delta) = \left\{x^n \in \mathcal{X}^n : \left|\frac{1}{n} l(\varphi_n(x^n)) - \frac{1}{n}\log_2 \frac{1}{P_{X^n}(x^n)}\right| \leq \delta\right\}. \tag{5}$$

Then, the mean-optimal FV codes have the following property.

Theorem 2: If $\left\{\frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)}\right\}_{n=1}^{\infty}$ is uniformly integrable, i.e., if it satisfies

$$\lim_{u\to\infty} \sup_{n\geq 1} \frac{1}{n} \sum_{x^n : \frac{1}{n}\log_2 \frac{1}{P_{X^n}(x^n)} \geq u} P_{X^n}(x^n) \log_2 \frac{1}{P_{X^n}(x^n)} = 0$$

then for an arbitrary constant $\delta > 0$ any mean-optimal FV code $\{(\varphi_n,\psi_n)\}_{n=1}^{\infty}$ satisfies

$$\lim_{n\to\infty} \Pr\left\{x^n \in \mathcal{W}_n^c(\delta)\right\} = 0$$

where $\mathcal{W}_n^c(\delta)$ denotes the complement of $\mathcal{W}_n(\delta)$.

Theorem 2 immediately yields the following corollary that is proved in Appendix II.

Corollary 1: If $\left\{\frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)}\right\}_{n=1}^{\infty}$ is uniformly integrable, then any mean-optimal FV code $\{(\varphi_n,\psi_n)\}_{n=1}^{\infty}$ satisfies

$$\lim_{n\to\infty} \mathcal{L}\left(\frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)}, \frac{1}{n} l(\varphi_n(X^n))\right) = 0$$

where $\mathcal{L}(S,T)$ denotes the Lévy distance between two real-valued random variables S and T defined by

$$\mathcal{L}(S,T) = \inf\Big\{\mu : \Pr\{S \leq \lambda - \mu\} - \mu \leq \Pr\{T \leq \lambda\} \leq \Pr\{S \leq \lambda + \mu\} + \mu \text{ for all real number } \lambda\Big\}. \tag{6}$$

Remark: Corollary 1 means that the distribution of $\frac{1}{n}l(\varphi_n(X^n))$ approaches the distribution of $\frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)}$ as $n \to \infty$ in the sense of the vanishing Lévy distance. We may use the variational distance or the normalized divergence distance instead of the Lévy distance. However, we cannot have a claim corresponding to Corollary 1 under such distances because the distributions of $\frac{1}{n}l(\varphi_n(X^n))$ and $\frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)}$ are often concentrated on mutually disjoint sets. Note that $\frac{1}{n}l(\varphi_n(X^n))$ takes rational values while $\frac{1}{n}\log_2\frac{1}{P_{X^n}(X^n)}$ takes real values in general. □

Now, we define the following two quantities called the spectral sup-entropy rate and the spectral inf-entropy rate, respectively:

$$\overline{H}(\boldsymbol{X}) = \inf\left\{R : \lim_{n\to\infty}\Pr\left\{\frac{1}{n}\log_2\frac{1}{P_{X^n}(X^n)} \geq R\right\} = 0\right\} \quad (7)$$

$$\underline{H}(\boldsymbol{X}) = \sup\left\{R : \lim_{n\to\infty}\Pr\left\{\frac{1}{n}\log_2\frac{1}{P_{X^n}(X^n)} \leq R\right\} = 0\right\}. \quad (8)$$

These two quantities play key roles in coding theorems from the information-spectrum approach. In fact, $\overline{H}(\boldsymbol{X})$ has an operational meaning as the infimum achievable fixed-to-fixed length (FF) coding rate for $\boldsymbol{X}$ with the vanishing decoding error probability [4], [5]. In addition, for a given FV code $\{(\varphi_n,\psi_n)\}_{n=1}^{\infty}$, we define $\overline{L}(\boldsymbol{X})$ and $\underline{L}(\boldsymbol{X})$ as follows:

$$\overline{L}(\boldsymbol{X}) = \inf\left\{R : \lim_{n\to\infty}\Pr\left\{\frac{1}{n}l(\varphi_n(X^n)) \geq R\right\} = 0\right\}$$

$$\underline{L}(\boldsymbol{X}) = \sup\left\{R : \lim_{n\to\infty}\Pr\left\{\frac{1}{n}l(\varphi_n(X^n)) \leq R\right\} = 0\right\}.$$

Then, we obtain the following corollary as a byproduct of Theorem 2.

Corollary 2: If $\left\{\frac{1}{n}\log_2\frac{1}{P_{X^n}(X^n)}\right\}_{n=1}^{\infty}$ is uniformly integrable, then any mean-optimal FV code $\{(\varphi_n,\psi_n)\}_{n=1}^{\infty}$ satisfies

$$\overline{L}(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}) \quad (9)$$

and

$$\underline{L}(\boldsymbol{X}) = \underline{H}(\boldsymbol{X}). \quad (10)$$

Theorem 2 is proved by using the following lemma characterizing a property on the uniformly integrable random variables. The uniform integrability, which was first introduced by Han [5] in the Shannon-theoretic field, is substantially used as a sufficient condition for guaranteeing the property (11) given in Lemma 1 below. See, for example, [1, Theorem 4.5.3] for a proof of the lemma.

Lemma 1: If $\left\{\frac{1}{n}\log_2\frac{1}{P_{X^n}(X^n)}\right\}_{n=1}^{\infty}$ is uniformly integrable, then it holds that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{x^n\in\mathcal{A}_n}P_{X^n}(x^n)\log_2\frac{1}{P_{X^n}(x^n)} = 0 \quad (11)$$

for any $\{\mathcal{A}_n\}_{n=1}^{\infty}$ satisfying $\mathcal{A}_n \subset \mathcal{X}^n$ for all $n \geq 1$ and

$$\lim_{n\to\infty}\Pr\{X^n \in \mathcal{A}_n\} = 0.$$

Remark: If $|\mathcal{X}|$ is finite, the claim of Lemma 1 is immediately obtained from the following inequality:

$$\begin{aligned}\frac{1}{n}&\sum_{x^n\in\mathcal{A}_n}P_{X^n}(x^n)\log_2\frac{1}{P_{X^n}(x^n)}\\ &\leq \frac{1}{n}\Pr\{X^n\in\mathcal{A}_n\}\cdot\log_2|\mathcal{A}_n|\\ &\quad - \frac{1}{n}\Pr\{X^n\in\mathcal{A}_n\}\log_2\Pr\{X^n\in\mathcal{A}_n\}\\ &\leq \Pr\{X^n\in\mathcal{A}_n\}\cdot\log_2|\mathcal{X}| + \frac{\log_2 e}{en}\end{aligned} \quad (12)$$

(see, e.g., [12, Lemma I.6.8]), where e denotes the base of the natural logarithm. Notice that, if $|\mathcal{X}|$ is countably infinite, (12) does not imply (11). In addition, it is also known that there exists a source with a countably infinite alphabet $\mathcal{X}$ not satisfying uniform integrability (see [5], [6]). □

Proof of Theorem 2: Fix $\delta > 0$ arbitrarily and define

$$\mathcal{U}_n(\delta) = \left\{x^n \in \mathcal{X}^n : \frac{1}{n}l(\varphi_n(x^n)) + \delta < \frac{1}{n}\log_2\frac{1}{P_{X^n}(x^n)}\right\} \quad (13)$$

$$\mathcal{V}_n(\delta) = \left\{x^n \in \mathcal{X}^n : \frac{1}{n}l(\varphi_n(x^n)) - \delta > \frac{1}{n}\log_2\frac{1}{P_{X^n}(x^n)}\right\}. \quad (14)$$

Since $\mathcal{W}_n(\delta)$ in (5) is written as $\mathcal{W}_n^c(\delta) = \mathcal{U}_n(\delta)\cup\mathcal{V}_n(\delta)$, it is sufficient to prove both

$$\lim_{n\to\infty}\Pr\{X^n\in\mathcal{U}_n(\delta)\} = 0 \quad (15)$$

$$\lim_{n\to\infty}\Pr\{X^n\in\mathcal{V}_n(\delta)\} = 0. \quad (16)$$

First we show

$$\Pr\{X^n\in\mathcal{U}_n(\delta)\} < 2^{-n\delta}, \qquad \text{for all } n\geq 1 \quad (17)$$

which clearly implies (15). We note that (L1) and $\mathcal{U}_n(\delta)\subseteq\mathcal{X}^n$ yield

$$1 \geq \sum_{x^n\in\mathcal{X}^n}2^{-l(\varphi_n(x^n))} \geq \sum_{x^n\in\mathcal{U}_n(\delta)}2^{-l(\varphi_n(x^n))}. \quad (18)$$

Since we have $-l(\varphi_n(x^n)) > \log_2 P_{X^n}(x^n) + n\delta$ for all $x^n\in\mathcal{U}_n(\delta)$, it follows that

$$\begin{aligned}\sum_{x^n\in\mathcal{U}_n(\delta)}2^{-l(\varphi_n(x^n))} &> \sum_{x^n\in\mathcal{U}_n(\delta)}2^{\log_2 P_{X^n}(x^n)+n\delta}\\ &= 2^{n\delta}\cdot\Pr\{X^n\in\mathcal{U}_n(\delta)\}.\end{aligned} \quad (19)$$

Then, the combination of (18) and (19) yields (17).

