Optimistic Shannon Coding Theorems for Arbitrary Single-User Systems

Po-Ning Chen and Fady Alajaji, *Member, IEEE*

***Abstract*—The conventional definitions of the source coding rate and of channel capacity require the existence of reliable codes for *all sufficiently large block lengths*. Alternatively, if it is required that good codes exist for *infinitely many block lengths*, then *optimistic* definitions of source coding rate and channel capacity are obtained.**

In this work, formulas for the optimistic minimum achievable fixed-length source coding rate and the minimum ε-achievable source coding rate for arbitrary finite-alphabet sources are established. The expressions for the optimistic capacity and the optimistic ε-capacity of arbitrary single-user channels are also provided. The expressions of the optimistic source coding rate and capacity are examined for the class of information stable sources and channels, respectively. Finally, examples for the computation of optimistic capacity are presented.

***Index Terms*—Error probability, optimistic channel capacity, optimistic source coding rate, Shannon theory, source-channel separation theorem.**

I. Introduction

The conventional definition of the minimum achievable fixed-length source coding rate T for a source $\boldsymbol{Z}$ [13, Definition 4] requires the existence of reliable source codes for *all sufficiently large block lengths*. Alternatively, if it is required that reliable codes exist for *infinitely many block lengths*, a new, more *optimistic* definition of source coding rate (denoted by $\underline{T}$) is obtained [13]. Similarly, the *optimistic capacity* $\overline{C}$ is defined by requiring the existence of reliable channel codes for infinitely many block lengths, as opposed to the definition of the conventional channel capacity C [14, Definition 1].

This concept of optimistic source coding rate and capacity has recently been investigated by Verdú *et al.* for *arbitrary* (not necessarily stationary, ergodic, information stable, etc.) sources and single-user channels [13], [14]. More specifically, they establish an additional *operational* characterization for the optimistic minimum achievable source coding rate ($\underline{T}$) by demonstrating that for a given channel, the classical statement of the source-channel separation theorem[1] holds for every channel if $\underline{T} = T$ [13]. In a dual fashion, they also show that for channels with $\overline{C} = C$, the classical separation theorem holds for every source. They also conjecture that $\underline{T}$ and $\overline{C}$ do not seem to admit a simple expression.

Manuscript received September 1, 1998; revised April 19, 1999. This work was supported in part under an ARC Grant from Queen's University, the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant OGP0183645, and the National Science Council of Taiwan, ROC, under Grant NSC 87-2213-E-009-139. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, MIT, Cambridge, MA, August 1998.

P.-N. Chen is with the Department of Communication Engineering, National Chiao-Tung University, HsinChu, Taiwan, ROC.

F. Alajaji is with the Department of Mathematics and Statistics, Queen's University, Kingston, Ont. K7L 3N6, Canada.

Communicated by I. Csiszár, Associate Editor for Shannon Theory.

Publisher Item Identifier S 0018-9448(99)07305-8.

[1] By the "classical statement of the source-channel separation theorem," we mean the following. Given a source $\boldsymbol{Z}$ with (conventional) source coding rate $T(\boldsymbol{Z})$ and channel $\boldsymbol{W}$ with capacity C, then $\boldsymbol{Z}$ can be reliably transmitted over $\boldsymbol{W}$ if $T(\boldsymbol{Z}) < C$. Conversely, if $T(\boldsymbol{Z}) > C$, then $\boldsymbol{Z}$ cannot be reliably transmitted over $\boldsymbol{W}$. By reliable transmissibility of the source over the channel, we mean that there exists a sequence of source-channel codes such that the decoding error probability vanishes as the block length $n \to \infty$ (cf. [13]).

In this work, we demonstrate that $\underline{T}$ and $\overline{C}$ do indeed have a general formula. The key to these results is the application of the generalized entropy/information rates introduced in [3] and [4] to the existing proofs by Verdú and Han [7], [14] of the direct and converse parts of the conventional coding theorems. We also provide a general expression for the optimistic minimum ε-achievable source coding rate and the optimistic ε-capacity.

In Section II, we briefly introduce the generalized sup/inf-information/entropy rates which will play a key role in proving our optimistic coding theorems. In Section III, we provide the optimistic source coding theorems. They are shown based on *two recent bounds* due to Han [7] on the error probability of a source code as a function of its size. Interestingly, these bounds constitute the natural counterparts of the upper bound provided by Feinstein's Lemma and the Verdú–Han lower bound to the error probability of a channel code. Furthermore, we show that for information-stable sources, the formula for $\underline{T}$ reduces to

$$\underline{T} = \liminf_{n\to\infty} \frac{1}{n} H(X^n).$$

This is in contrast to the expression for T, which is known to be

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

The above result leads us to observe that for sources that are both stationary and information-stable, the classical separation theorem is valid for every channel.