Next, we prove (16) under the uniform integrability of $\left\{\frac{1}{n}\log_2\frac{1}{P_{X^n}(X^n)}\right\}_{n=1}^{\infty}$. Setting $\tau_n = \frac{\log_2 n}{n}$, (17) guarantees that

$$\Pr\{X^n\in\mathcal{U}_n(\tau_n)\} \leq \frac{1}{n} \to 0, \qquad \text{as } n\to\infty.$$

Since $\mathcal{U}_n(\tau_n)$ and $\mathcal{V}_n(\delta)$ are disjoint, we have

$$\begin{aligned}\frac{1}{n}&E[l(\varphi_n(X^n))] - \frac{1}{n}H(X^n)\\ &= \left[\sum_{x^n\in\mathcal{U}_n(\tau_n)} + \sum_{x^n\in\mathcal{V}_n(\delta)} + \sum_{x^n\in\mathcal{U}_n^c(\tau_n)\cap\mathcal{V}_n^c(\delta)}\right]\\ &\quad\cdot P_{X^n}(x^n)\left[\frac{1}{n}l(\varphi_n(x^n)) - \frac{1}{n}\log_2\frac{1}{P_{X^n}(x^n)}\right].\end{aligned} \quad (20)$$

Notice that the first sum on the right-hand side of (20) is nonpositive from the definition of $\mathcal{U}_n(\tau_n)$ and its absolute value is evaluated as

$$\begin{aligned}&\left|\frac{1}{n}\sum_{x^n\in\mathcal{U}_n(\tau_n)}P_{X^n}(x^n)\left[l(\varphi_n(x^n)) - \log_2\frac{1}{P_{X^n}(x^n)}\right]\right|\\ &\qquad\leq \frac{1}{n}\sum_{x^n\in\mathcal{U}_n(\tau_n)}P_{X^n}(x^n)\log_2\frac{1}{P_{X^n}(x^n)}.\end{aligned} \quad (21)$$

Since $\Pr\{X^n \in \mathcal{U}_n(\tau_n)\} \to 0$ as $n \to \infty$, Lemma 1 guarantees that the right-hand side of (21) tends to zero as $n \to \infty$. On the other hand, the second sum in (20) is lower-bounded as

$$\sum_{x^n \in \mathcal{V}_n(\delta)} P_{X^n}(x^n)\left[\frac{1}{n} l(\varphi_n(x^n)) - \frac{1}{n}\log_2 \frac{1}{P_{X^n}(x^n)}\right] > \delta \cdot \Pr\{X^n \in \mathcal{V}_n(\delta)\} \quad (22)$$

from the definition of $\mathcal{V}_n(\delta)$. The third sum in (20) is evaluated in a following way:

$$\begin{aligned}\sum_{x^n \in \mathcal{U}_n^c(\tau_n) \cap \mathcal{V}_n^c(\delta)} & P_{X^n}(x^n)\left[\frac{1}{n} l(\varphi_n(x^n)) - \frac{1}{n}\log_2 \frac{1}{P_{X^n}(x^n)}\right] \\ &\geq -\tau_n \cdot \Pr\{X^n \in \mathcal{U}_n^c(\tau_n) \cap \mathcal{V}_n^c(\delta)\} \\ &\geq -\tau_n \end{aligned} \quad (23)$$

where $x^n \in \mathcal{U}_n^c(\tau_n)$ is used to obtain the first inequality in (23). By substituting (21)–(23) into (20), we have

$$\frac{1}{n} E[l(\varphi_n(X^n))] - \frac{1}{n} H(X^n) > \delta \cdot \Pr\{X^n \in \mathcal{V}_n(\delta)\} + o(1) \quad (24)$$

where $o(1) \to 0$ as $n \to \infty$. Clearly, (24) implies that

$$\limsup_{n\to\infty}\left[\frac{1}{n} E[l(\varphi_n(X^n))] - \frac{1}{n} H(X^n)\right] \geq \delta \cdot \limsup_{n\to\infty} \Pr\{X^n \in \mathcal{V}_n(\delta)\} \quad (25)$$

which develops (16) since the left-hand side of (25) is assumed to be nonpositive from (L2). □

Proof of Corollary 2: Equations (9) and (10) are proved by establishing all of $\overline{H}(\boldsymbol{X}) \leq \overline{L}(\boldsymbol{X})$, $\overline{L}(\boldsymbol{X}) \leq \overline{H}(\boldsymbol{X})$, $\underline{H}(\boldsymbol{X}) \leq \underline{L}(\boldsymbol{X})$, and $\underline{L}(\boldsymbol{X}) \leq \underline{H}(\boldsymbol{X})$. However, only $\overline{H}(\boldsymbol{X}) \leq \overline{L}(\boldsymbol{X})$ is proved here since the others can be obtained similarly.

Fix an mean-optimal FV code $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ arbitrarily. Suppose that $\overline{H}(\boldsymbol{X}) \leq \overline{L}(\boldsymbol{X})$ does not hold, i.e., there exists a constant $\delta > 0$ satisfying $\overline{H}(\boldsymbol{X}) \geq \overline{L}(\boldsymbol{X}) + 3\delta$. Since the definition of $\overline{L}(\boldsymbol{X})$ implies that

$$\Pr\left\{\frac{1}{n} l(\varphi_n(X^n)) \leq \overline{L}(\boldsymbol{X}) + \delta\right\} \to 1, \qquad \text{as } n \to \infty$$

we have

$$\Pr\left\{\frac{1}{n} l(\varphi_n(X^n)) \leq \overline{H}(\boldsymbol{X}) - 2\delta\right\} \to 1, \qquad \text{as } n \to \infty.$$

Define

$$\mathcal{B}_n(\delta) = \left\{x^n \in \mathcal{X}^n : \frac{1}{n} l(\varphi_n(x^n)) \leq \overline{H}(\boldsymbol{X}) - 2\delta\right\}.$$

It is important to note that

$$\frac{1}{n}\log_2 \frac{1}{P_{X^n}(x^n)} \leq \overline{H}(\boldsymbol{X}) - \delta$$

for any $x^n \in \mathcal{B}_n(\delta) \cap \mathcal{W}_n(\delta)$, where $\mathcal{W}_n(\delta)$ is defined in (5). In addition, since $\Pr\{X^n \in \mathcal{B}_n(\delta)\} \to 1$ and $\Pr\{X^n \in \mathcal{W}_n(\delta)\} \to 1$ as $n \to \infty$, we have

$$\Pr\left\{\frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)} \leq \overline{H}(\boldsymbol{X}) - \delta\right\} \to 1, \quad \text{as } n \to \infty. \quad (26)$$

We now obtain $\overline{H}(\boldsymbol{X}) \leq \overline{L}(\boldsymbol{X})$ because (26) contradicts the definition of $\overline{H}(\boldsymbol{X})$. □

III. ε-ERROR FV CODES

In this section, FV codes satisfying $\limsup_{n\to\infty} \varepsilon_n \leq \varepsilon$ for an arbitrarily given $\varepsilon \in [0,1)$ are considered, where ε_n denotes the decoding error probability defined in (4). First, we investigate the infimum achievable FV coding rate.

Definition 3: Let $\varepsilon \in [0,1)$ be a constant arbitrarily given. A rate R is called ε-achievable if there exists a prefix-free FV codes $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ satisfying

$$\limsup_{n\to\infty} \frac{1}{n} E[l(\varphi_n(X^n)] \leq R$$

and

$$\limsup_{n\to\infty} \varepsilon_n \leq \varepsilon.$$

In particular, the infimum of the ε-achievable rate is called the infimum ε-achievable FV coding rate and is denoted by $R_{[\varepsilon]}(\boldsymbol{X})$.

Since $R_{[1]}(\boldsymbol{X}) = 0$ is trivial, we do not consider the case of $\varepsilon = 1$ throughout this section.

Han [5], [6] shows that $R_{[0]}(\boldsymbol{X})$ is expressed in the following formula:

$$R_{[0]}(\boldsymbol{X}) = H^*(\boldsymbol{X}) \stackrel{\text{def}}{=} \lim_{\gamma\downarrow 0}\limsup_{n\to\infty} \frac{1}{n} H_{[\gamma]}(X^n) \quad (27)$$

where

$$H_{[\gamma]}(X^n) = \inf_{A_n : \Pr\{X^n \in A_n\} \geq 1-\gamma} \sum_{x^n \in A_n} \frac{P_{X^n}(x^n)}{\Pr\{X^n \in A_n\}} \cdot \log_2 \frac{\Pr\{X^n \in A_n\}}{P_{X^n}(x^n)}.$$

Notice that setting $\varepsilon = 0$ means that $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ is required to have the vanishing decoding error probability. Han [5], [6] also shows that the right-hand side of (27) is equal to

$$H(\boldsymbol{X}) = \limsup_{n\to\infty} \frac{1}{n} H(X^n)$$

i.e., the infimum achievable FV coding rate if $\left\{\frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)}\right\}_{n=1}^\infty$ satisfies the uniform integrability.

We have the following theorem that can be regarded as an extension of Han's results.

Theorem 3: For any constant $\varepsilon \in [0,1)$ it holds that

$$R_{[\varepsilon]}(\boldsymbol{X}) = G_{[\varepsilon]}(\boldsymbol{X}) \stackrel{\text{def}}{=} \lim_{\gamma\downarrow 0}\limsup_{n\to\infty} \frac{1}{n} G_{[\varepsilon+\gamma]}(X^n)$$

where

$$G_{[\varepsilon]}(X^n) = \inf_{A_n : \Pr\{X^n \in A_n\} \geq 1-\varepsilon} \sum_{x^n \in A_n} P_{X^n}(x^n) \cdot \log_2 \frac{\Pr\{X^n \in A_n\}}{P_{X^n}(x^n)}.$$

Theorem 3 can be proved similarly to the development of (27). But we give the proof of Theorem 3 in Appendix I for readers' convenience.

Note that $H^*(\boldsymbol{X})$ obviously coincides with $G_{[0]}(\boldsymbol{X})$. However, the definitions of $H^*(\boldsymbol{X})$ and $H_{[\gamma]}(X^n)$ are not adequate for characterizing $R_{[\varepsilon]}(\boldsymbol{X})$ because it cannot be generalized to the case of $\varepsilon \in [0,1)$.

The following theorem gives a lower and an upper bounds of $G_{[\varepsilon]}(\boldsymbol{X})$. In fact, the upper bound in (28) corresponds to the infimum ε-achievable FF coding rate [5].

Theorem 4: For any constant $\varepsilon \in [0,1)$ it holds that

$$(1-\varepsilon)\underline{H}(\boldsymbol{X}) \leq G_{[\varepsilon]}(\boldsymbol{X}) \leq \inf\{R : F(R) \leq \varepsilon\} \quad (28)$$

where $\underline{H}(\boldsymbol{X})$ is defined in (8) and

$$F(R) = \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)} \geq R\right\}.$$

Proof: We first develop the upper bound of $G_{[\varepsilon]}(\boldsymbol{X})$ in (28). Fix an $\varepsilon \in [0,1)$ arbitrarily and set $R_0 = \inf\{R : F(R) \leq \varepsilon\}$. Then, the definition of R_0 implies that

$$F(R_0 + \gamma) \leq \varepsilon \quad (29)$$

for any $\gamma > 0$. Furthermore, (29) guarantees the existence of an integer $n_0 = n_0(\gamma)$ such that

$$\Pr\left\{\frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)} \geq R_0 + \gamma\right\} \leq \varepsilon + \gamma, \qquad \text{for all } n \geq n_0. \tag{30}$$

Now, define $A_n \in \mathcal{X}^n$ and $\Gamma_n(A_n)$ by

$$A_n = \left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log_2 \frac{1}{P_{X^n}(x^n)} < R_0 + \gamma\right\}$$

and

$$\Gamma_n(A_n) = \frac{1}{n}\sum_{x^n \in A_n} P_{X^n}(x^n)\log_2 \frac{\Pr\{X^n \in A_n\}}{P_{X^n}(x^n)} \tag{31}$$

respectively. Since (30) implies that $\Pr\{X^n \in A_n\} \geq 1 - \varepsilon - \gamma$ for all $n \geq n_0$, the definitions of $G_{[\varepsilon+\gamma]}(X^n)$ and $\Gamma_n(A_n)$ yield

$$\frac{1}{n}G_{[\varepsilon+\gamma]}(X^n) \leq \Gamma_n(A_n), \qquad \text{for all } n \geq n_0. \tag{32}$$

In addition, notice that $\Gamma_n(A_n)$ is evaluated in the following way for all $n \geq 1$:

$$\begin{aligned}\Gamma_n(A_n) &= \frac{1}{n}\Pr\{X^n \in A_n\}\log_2 \Pr\{X^n \in A_n\} \\ &\quad + \frac{1}{n}\sum_{x^n \in A_n} P_{X^n}(x^n)\log_2 \frac{1}{P_{X^n}(x^n)} \\ &< \frac{1}{n}\Pr\{X^n \in A_n\}\log_2 \Pr\{X^n \in A_n\} \\ &\quad + \Pr\{X^n \in A_n\}\cdot(R_0 + \gamma) \\ &\leq R_0 + \gamma \end{aligned} \tag{33}$$

where the inequality in (33) is obtained from the definition of A_n. Then, it follows from (32) and (33) that

$$\limsup_{n\to\infty} \frac{1}{n}G_{[\varepsilon+\gamma]}(X^n) \leq R_0 + \gamma. \tag{34}$$

Since $\gamma > 0$ is arbitrary, the upper bound in (28) is obtained by letting $\gamma \downarrow 0$ in (34).

The lower bound in (28) can be developed similarly to the proof of [5, Theorem 1.7.2]. For an arbitrary $\gamma > 0$, define

$$S_n = \left\{x^n \in \mathcal{X}^n : \frac{1}{n}\log_2 \frac{1}{P_{X^n}(x^n)} \geq \underline{H}(\boldsymbol{X}) - \gamma\right\}.$$

From the definitions of S_n and $\underline{H}(\boldsymbol{X})$ it is clear that

$$\Pr\{X^n \in S_n\} \to 1, \qquad \text{as } n \to \infty. \tag{35}$$

Let A_n be an arbitrary subset of $\mathcal{X}^n$ satisfying $\Pr\{X^n \in A_n\} \geq 1-\varepsilon-\gamma$. Then, (35) guarantees the existence of an integer $n_1 = n_1(\gamma)$ satisfying

$$\Pr\{X^n \in S_n \cap A_n\} \geq 1 - \varepsilon - 2\gamma, \qquad \text{for all } n \geq n_1. \tag{36}$$

We evaluate $\Gamma_n(A_n)$ in the following manner for all $n \geq n_1$:

$$\begin{aligned}\Gamma_n(A_n) &\overset{1)}{\geq} \frac{1}{n}\Pr\{X^n \in A_n\}\log_2 \Pr\{X^n \in A_n\} \\ &\quad + \frac{1}{n}\sum_{x^n \in A_n\cap S_n} P_{X^n}(x^n)\log_2 \frac{1}{P_{X^n}(x^n)} \\ &\overset{2)}{\geq} -\frac{\log_2 e}{en} + \frac{1}{n}\sum_{x^n \in A_n\cap S_n} P_{X^n}(x^n)\log_2 \frac{1}{P_{X^n}(x^n)} \\ &\overset{3)}{\geq} -\frac{\log_2 e}{en} + (1-\varepsilon-2\gamma)\cdot(\underline{H}(\boldsymbol{X}) - \gamma) \end{aligned} \tag{37}$$

where the marked inequalities in (37) follow from

1) $P_{X^n}(x^n)\log_2 \frac{1}{P_{X^n}(x^n)} \geq 0$, for all $x^n \in A_n \backslash S_n$;
2) $x\log_2 x \geq -\frac{\log_2 e}{e}$, for all $x \in [0,1]$;
3) the definition of S_n and (36).

Since $A_n \subseteq \mathcal{X}^n$ in (37) is an arbitrary set satisfying $\Pr\{X^n \in A_n\} \geq 1 - \varepsilon - \gamma$ and the right-hand side of (37) no longer depends on A_n, (37) leads to

$$(1-\varepsilon-2\gamma)(\underline{H}(\boldsymbol{X}) - \gamma) - \frac{\log_2 e}{en} \leq \frac{1}{n}G_{[\varepsilon+\gamma]}(X^n)$$

for all $n \geq n_1$, which implies that

$$(1-\varepsilon-2\gamma)(\underline{H}(\boldsymbol{X}) - \gamma) \leq \limsup_{n\to\infty} \frac{1}{n}G_{[\varepsilon+\gamma]}(X^n). \tag{38}$$

The lower bound of $G_{[\varepsilon]}(\boldsymbol{X})$ in (28) immediately follows from (38) by letting $\gamma \downarrow 0$. □

We can explicitly express $R_{[\varepsilon]}(\boldsymbol{X})$ for the following two sources. The proofs of (39) and (40) below are given in Appendix III.

Example 1 (i.i.d. Source): Let X^n be n i.i.d. random variables subject to a probability distribution P over a finite alphabet $\mathcal{X}$ satisfying $H(P) > 0$, where $H(P)$ denotes the entropy of base 2. Define $\boldsymbol{X} = \{X^n\}_{n=1}^\infty$. Then, for any $\varepsilon \in [0,1)$ we have

$$R_{[\varepsilon]}(\boldsymbol{X}) = (1-\varepsilon)H(P). \tag{39}$$

Example 2 (Mixed Source): Let X_1^n and X_2^n be n i.i.d. random variables subject to probability distributions P_1 and P_2 over a finite alphabet $\mathcal{X}$, respectively. Assume that $H(P_1) < H(P_2)$. Let X^n be n random variables satisfying

$$P_{X^n}(x^n) = \alpha_1 P_{X_1^n}(x^n) + \alpha_2 P_{X_2^n}(x^n)$$

for some α_1 and α_2 satisfying $\alpha_1 > 0$, $\alpha_2 > 0$, and $\alpha_1 + \alpha_2 = 1$. Define $\boldsymbol{X} = \{X^n\}_{n=1}^\infty$. Then, for any $\varepsilon \in [0,1)$, we have

$$R_{[\varepsilon]}(\boldsymbol{X}) = \begin{cases} \alpha_1 H(P_1) + (\alpha_2 - \varepsilon)H(P_2), & \text{if } 0 \leq \varepsilon < \alpha_2 \\ (1-\varepsilon)H(P_1), & \text{otherwise.}\end{cases} \tag{40}$$

In this section, as well we are interested in asymptotic behavior of

$$\frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)} \quad \text{and} \quad \frac{1}{n}l(\varphi_n(X^n))$$

for a class of FV codes $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ achieving $R_{[\varepsilon]}(\boldsymbol{X})$ for an arbitrarily given $\varepsilon \in [0,1)$. We introduce the following class of FV codes.