In Section IV, we present (without proving) the general optimistic channel coding theorems, and prove that for the class of information-stable channels the expression of $\overline{C}$ becomes

$$\overline{C} = \limsup_{n\to\infty} \sup_{X^n} \frac{1}{n} I(X^n; Y^n)$$

while the expression of C is

$$C = \liminf_{n\to\infty} \sup_{X^n} \frac{1}{n} I(X^n; Y^n).$$

Finally, in Section V, we present examples for the computation of C and $\overline{C}$ for information-stable as well as information-unstable channels.

II. ε-Inf/Sup-Information/Entropy Rates

Consider an input process $\boldsymbol{X}$ defined by a sequence of finite-dimensional distributions [14] $\boldsymbol{X} \triangleq \{X^n = (X_1^{(n)}, \cdots, X_n^{(n)})\}_{n=1}^{\infty}$. Denote by $\boldsymbol{Y} \triangleq \{Y^n = (Y_1^{(n)}, \cdots, Y_n^{(n)})\}_{n=1}^{\infty}$ the corresponding output process induced by $\boldsymbol{X}$ via the channel

$$\boldsymbol{W} \triangleq \{W^n = P_{Y^n|X^n}: \mathcal{X}^n \to \mathcal{Y}^n\}_{n=1}^{\infty}$$

which is an arbitrary sequence of n-dimensional conditional distributions from $\mathcal{X}^n$ to $\mathcal{Y}^n$, where $\mathcal{X}$ and $\mathcal{Y}$ are the input and output alphabets, respectively. We assume throughout this correspondence that $\mathcal{X}$ and $\mathcal{Y}$ are finite.

In [8] and [14], Han and Verdú introduce the notions of inf/sup-information/entropy rates and illustrate the key role these information measures play in proving a general lossless (block) source coding theorem and a general channel coding theorem.

The *inf-information rate* $\underline{I}(\boldsymbol{X}; \boldsymbol{Y})$ (resp., *sup-information rate* $\overline{I}(\boldsymbol{X}; \boldsymbol{Y})$) between processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ is defined in [8] as the *liminf in probability* (resp., *limsup in probability*) of the sequence of normalized information densities $(1/n)\, i_{X^nW^n}(X^n; Y^n)$, where

$$\frac{1}{n} i_{X^nW^n}(a^n; b^n) \triangleq \frac{1}{n} \log \frac{P_{Y^n|X^n}(b^n|a^n)}{P_{Y^n}(b^n)}.$$

When $\boldsymbol{X}$ is equal to $\boldsymbol{Y}$, $\overline{I}(\boldsymbol{X}; \boldsymbol{X})$ (resp., $\underline{I}(\boldsymbol{X}; \boldsymbol{X})$) is referred to as the *sup* (resp., *inf*) *entropy rate* of $\boldsymbol{X}$ and is denoted by $\overline{H}(\boldsymbol{X})$ (resp., $\underline{H}(\boldsymbol{X})$).

The *liminf in probability* of a sequence of random variables is defined as follows [8]: if A_n is a sequence of random variables, then its *liminf in probability* is the largest extended real number $\underline{U}$ such that

$$\lim_{n\to\infty} \Pr[A_n < \underline{U}] = 0. \tag{1}$$

Similarly, its *limsup in probability* is the smallest extended real number $\overline{U}$ such that

$$\lim_{n\to\infty} \Pr[A_n > \overline{U}] = 0. \tag{2}$$

Note that these two quantities are always defined; if they are equal, then the sequence of random variables converges in probability to a constant.

It is straightforward to deduce that (1) and (2) are, respectively, equivalent to

$$\liminf_{n\to\infty} \Pr[A_n < \underline{U}] = \limsup_{n\to\infty} \Pr[A_n < \underline{U}] = 0 \tag{3}$$

and

$$\liminf_{n\to\infty} \Pr[A_n < \overline{U}] = \limsup_{n\to\infty} \Pr[A_n > \overline{U}] = 0. \tag{4}$$

We can observe, however, that there might exist cases of interest where *only* the liminf's of the probabilities in (3) and (4) are equal to zero, while the limsup's do *not* vanish. There are also other cases where *both* the liminf's and limsup's in (3) and (4) do not vanish, but they are upper-bounded by a prescribed threshold ε. Furthermore, there are situations where the interval $[\underline{U}, \overline{U}]$ does not contain only one point; e.g., when A_n converges in distribution to another random variable. This remark constitutes the motivation to the recent work in [3] and [4], where generalized versions of the inf/sub-information/entropy rates are established.