Definition 4: Let $\varepsilon \in [0,1)$ be an arbitrary constant. An FV code $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ is called ε-mean-optimal if it satisfies all of

(E1) $\sum_{y^* \in \mathcal{C}_n} 2^{-l(y^*)} \leq 1$, for all $n \geq 1$;

(E2) $\lim_{\gamma\downarrow 0}\limsup_{n\to\infty} \left[\frac{1}{n}E[l(\varphi_n(x^n))] - \frac{1}{n}G_{[\varepsilon+\gamma]}(X^n)\right] \leq 0$;

(E3) $\limsup_{n\to\infty} \varepsilon_n \leq \varepsilon$;

where $\mathcal{C}_n$ denotes the range of φ_n.

Remark: The meaning of condition (E2) in Definition 4 can be understood from the following argument. If we consider FV codes satisfying (E3), for any constant $\gamma > 0$ it holds that $\varepsilon_n \leq \varepsilon + \gamma$ for all sufficiently large n. Then, Lemma 2 in Appendix I tells us that for any constant $\delta > 0$ there exists an FV code satisfying

$$\frac{1}{n}E[l(\varphi_n(X^n))] \leq \frac{1}{n}G_{[\varepsilon+\gamma]}(X^n) + \delta \tag{41}$$

for all sufficiently large n and $\varepsilon_n \le \varepsilon + \gamma$ for all $n \ge 1$. Since (41) leads to

$$\limsup_{n\to\infty}\left[\frac{1}{n}E\left[l\left(\varphi_n(X^n)\right)\right] - \frac{1}{n}G_{[\varepsilon+\gamma]}(X^n)\right] \le 0$$

and $\gamma > 0$ can be made arbitrarily small, we have condition (E2) by letting $\gamma \downarrow 0$.

It is also important to note that any FV code satisfying (E2) achieves $G_{[\varepsilon]}(\boldsymbol{X})$ in Theorem 3. In fact, it is easy to verify that

$$\begin{aligned} 0 &\ge \lim_{\gamma\downarrow 0}\limsup_{n\to\infty}\left\{\frac{1}{n}E\left[l\left(\varphi_n(X^n)\right)\right] - \frac{1}{n}G_{[\varepsilon+\gamma]}(X^n)\right\} \\ &\ge \lim_{\gamma\downarrow 0}\left[\limsup_{n\to\infty}\frac{1}{n}E\left[l\left(\varphi_n(X^n)\right)\right] - \limsup_{n\to\infty}\frac{1}{n}G_{[\varepsilon+\gamma]}(X^n)\right] \\ &= \limsup_{n\to\infty}\frac{1}{n}E\left[l\left(\varphi_n(X^n)\right)\right] - \lim_{\gamma\downarrow 0}\limsup_{n\to\infty}\frac{1}{n}G_{[\varepsilon+\gamma]}(X^n) \end{aligned}$$

which implies that

$$\limsup_{n\to\infty}\frac{1}{n}E\left[l\left(\varphi_n(X^n)\right)\right] = G_{[\varepsilon]}(\boldsymbol{X})$$

i.e., $G_{[\varepsilon]}(\boldsymbol{X})$ is achievable. □

The following theorem characterizes a relationship between

$$\frac{1}{n}\log_2\frac{1}{P_{X^n}(X^n)} \quad \text{and} \quad \frac{1}{n}l\left(\varphi_n(X^n)\right)$$

of an ε-mean-optimal FV code.

Theorem 5: Let $\varepsilon \in [0,1)$ be an arbitrary constant and $\{(\varphi_n,\psi_n)\}_{n=1}^{\infty}$ be any ε-mean-optimal FV code. Define D_n by

$$D_n = \left\{x^n \in \mathcal{X}^n : \psi_n\left(\varphi_n(x^n)\right) = x^n\right\}. \tag{42}$$

Then, for any $\delta > 0$ it holds that

$$\lim_{n\to\infty}\Pr\left\{x^n \in \mathcal{W}_n^c(\delta)\cap D_n\right\} = 0$$

where $\mathcal{W}_n(\delta)$ is the set defined in (5).

Remark: Notice that Theorem 5 holds without any assumption on the source such as the uniform integrability of $\left\{\frac{1}{n}\log_2\frac{1}{P_{X^n}(X^n)}\right\}_{n=1}^{\infty}$. However, this does not mean that Theorem 5 includes Theorem 2 as a special case of $\varepsilon = 0$. While we are interested in FV codes with $\varepsilon_n = 0$ for all $n \ge 1$ in Theorem 2, FV codes treated in Theorem 5 with $\varepsilon = 0$ satisfy only $\varepsilon_n \to 0$ as $n \to \infty$. It is also important to notice that the claim of Theorem 2 is stronger than the claim of Theorem 5. While Theorem 2 considers codeword lengths for all sequences $x^n \in \mathcal{X}^n$, Theorem 5 only treats codeword lengths for sequences belonging to D_n. □

Theorem 5 yields the following corollary that is proved in Appendix II.

Corollary 3: Let $\varepsilon \in [0,1)$ be an arbitrary constant. Then, any ε-mean-optimal FV code $\{(\varphi_n,\psi_n)\}_{n=1}^{\infty}$ satisfies

$$\limsup_{n\to\infty}\mathcal{L}\left(\frac{1}{n}\log_2\frac{1}{P_{X^n}(X^n)}, \frac{1}{n}l\left(\varphi_n(X^n)\right)\right) \le \varepsilon.$$

Proof of Theorem 5: For proving the theorem it suffices to prove both

$$\lim_{n\to\infty}\Pr\left\{X^n \in \mathcal{U}_n(\delta)\cap D_n\right\} = 0 \tag{43}$$

$$\lim_{n\to\infty}\Pr\left\{X^n \in \mathcal{V}_n(\delta)\cap D_n\right\} = 0 \tag{44}$$

for an arbitrarily fixed $\delta > 0$, where $\mathcal{U}_n(\delta)$ and $\mathcal{V}_n(\delta)$ are the sets defined in (13) and (14), respectively. We first prove (43). It is important to notice that φ_n is one-to-one in D_n and $\{\varphi_n(x^n) : x^n \in D_n\}$ satisfies the prefix condition. Therefore, it follows that

$$\begin{aligned} 1 &\ge \sum_{x^n\in D_n} 2^{-l(\varphi_n(x^n))} \\ &\ge \sum_{x^n\in \mathcal{U}_n(\delta)\cap D_n} 2^{-l(\varphi_n(x^n))} \\ &> 2^{n\delta}\cdot\Pr\left\{X^n \in \mathcal{U}_n(\delta)\cap D_n\right\}, \qquad \text{for all } n \ge 1 \end{aligned} \tag{45}$$

where the last inequality follows from the definition of $\mathcal{U}_n(\delta)$. Dividing both sides of (45) by $2^{n\delta}$ establishes (43).

In order to prove (44) we partition $\mathcal{X}^n$ into four disjoint subsets, $D_n^c, \mathcal{U}_n(\tau_n)\cap D_n, \mathcal{V}_n(\delta)\cap D_n$, and $\mathcal{U}_n^c(\tau_n)\cap\mathcal{V}_n^c(\delta)\cap D_n$, where $\tau_n = \frac{\log_2 n}{n}$. By using these four subsets $\frac{1}{n}E\left[l\left(\varphi_n(X^n)\right)\right]$ is evaluated in the following way:

$$\begin{aligned} \frac{1}{n}E\left[l\left(\varphi_n(X^n)\right)\right] &= \frac{1}{n}\left[\sum_{x^n\in D_n^c} + \sum_{x^n\in\mathcal{U}_n(\tau_n)\cap D_n} + \sum_{x^n\in\mathcal{V}_n(\delta)\cap D_n} + \sum_{x^n\in\mathcal{U}_n^c(\tau_n)\cap\mathcal{V}_n^c(\delta)\cap D_n}\right] P_{X^n}(x^n)l\left(\varphi_n(x^n)\right) \\ &\ge \frac{1}{n}\sum_{x^n\in\mathcal{V}_n(\delta)\cap D_n} P_{X^n}(x^n)l\left(\varphi_n(x^n)\right) \\ &\quad + \frac{1}{n}\sum_{x^n\in\mathcal{U}_n^c(\tau_n)\cap\mathcal{V}_n^c(\delta)\cap D_n} P_{X^n}(x^n)l\left(\varphi_n(x^n)\right). \end{aligned} \tag{46}$$

From the definition of $\mathcal{V}_n(\delta)$, the first sum in (46) is bounded as follows:

$$\begin{aligned} &\frac{1}{n}\sum_{x^n\in\mathcal{V}_n(\delta)\cap D_n} P_{X^n}(x^n)l\left(\varphi_n(x^n)\right) \\ &\quad > \sum_{x^n\in\mathcal{V}_n(\delta)\cap D_n} P_{X^n}(x^n)\left[\frac{1}{n}\log_2\frac{1}{P_{X^n}(x^n)} + \delta\right] \\ &\quad = \frac{1}{n}\sum_{x^n\in\mathcal{V}_n(\delta)\cap D_n} P_{X^n}(x^n)\log_2\frac{1}{P_{X^n}(x^n)} \\ &\qquad + \delta\cdot\Pr\left\{X^n \in \mathcal{V}_n(\delta)\cap D_n\right\}. \end{aligned} \tag{47}$$

The second term in (46) is evaluated as

$$\begin{aligned} &\frac{1}{n}\sum_{x^n\in\mathcal{U}_n^c(\tau_n)\cap\mathcal{V}_n^c(\delta)\cap D_n} P_{X^n}(x^n)l\left(\varphi_n(x^n)\right) \\ &\quad \ge \frac{1}{n}\sum_{x^n\in\mathcal{U}_n^c(\tau_n)\cap\mathcal{V}_n^c(\delta)\cap D_n} P_{X^n}(x^n)\log_2\frac{1}{P_{X^n}(x^n)} - \tau_n \end{aligned} \tag{48}$$

where $x^n \in \mathcal{U}_n^c(\tau_n)$ is used to obtain the inequality. By substituting (47) and (48) into (46) and noticing that $\mathcal{U}_n(\tau_n)\cap\mathcal{V}_n(\delta) = \phi$, we have the following lower bound on the expected codeword length:

$$\begin{aligned} \frac{1}{n}E\left[l\left(\varphi_n(X^n)\right)\right] &> \frac{1}{n}\sum_{x^n\in\mathcal{U}_n^c(\tau_n)\cap D_n} P_{X^n}(x^n)\log_2\frac{1}{P_{X^n}(x^n)} \\ &\quad + \delta\cdot\Pr\left\{X^n \in \mathcal{V}_n(\delta)\cap D_n\right\} - \tau_n. \end{aligned} \tag{49}$$

We notice here that

$$\begin{aligned} &\Pr\left\{X^n \in \mathcal{U}_n^c(\tau_n)\cap D_n\right\} \\ &\quad = \Pr\left\{X^n \in D_n\right\} - \Pr\left\{X^n \in \mathcal{U}_n(\tau_n)\cap D_n\right\}. \end{aligned}$$

Since (45) guarantees that $\Pr\{X^n \in \mathcal{U}_n(\tau_n) \cap D_n\} \to 0$ as $n \to \infty$, (E3) and the definition of D_n guarantee that for any $\gamma > 0$ there exists an integer $n_1 = n_1(\gamma)$ satisfying

$$\Pr\{X^n \in \mathcal{U}_n^c(\tau_n) \cap D_n\} \geq 1 - \varepsilon - 2\gamma, \qquad \text{for all } n \geq n_1.$$

Therefore, the first term on the right-hand side of (49) is evaluated in the following way:

$$\begin{aligned} &\frac{1}{n} \sum_{x^n \in \mathcal{U}_n^c(\tau_n) \cap D_n} P_{X^n}(x^n) \log_2 \frac{1}{P_{X^n}(x^n)} \\ &\quad \geq \frac{1}{n} \sum_{x^n \in \mathcal{U}_n^c(\tau_n) \cap D_n} P_{X^n}(x^n) \cdot \log_2 \frac{\Pr\{X^n \in \mathcal{U}_n^c(\tau_n) \cap D_n\}}{P_{X^n}(x^n)} \\ &\quad \geq \frac{1}{n} G_{[\varepsilon+2\gamma]}(X^n), \qquad \text{for all } n \geq n_1 \end{aligned} \tag{50}$$

where the second inequality in (50) follows from the definition of $G_{[\varepsilon+2\gamma]}(X^n)$. Then, (49) and (50) lead to

$$\frac{1}{n} E[l(\varphi_n(X^n))] - \frac{1}{n} G_{[\varepsilon+2\gamma]}(X^n) > \delta \cdot \Pr\{X^n \in \mathcal{V}_n(\delta) \cap D_n\} - \tau_n$$

which implies that

$$\limsup_{n\to\infty} \left[\frac{1}{n} E[l(\varphi_n(X^n))] - \frac{1}{n} G_{[\varepsilon+2\gamma]}(X^n)\right] \geq \delta \cdot \limsup_{n\to\infty} \Pr\{X^n \in \mathcal{V}_n(\delta) \cap D_n\}.$$

Since $\gamma > 0$ is arbitrarily, it follows from (E2) that

$$\begin{aligned} 0 &\geq \lim_{\gamma \downarrow 0} \limsup_{n\to\infty} \left[\frac{1}{n} E[l(\varphi_n(X^n))] - \frac{1}{n} G_{[\varepsilon+2\gamma]}(X^n)\right] \\ &\geq \delta \cdot \limsup_{n\to\infty} \Pr\{X^n \in \mathcal{V}_n(\delta) \cap D_n\}. \end{aligned} \tag{51}$$

Now, (44) is immediate from (51). □

APPENDIX I
PROOF OF THEOREM 3

We first give the following lemma which is valid for all $n \geq 1$ and $\varepsilon \in [0, 1)$.

Lemma 2: Let $\varepsilon \in [0, 1)$ be an arbitrarily given constant. Then, for any FV code $\{(\varphi_n, \psi_n)\}_{n=1}^{\infty}$ with $\varepsilon_n \leq \varepsilon$ for all $n \geq 1$ satisfies

$$E[l(\varphi_n(X^n))] \geq G_{[\varepsilon]}(X^n) \tag{52}$$

for all $n \geq 1$. In addition, there exists an FV code $\{(\varphi_n, \psi_n)\}_{n=1}^{\infty}$ satisfying $\varepsilon_n \leq \varepsilon$ for all $n \geq 1$ and

$$\frac{1}{n} E[l(\varphi_n(X^n))] \leq \frac{1}{n} G_{[\varepsilon]}(X^n) + \delta \tag{53}$$

for all sufficiently large n, where $\delta > 0$ is an arbitrarily given constant.

Proof: We first prove that any FV code with $\varepsilon_n \leq \varepsilon$ for all $n \geq 1$ satisfies (52). Fix an FV-code $\{(\varphi_n, \psi_n)\}_{n=1}^{\infty}$ satisfying $\varepsilon_n \leq \varepsilon$ for all $n \geq 1$ arbitrarily. Define D_n by (42). Then, it clearly holds that $\Pr\{X^n \in D_n\} \geq 1 - \varepsilon$ for all $n \geq 1$. We evaluate $E[l(\varphi_n(X^n))]$ in the following way:

$$\begin{aligned} E[l(\varphi_n(X^n))] &\geq \sum_{x^n \in D_n} P_{X^n}(x^n) l(\varphi_n(x^n)) \\ &= \Pr\{X^n \in D_n\} E[l(\varphi_n(X^n)) | X^n \in D_n] \end{aligned} \tag{54}$$

where

$$E[l(\varphi_n(X^n)) | X^n \in D_n] = \sum_{x^n \in D_n} \frac{P_{X^n}(x^n)}{\Pr\{X^n \in D_n\}} l(\varphi_n(x^n)).$$

Since $P_{X^n}(x^n)/\Pr\{X^n \in D_n\}$ is a probability distribution over D_n and $\{\varphi_n(x^n) : x^n \in D_n\}$ satisfies the prefix condition, it holds that

$$E[l(\varphi_n(X^n)) | X^n \in D_n] \geq \sum_{x^n \in D_n} \frac{P_{X^n}(x^n)}{\Pr\{X^n \in D_n\}} \log_2 \frac{\Pr\{X^n \in D_n\}}{P_{X^n}(x^n)}. \tag{55}$$

The combination of (54) and (55) leads to

$$\begin{aligned} E[l(\varphi_n(X^n))] &\geq \sum_{x^n \in D_n} P_{X^n}(x^n) \log_2 \frac{\Pr\{X^n \in D_n\}}{P_{X^n}(x^n)} \\ &\geq G_{[\varepsilon]}(X^n), \end{aligned}$$

where the last inequality follows from the definition of $G_{[\varepsilon]}(X^n)$ and $\Pr\{X^n \in D_n\} \geq 1 - \varepsilon$.