Definition 2.1—Inf/Sup Spectrums [3], [4]: If $\{A_n\}_{n=1}^{\infty}$ is a sequence of random variables taking values in a finite set $\mathcal{A}$, then its *inf-spectrum* $\underline{u}(\cdot)$ and its *sup-spectrum* $\overline{u}(\cdot)$ are defined by

$$\underline{u}(\theta) \triangleq \liminf_{n\to\infty} \Pr\{A_n \le \theta\}$$

and

$$\overline{u}(\theta) \triangleq \limsup_{n\to\infty} \Pr\{A_n \le \theta\}.$$

In other words, $\underline{u}(\cdot)$ and $\overline{u}(\cdot)$ are, respectively, the liminf and the limsup of the cumulative distribution function (CDF) of A_n. Note that by definition, the CDF of A_n—$\Pr\{A_n \le \theta\}$—is nondecreasing and right-continuous. However, for $\underline{u}(\cdot)$ and $\overline{u}(\cdot)$, only the *nondecreasing* property remains.

Definition 2.2—Quantile of Inf/Sup-Spectrum [3], [4]: For any $0 \le \varepsilon \le 1$, the *quantiles* $\underline{U}_\varepsilon$ and $\overline{U}_\varepsilon$ of the sup-spectrum and the inf-spectrum are defined by

$$\underline{U}_\varepsilon \triangleq \sup\{\theta: \overline{u}(\theta) \le \varepsilon\}$$

and

$$\overline{U}_\varepsilon \triangleq \sup\{\theta: \underline{u}(\theta) \le \varepsilon\}$$

respectively. It follows from the above definitions that $\underline{U}_\varepsilon$ and $\overline{U}_\varepsilon$ are right-continuous and nondecreasing in ε. Note that Han and Verdú's liminf/limsup in probability of A_n are *special cases* of $\underline{U}_\varepsilon$ and $\overline{U}_\varepsilon$. More specifically, the following hold:

$$\underline{U} = \underline{U}_0 \quad \text{and} \quad \overline{U} = \overline{U}_{1-}$$

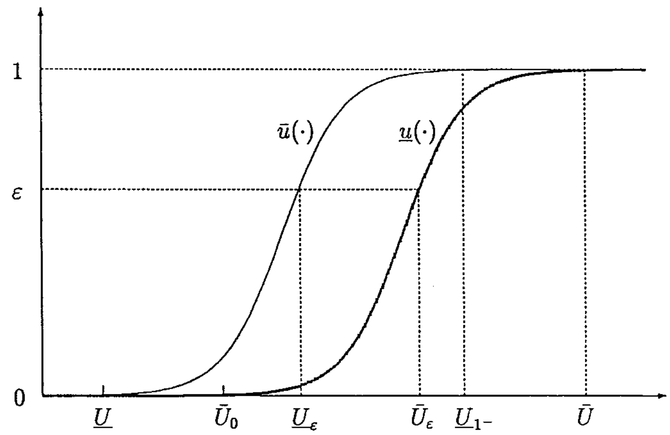

Fig. 1. The asymptotic CDF's of a sequence of random variables $\{A_n\}_{n=1}^{\infty}$: $\overline{u}(\cdot)$ = sup-spectrum and $\underline{u}(\cdot)$ = inf-spectrum.

where the superscript "$-$" denotes a strict inequality in the definition of $\overline{U}_{1-}$; i.e.,

$$\overline{U}_{\varepsilon-} \triangleq \sup\{\theta: \underline{u}(\theta) < \varepsilon\}.$$

Note also that $\underline{U} \le \underline{U}_\varepsilon \le \overline{U}_\varepsilon \le \overline{U}$. Remark that $\underline{U}_\varepsilon$ and $\overline{U}_\varepsilon$ always exist. For a better understanding of the quantities defined above, we depict them in Fig. 1. If we replace A_n by the normalized information (resp., entropy) density, we get the following definitions.