Next, we prove the existence of an FV code satisfying (53). We first notice that for any $\gamma > 0$, there exists a set $A_n \subseteq \mathcal{X}^n$ satisfying $\Pr\{X^n \in A_n\} \geq 1 - \varepsilon$ and

$$\sum_{x^n \in A_n} P_{X^n}(x^n) \log_2 \frac{\Pr\{X^n \in A_n\}}{P_{X^n}(x^n)} \leq G_{[\varepsilon]}(X^n) + \gamma. \tag{56}$$

Without loss of generality, we can assume that $P_{X^n}(x^n) > 0$ for all $x^n \in A_n$. Since $P_{X^n}(x^n)/\Pr\{X^n \in A_n\}$ gives a probability distribution over A_n, we can construct the Shannon–Fano–Elias code $\tilde{\varphi}_n : A_n \to \{0, 1\}^*$. We define an encoder $\varphi_n : \mathcal{X}^n \to \{0, 1\}^*$ by

$$\varphi_n(x^n) = \begin{cases} 0\tilde{\varphi}_n(x^n), & \text{if } x^n \in A_n \\ 1, & \text{otherwise} \end{cases} \tag{57}$$

for each $n \geq 1$. Note that $\{\varphi_n(x^n) : \varphi_n(x^n) \in \mathcal{X}^n\}$ satisfies the prefix condition because the Shannon–Fano–Elias code satisfies the prefix condition. We define a decoder ψ_n in such a way that every $x^n \in A_n$ is correctly decodable. Clearly, we have $\varepsilon_n \leq \Pr\{X^n \notin A_n\} \leq \varepsilon$ for all $n \geq 1$. On the other hand, $\frac{1}{n} E[l(\varphi_n(X^n))]$ is evaluated in the following way:

$$\begin{aligned} \frac{1}{n} E[l(\varphi_n(X^n))] &= \frac{1}{n} \sum_{x^n \in A_n} P_{X^n}(x^n) l(\varphi_n(x^n)) + \frac{1}{n} \sum_{x^n \in A_n^c} P_{X^n}(x^n) l(\varphi_n(x^n)) \\ &\overset{1)}{<} \frac{1}{n} \sum_{x^n \in A_n} P_{X^n}(x^n) \left(\log_2 \frac{\Pr\{X^n \in A_n\}}{P_{X^n}(x^n)} + 3\right) + \frac{1}{n} \sum_{x^n \in A_n^c} P_{X^n}(x^n) \cdot 1 \\ &\leq \frac{1}{n} \sum_{x^n \in A_n} P_{X^n}(x^n) \log_2 \frac{\Pr\{X^n \in A_n\}}{P_{X^n}(x^n)} + \frac{3}{n} \\ &\overset{2)}{\leq} \frac{1}{n} G_{[\varepsilon]}(X^n) + \frac{3 + \gamma}{n} \\ &\leq \frac{1}{n} G_{[\varepsilon]}(X^n) + \delta, \qquad \text{for all sufficiently } n \end{aligned}$$

where the marked inequalities follow from

1) $l(\tilde{\varphi}_n(x^n)) < \log_2 \frac{\Pr\{X^n \in A_n\}}{P_{X^n}(x^n)} + 2$ [2] and (57);
2) Equation (56);

which completes the proof of the lemma. □

Proof of Theorem 3: First, we establish the converse part of Theorem 3. To this end, let $\{(\varphi_n, \psi_n)\}_{n=1}^{\infty}$ be an arbitrarily fixed FV code satisfying $\limsup_{n\to\infty} \varepsilon_n \le \varepsilon$. Clearly, for any $\gamma > 0$ there exists an integer $n_0 = n_0(\gamma)$ such that $\varepsilon_n \le \varepsilon + \gamma$ for all $n \ge n_0$. Then, the first claim of Lemma 2 tells us that

$$E\left[l\left(\varphi_n(X^n)\right)\right] \ge G_{[\varepsilon+\gamma]}(X^n), \qquad \text{for all } n \ge n_0$$

which leads to

$$\limsup_{n\to\infty} \frac{1}{n} E\left[l\left(\varphi_n(X^n)\right)\right] \ge \limsup_{n\to\infty} \frac{1}{n} G_{[\varepsilon+\gamma]}(X^n).$$

Since $\gamma > 0$ is arbitrary, we have

$$\limsup_{n\to\infty} \frac{1}{n} E\left[l\left(\varphi_n(X^n)\right)\right] \ge \lim_{\gamma\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} G_{[\varepsilon+\gamma]}(X^n).$$

Next, we establish the direct part. Let $\delta > 0$ be an arbitrary constant and $\{\mu_i\}_{i=0}^{\infty}$ an arbitrary sequence satisfying

$$1 - \varepsilon = \mu_0 > \mu_1 > \cdots > \mu_n > \cdots > 0$$

and $\mu_n \to 0$ as $n \to \infty$. Then, the second claim of Lemma 2 tells us that for each $i \ge 1$ there exists an FV code $\{(\varphi_{\mu_i,n}, \psi_{\mu_i,n})\}_{n=1}^{\infty}$ satisfying

$$\Pr\left\{\psi_{\mu_i,n}\left(\varphi_{\mu_i,n}(X^n)\right) \ne X^n\right\} \le \varepsilon + \mu_i, \qquad \text{for all } n \ge 1 \quad (58)$$

and

$$\frac{1}{n} E\left[l\left(\varphi_{\mu_i,n}(X^n)\right)\right] \le \frac{1}{n} G_{[\varepsilon+\mu_i]}(X^n) + \delta, \quad \text{for all } n \ge n_i \quad (59)$$

where n_i is an integer dependent on δ. We define a sequence $\{N_i\}_{i=0}^{\infty}$ by $N_0 = 1$ and $N_i = \max\{N_{i-1}+1, n_i\}$ for $i \ge 1$. Notice that $\{N_i\}_{i=0}^{\infty}$ is monotone increasing and $N_i \ge n_i$ for each $i \ge 1$. By using $\{N_i\}_{i=0}^{\infty}$, we define $\varphi_n = \varphi_{\mu_i,n}$ and $\psi_n = \psi_{\mu_i,n}$ for all $N_i \le n < N_{i+1}$. For n satisfying $N_0 \le n < N_1$ φ_n and ψ_n can be defined arbitrarily. Then, it follows from (58) and (59) that

$$\varepsilon_n = \Pr\left\{\psi_n\left(\varphi_n(X^n)\right) \ne X^n\right\} \le \varepsilon + \mu_i \quad (60)$$

and

$$\frac{1}{n} E\left[l\left(\varphi_n(X^n)\right)\right] \le \frac{1}{n} G_{[\varepsilon+\mu_i]}(X^n) + \delta \quad (61)$$

for all $N_i \le n < N_{i+1}$. If we define $i_n = \max\{i : n \ge N_i\}$ for each $n \ge 1$, (60) and (61) can be written as

$$\varepsilon_n \le \varepsilon + \mu_{i_n} \quad (62)$$

and

$$\frac{1}{n} E\left[l\left(\varphi_n(X^n)\right)\right] \le \frac{1}{n} G_{[\varepsilon+\mu_{i_n}]}(X^n) + \delta \quad (63)$$

respectively. Clearly, both (62) and (63) are valid for all $n \ge 1$. We notice here that $i_n \to \infty$ as $n \to \infty$ and, therefore, $\mu_{i_n} \to 0$ as $n \to \infty$ from the definitions of $\{N_i\}_{i=0}^{\infty}$ and $\{\mu_i\}_{i=0}^{\infty}$. Then, it follows from (62) and (63) that $\{(\varphi_n, \psi_n)\}_{n=1}^{\infty}$ satisfies

$$\limsup_{n\to\infty} \varepsilon_n \le \varepsilon$$

and

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} E\left[l\left(\varphi_n(X^n)\right)\right] &\le \limsup_{n\to\infty} \frac{1}{n} G_{[\varepsilon+\mu_{i_n}]}(X^n) + \delta \\ &= \lim_{\gamma\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} G_{[\varepsilon+\gamma]}(X^n) + \delta \quad (64)\end{aligned}$$

where the equality in (64) holds because $\limsup_{n\to\infty} \frac{1}{n} G_{[\varepsilon]}(X^n)$ is monotone nonincreasing in ε. In fact, for any fixed $\gamma > 0$, there exists an integer n_0 such that $\mu_{i_n} \le \gamma$ for all $n \ge n_0$ and therefore,

$$\frac{1}{n} G_{[\varepsilon+\gamma]}(X^n) \le \frac{1}{n} G_{[\varepsilon+\mu_{i_n}]}(X^n), \qquad \text{for all } n \ge n_0.$$

This leads to

$$\lim_{\gamma\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} G_{[\varepsilon+\gamma]}(X^n) \le \limsup_{n\to\infty} \frac{1}{n} G_{[\varepsilon+\mu_{i_n}]}(X^n)$$

by letting $n \to \infty$ and then $\gamma \downarrow 0$. On the other hand, for an integer M define $\tilde{\mu}_n = \max\{\mu_{i_M}, \mu_{i_n}\}$. Since $\tilde{\mu}_{i_n} = \mu_{i_n}$ for all $1 \le n \le M$ and $\tilde{\mu}_{i_n} = \mu_{i_M}$ for all $n > M$, we have

$$\limsup_{n\to\infty} \frac{1}{n} G_{[\varepsilon+\tilde{\mu}_{i_n}]}(X^n) \le \lim_{\gamma\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} G_{[\varepsilon+\gamma]}(X^n)$$

which leads to

$$\limsup_{n\to\infty} \frac{1}{n} G_{[\varepsilon+\mu_{i_n}]}(X^n) \le \lim_{\gamma\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} G_{[\varepsilon+\gamma]}(X^n)$$

by letting $M \to \infty$. Now the proof of the direct part of Theorem 3 is completed since δ in (64) is arbitrary.

Appendix II
Proofs of Corollaries 1 and 3

We may use a known result (see, e.g., [1, Sec. 4.4, Problem 8]) for proving Corollaries 1 and 3. However, we give proofs of the two corollaries in order to make this paper self-contained.

For a real number $\lambda \in \boldsymbol{R}$ define $E_n(\lambda)$ and $F_n(\lambda)$ as follows:

$$E_n(\lambda) = \left\{ x^n \in \mathcal{X}^n : \frac{1}{n} \log_2 \frac{1}{P_{X^n}(x^n)} \le \lambda \right\}$$

$$F_n(\lambda) = \left\{ x^n \in \mathcal{X}^n : \frac{1}{n} l\left(\varphi_n(x^n)\right) \le \lambda \right\}.$$

The following lemma plays a key role.

Lemma 3: For an arbitrary constant $\delta > 0$ define $\mathcal{W}_n(\delta)$ by (5). Then, for any $\lambda > 0$ it holds that

$$\begin{aligned}&\Pr\left\{X^n \in E_n(\lambda-\delta)\right\} - \Pr\left\{X^n \in \mathcal{W}_n^c(\delta)\right\} \\ &\quad \le \Pr\left\{X^n \in F_n(\delta)\right\} \\ &\quad \le \Pr\left\{X^n \in E_n(\lambda+\delta)\right\} + \Pr\left\{X^n \in \mathcal{W}_n^c(\delta)\right\}. \quad (65)\end{aligned}$$

Proof: We first note that the definitions of $\mathcal{W}_n(\delta)$, $E_n(\lambda)$, and $F_n(\lambda)$ imply that

$$\begin{aligned}\left(\mathcal{W}_n(\delta) \cap E_n(\lambda-\delta)\right) &\subseteq \left(\mathcal{W}_n(\delta) \cap F_n(\lambda)\right) \\ &\subseteq \left(\mathcal{W}_n(\delta) \cap E_n(\lambda+\delta)\right). \quad (66)\end{aligned}$$

Then, the upper bound of $\Pr\left\{X^n \in F_n(\delta)\right\}$ in (65) can be developed in the following way:

$$\begin{aligned}\Pr\left\{X^n \in F_n(\lambda)\right\} &= \Pr\left\{X^n \in F_n(\lambda) \cap \mathcal{W}_n(\delta)\right\} \\ &\quad + \Pr\left\{X^n \in F_n(\lambda) \cap \mathcal{W}_n^c(\delta)\right\} \\ &\le \Pr\left\{X^n \in E_n(\lambda+\delta) \cap \mathcal{W}_n(\delta)\right\} \\ &\quad + \Pr\left\{X^n \in F_n(\lambda) \cap \mathcal{W}_n^c(\delta)\right\} \\ &\le \Pr\left\{X^n \in E_n(\lambda+\delta)\right\} \\ &\quad + \Pr\left\{X^n \in \mathcal{W}_n^c(\delta)\right\}, \quad (67)\end{aligned}$$

where the first inequality in (67) follows from (66). On the other hand, the lower bound of $\Pr\{X^n \in F_n(\delta)\}$ in (65) is developed in the following manner:

$$\begin{aligned}\Pr\{X^n \in F_n(\lambda)\} &= \Pr\{X^n \in F_n(\lambda) \cap \mathcal{W}_n(\delta)\} \\ &\quad + \Pr\{X^n \in F_n(\lambda) \cap \mathcal{W}_n^c(\delta)\} \\ &\geq \Pr\{X^n \in E_n(\lambda-\delta) \cap \mathcal{W}_n(\delta)\} \\ &\quad + \Pr\{X^n \in F_n(\lambda) \cap \mathcal{W}_n^c(\delta)\} \\ &\geq \Pr\{X^n \in E_n(\lambda-\delta) \cap \mathcal{W}_n(\delta)\} \\ &\geq \Pr\{X^n \in E_n(\lambda-\delta)\} \\ &\quad - \Pr\{X^n \in \mathcal{W}_n^c(\delta)\} \end{aligned} \quad (68)$$

where the first inequality in (68) follows from (66). Now the claim of this lemma is immediate from the combination of (67) and (68). □

Proof of Corollary 1: Theorem 2 guarantees that for any $\delta > 0$ there exists an integer $n_0 = n_0(\delta)$ such that

$$\Pr\{X^n \in \mathcal{W}_n^c(\delta)\} \leq \delta, \qquad \text{for all } n \geq n_0.$$

Thus, it follows from Lemma 3 that

$$\begin{aligned}\Pr\{X^n \in E_n(\lambda-\delta)\} - \delta &\leq \Pr\{X^n \in F_n(\lambda)\} \\ &\leq \Pr\{X^n \in E_n(\lambda+\delta)\} + \delta \\ &\qquad \text{for all } n \geq n_0 \end{aligned} \quad (69)$$

which, together with the definition of $\mathcal{L}(\cdot,\cdot)$ in (6), means that

$$\mathcal{L}\left(\frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)}, \frac{1}{n} l(\varphi_n(X^n))\right) \leq \delta, \qquad \text{for all } n \geq n_0.$$

Since $\delta > 0$ in (69) is arbitrary, the claim of Corollary 1 is established. □

Proof of Corollary 3: Let D_n be the set defined by (42). It is important to notice that we can prove

$$\begin{aligned}&\Pr\{X^n \in E_n(\lambda-\delta)\} - \Pr\{X^n \in (\mathcal{W}_n(\delta) \cap D_n)^c\} \\ &\quad \leq \Pr\{X^n \in F_n(\delta)\} \\ &\quad \leq \Pr\{X^n \in E_n(\lambda+\delta)\} + \Pr\{X^n \in (\mathcal{W}_n(\delta) \cap D_n)^c\} \end{aligned} \quad (70)$$

by using the same argument in the proof of Lemma 3 and replacing $\mathcal{W}_n(\delta)$ with $\mathcal{W}_n(\delta) \cap D_n$. We now use Theorem 5 for evaluating $\Pr\{X^n \in (\mathcal{W}_n(\delta) \cap D_n)^c\}$. Since it holds that

$$\begin{aligned}&\Pr\{X^n \in (\mathcal{W}_n(\delta) \cap D_n)^c\} \\ &\qquad = \Pr\{X^n \in \mathcal{W}_n^c(\delta) \cap D_n\} + \Pr\{X^n \in D_n^c\} \end{aligned} \quad (71)$$

and $\{(\varphi_n, \psi_n)\}_{n=1}^{\infty}$ is assumed to be ε-mean-optimal, Theorem 5 guarantees that for any $\delta > 0$ there exists an integer $n_1 = n_1(\delta)$ satisfying both

$$\Pr\{X^n \in \mathcal{W}_n^c(\delta) \cap D_n\} \leq \delta/2, \qquad \text{for all } n \geq n_1 \quad (72)$$

and

$$\Pr\{X^n \in D_n\} \leq \varepsilon + \delta/2, \qquad \text{for all } n \geq n_1. \quad (73)$$

Hence, by combining (71)–(73) we have

$$\Pr\{X^n \in (\mathcal{W}_n(\delta) \cap D_n)^c\} \leq \varepsilon + \delta, \qquad \text{for all } n \geq n_1. \quad (74)$$

In addition, notice that we have

$$\Pr\{X^n \in E_n(\lambda-\delta)\} \geq \Pr\{X^n \in E_n(\lambda-\varepsilon-\delta)\} \quad (75)$$

$$\Pr\{X^n \in E_n(\lambda+\delta)\} \leq \Pr\{X^n \in E_n(\lambda+\varepsilon+\delta)\} \quad (76)$$

from the definition of $E_n(\lambda)$. Then, the substitution of (74)–(76) into (70) yields

$$\begin{aligned}&\Pr\{X^n \in E_n(\lambda-\varepsilon-\delta)\} - \varepsilon - \delta \\ &\quad \leq \Pr\{X^n \in F_n(\delta)\} \\ &\quad \leq \Pr\{X^n \in E_n(\lambda+\varepsilon+\delta)\} + \varepsilon + \delta, \qquad \text{for all } n \geq n_1 \end{aligned} \quad (77)$$

which means that

$$\mathcal{L}\left(\frac{1}{n}\log_2 \frac{1}{P_{X^n}(X^n)}, \frac{1}{n} l(\varphi_n(X^n))\right) \leq \varepsilon + \delta, \quad \text{for all } n \geq n_1.$$

Since $\delta > 0$ in (77) is arbitrary, the claim of Corollary 3 is established. □

Appendix III
Proofs of (39) and (40)

Proof of (39): Since $\underline{H}(\boldsymbol{X}) = H(P)$ is obvious from the definition of $\underline{H}(\boldsymbol{X})$, in view of Theorems 3 and 4 it is sufficient to establish

$$G_{[\varepsilon]}(\boldsymbol{X}) \leq (1-\varepsilon)H(P) \quad (78)$$

for arbitrarily fixed $\varepsilon \in (0,1)$. Note that (78) is obvious from the ordinary source coding theorem if $\varepsilon = 0$. We fix a constant δ satisfying $0 < \delta < \min\{\varepsilon, H(P)\}$ arbitrarily and define

$$T_n = \left\{x^n \in \mathcal{X}^n : \left|\frac{1}{n}\log_2 \frac{1}{P_{X^n}(x^n)} - H(P)\right| \leq \delta\right\}.$$

Then, the weak law of large numbers guarantees that $\Pr\{X^n \in T_n\} \geq 1-\delta$ for all sufficiently large n. Letting $\gamma > 0$ be a sufficiently small constant, we note that we can choose a subset $A_n \subseteq T_n$ satisfying

$$\begin{aligned}1-\varepsilon-\gamma &\leq \Pr\{X^n \in A_n\} \\ &\leq 1-\varepsilon-\gamma + \max_{x^n \in T_n \setminus A_n} P_{X^n}(x^n) \end{aligned} \quad (79)$$

for all sufficiently large n. Denoting $\Gamma_n(A_n)$ the function defined in (31), it follows that

$$\begin{aligned}\frac{1}{n}G_{[\varepsilon+\gamma]}(X^n) &\overset{1)}{\leq} \Gamma_n(A_n) \\ &\overset{2)}{\leq} \sum_{x^n \in A_n} P_{X^n}(x^n)(H(P)+\delta) \\ &\overset{3)}{\leq} (1-\varepsilon-\gamma+2^{-n(H(P)-\delta)})(H(P)+\delta) \end{aligned} \quad (80)$$

where the marked inequalities in (80) follow from

1) the definition of $G_{[\varepsilon+\gamma]}(X^n)$ and (79);
2) $\Pr\{X^n \in A_n\} \log_2 \Pr\{X^n \in A_n\} \leq 0$ and
$$\frac{1}{n}\log_2 \frac{1}{P_{X^n}(x^n)} \leq H(P)+\delta, \qquad \text{for all } x^n \in A_n \subset T_n;$$
3) $P_{X^n}(x^n) \leq 2^{-n(H(P)-\delta)}$, for all $x^n \in T_n \setminus A_n$.

By taking the limit superior of both sides of (80) and letting $\gamma \downarrow 0$, we have $G_{[\varepsilon]}(\boldsymbol{X}) \leq (1-\varepsilon)H(P)+\delta$. This establishes (39) because $\delta > 0$ can be made arbitrarily small. □

Proof of (40): Since

$$G_{[\varepsilon]}(\boldsymbol{X}) \leq \begin{cases}\alpha_1 H(P_1) + (\alpha_2-\varepsilon)H(P_2), & \text{if } 0 \leq \varepsilon < \alpha_2 \\ (1-\varepsilon)H(P_1), & \text{otherwise}\end{cases} \quad (81)$$

can be obtained similarly to (39), we prove only the inequality opposite to (81). Theorem 4 and $\underline{H}(\boldsymbol{X}) = H(P_1)$ tell us that it is sufficient to prove

$$G_{[\varepsilon]}(\boldsymbol{X}) \geq \alpha_1 H(P_1) + (\alpha_2-\varepsilon)H(P_2)$$

for an arbitrarily fixed $0 \le \varepsilon < \alpha_2$. To this end, we fix a constant $\delta > 0$ and define $T_n^{(1)}$ and $T_n^{(2)}$ by

$$T_n^{(1)} = \left\{ x^n \in \mathcal{X}^n : \left| \frac{1}{n} \log_2 \frac{1}{P_{X^n}(x^n)} - H(P_1) \right| < \delta \right\}$$
$$T_n^{(2)} = \left\{ x^n \in \mathcal{X}^n : \left| \frac{1}{n} \log_2 \frac{1}{P_{X^n}(x^n)} - H(P_2) \right| < \delta \right\}$$

respectively. Notice that we can choose a sufficiently small $\delta > 0$ such that

$$\Pr\left\{ X^n \in T_n^{(1)} \right\} = \alpha_1 - \eta_n$$
$$\Pr\left\{ X^n \in T_n^{(2)} \right\} = \alpha_2 - \eta_n$$

and $T_n^{(1)} \cap T_n^{(2)} = \phi$, where $\eta_n \to 0$ as $n \to \infty$. We note that for any $A_n \subseteq \mathcal{X}^n$ such that $\Pr\{X^n \in A_n\} \ge 1 - \varepsilon - \gamma$, where $\gamma > 0$ is an arbitrary constant, $\Gamma_n(A_n)$ defined in (31) is lower-bounded in the following way:

$$\begin{aligned}
\Gamma_n(A_n) &= \frac{1}{n} \Pr\{X^n \in A_n\} \log_2 \Pr\{X^n \in A_n\} \\
&\quad + \frac{1}{n} \sum_{x^n \in A_n} P_{X^n}(x^n) \log_2 \frac{1}{P_{X^n}(x^n)} \\
&\ge \frac{1}{n} \sum_{x^n \in A_n \cap T_n^{(1)}} P_{X^n}(x^n) \log_2 \frac{1}{P_{X^n}(x^n)} \\
&\quad + \frac{1}{n} \sum_{x^n \in A_n \cap T_n^{(2)}} P_{X^n}(x^n) \log_2 \frac{1}{P_{X^n}(x^n)} + o(1) \\
&\overset{4)}{\ge} \Pr\left\{ X^n \in A_n \cap T_n^{(1)} \right\} H(P_1) \\
&\quad + \Pr\left\{ X^n \in A_n \cap T_n^{(2)} \right\} H(P_2) - \delta + o(1) \\
&\overset{5)}{\ge} (\alpha_1 - \eta_n) H(P_1) \\
&\quad + (\alpha_2 - \varepsilon - \gamma - \eta_n) H(P_2) - \delta
\end{aligned} \tag{82}$$

where $o(1) \to 0$ as $n \to \infty$ and the marked inequalities follow from

4) the definitions of $T_n^{(1)}$ and $T_n^{(2)}$;
5) $H(P_1) < H(P_2)$,

$$\Pr\left\{ X^n \in A_n \cap T_n^{(1)} \right\} + \Pr\left\{ X^n \in A_n \cap T_n^{(2)} \right\} \ge 1 - \varepsilon - \gamma - 2\eta_n$$

and

$$\Pr\left\{ X^n \in A_n \cap T_n^{(1)} \right\} \le \alpha_1 - \eta_n.$$

Since the right-hand side of (82) no longer depends on A_n, (82) leads to

$$\limsup_{n \to \infty} \frac{1}{n} G_{[\varepsilon+\gamma]}(X^n) \ge \alpha_1 H(P_1) + (\alpha_2 - \varepsilon - \gamma) H(P_2) - \delta.$$

By letting $\gamma \downarrow 0$, we have $G_{[\varepsilon]}(\boldsymbol{X}) \le \alpha_1 H(P_1) + (\alpha_2 - \varepsilon) H(P_2) - \delta$. Now we have (40) because $\delta > 0$ can be made arbitrarily small. □

ACKNOWLEDGMENT

The authors would like to acknowledge the Associate Editor Imre Csiszár and the anonymous referees for their helpful comments. The authors also wish to thank Mr. Naoto Yamaguchi, who was a master course student at the Graduate School of Engineering, University of Tokyo, for interesting discussions.

REFERENCES

[1] K. L. Chung, *A Course in Probability Theory*, 2nd ed. New York: Academic, 1974.
[2] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1989.
[3] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
[4] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.
[5] T. S. Han, *Information-Spectrum Methods in Information Theory*. Berlin, Germany: Springer-Verlag, 2003.
[6] ——, "Weak variable-length source coding theorem," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1217–1226, Jul. 2000.
[7] D. A. Huffman, "A method for construction of minimum redundancy codes," *Proc. IRE*, vol. 40, pp. 1098–1101, 1952.
[8] J. C. Kieffer, "Finite-state adaptive block to variable-length noiseless coding of a nonstationary information sources," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1259–1263, Nov. 1989.
[9] G. Katona and T. Nemetz, "Huffman codes and self-information," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 3, pp. 337–340, May 1976.
[10] T. Nemetz and J. Simon, "Self-information and optimal codes," in *Topics in Information Theory*. Amsterdam, The Netherlands: North-Holland, 1977, pp. 457–468.
[11] D. Ornstein and P. Shields, "Universal almost sure data compression," *Ann. Probab.*, vol. 18, no. 2, pp. 441–452, Apr. 1990.
[12] P. C. Shields, *The Ergodic Theory of Discrete Sample Path*. Providence, RI: Amer. Math. Soc., 1996.
[13] Y. Steinberg and S. Verdú, "Simulation of random processes and rate-distortion theory," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 63–86, Jan. 1996.
[14] S. Vembu and S. Verdú, "Generating random bits from an arbitrary source: Fundamental limits," *IEEE Trans. Inf. Theory*, vol. 41, no. 5, pp. 1322–1332, Sep. 1995.
[15] K. Visweswariah, S. R. Kulkarni, and S. Verdú, "Source codes as random number generators," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 462–471, Mar. 1998.
[16] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 5, pp. 530–536, Sep. 1978.

Results on the Nonlinear Span of Binary Sequences

Panagiotis Rizomiliotis, *Member, IEEE,* and
Nicholas Kalouptsidis, *Senior Member, IEEE*

***Abstract*—The problem of finding the length of a shortest feedback shift register that generates a given finite-length sequence is considered. An efficient algorithm for the determination of the span is proposed, that takes advantage of the special block structure of the associated system of linear equations. The span distribution of finite-length binary sequences is also studied.**

***Index Terms*—Binary sequences, nonlinear feedback functions, shift registers, span.**

I. INTRODUCTION

The binary sequences produced by finite-state machines find various applications in modern communications [13]. Depending on the application, the sequences are required to possess certain properties. When

Manuscript received May 6, 2003; revised July 16, 2004. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Chicago, IL, June/July 2004.

The authors are with the Department of Informatics and Telecommunications, Division of Communications and Signal Processing, National and Kapodistrian University of Athens, 15784 Athens, Greece (e-mail: rizop@di.uoa.gr; kalou@di.uoa.gr).

Communicated by K. G. Paterson, Associate Editor for Sequences.
Digital Object Identifier 10.1109/TIT.2005.844090