Definition 2.3—ε-Inf/Sup-Information Rates [3], [4]: The ε-inf-information rate $\underline{I}_\varepsilon(\boldsymbol{X}; \boldsymbol{Y})$ (resp., ε-sup-information rate $\overline{I}_\varepsilon(\boldsymbol{X}; \boldsymbol{Y})$) between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is defined as the quantile of the sup-spectrum (resp., inf-spectrum) of the normalized information density. More specifically

$$\underline{I}_\varepsilon(\boldsymbol{X}; \boldsymbol{Y}) \triangleq \sup\{\delta: \bar{i}_{\boldsymbol{XW}}(\delta) \le \varepsilon\}$$

where

$$\bar{i}_{\boldsymbol{XW}}(\delta) \triangleq \limsup_{n\to\infty} \Pr\left\{\frac{1}{n} i_{X^nW^n}(X^n; Y^n) \le \delta\right\}$$

and

$$\overline{I}_\varepsilon(\boldsymbol{X}; \boldsymbol{Y}) \triangleq \sup\{\delta: \underline{i}_{\boldsymbol{XW}}(\delta) \le \varepsilon\}$$

where

$$\underline{i}_{\boldsymbol{XW}}(\delta) \triangleq \limsup_{n\to\infty} \Pr\left\{\frac{1}{n} i_{X^nW^n}(X^n; Y^n) \le \delta\right\}.$$

Definition 2.4—ε-Inf/Sup-Entropy Rates [3], [4]: The ε-inf-entropy rate $\underline{H}_\varepsilon(\boldsymbol{X})$ (resp., ε-sup-entropy rate $\overline{H}_\varepsilon(\boldsymbol{X})$) for a source $\boldsymbol{X}$ is defined as the quantile of the sup-spectrum (resp., inf-spectrum) of the normalized entropy density. More specifically

$$\underline{H}_\varepsilon(\boldsymbol{X}) \triangleq \sup\{\delta: \bar{h}_{\boldsymbol{X}}(\delta) \le \varepsilon\}$$

where

$$\bar{h}_{\boldsymbol{X}}(\delta) \triangleq \limsup_{n\to\infty} \Pr\left\{\frac{1}{n} h_{X^n}(X^n) \le \delta\right\}$$

and

$$\overline{H}_\varepsilon(\boldsymbol{X}) \triangleq \sup\{\delta: \underline{h}_{\boldsymbol{X}}(\delta) \le \varepsilon\}$$

where

$$\underline{h}_{\boldsymbol{X}}(\delta) \triangleq \limsup_{n\to\infty} \Pr\left\{\frac{1}{n} h_{X^n}(X^n) \le \delta\right\}$$

and

$$\frac{1}{n} h_{X^n}(X^n) \triangleq \frac{1}{n} \log(1/P_{X^n}(X^n)).$$

III. Optimistic Source Coding Theorems

In [13], Vembu *et al.* characterize the sources for which the classical separation theorem holds for *every channel*. They demonstrate that for a given source $\boldsymbol{X}$, the separation theorem holds for every channel if its optimistic minimum achievable source coding rate ($\underline{T}(\boldsymbol{X})$) coincides with its conventional (or pessimistic) minimum achievable source coding rate ($T(\boldsymbol{X})$); i.e., if $\underline{T}(\boldsymbol{X}) = T(\boldsymbol{X})$.

We herein establish a general formula for $\underline{T}(\boldsymbol{X})$. We prove that for any source $\boldsymbol{X}$

$$\underline{T}(\boldsymbol{X}) = \underline{H}_{1^-}(\boldsymbol{X}).$$

We also provide the general expression for the optimistic minimum ε-achievable source coding rate. We show these results based on *two new bounds* due to Han (one upper bound and one lower bound) on the error probability of a source code [7, Ch. 1]. The upper bound (Lemma 3.1) consists of the *counterpart* of Feinstein's lemma for channel codes (cf., for example, [14, Theorem 1]), while the lower bound (Lemma 3.2) consists of the *counterpart* of the Verdú–Han lower bound on the error probability of a channel code [14, Theorem 4]. As in the case of the channel coding bounds, both source coding bounds (Lemmas 3.1 and 3.2) hold for arbitrary sources and for arbitrary fixed block length.

Definition 3.5: An (n, M) fixed-length source code for X^n is a collection of M n-tuples $\mathcal{C}_n = \{c_1^n, \cdots, c_M^n\}$. The error probability of the code is $P_e^{(n)} \triangleq \Pr[X^n \notin \mathcal{C}_n]$.

Definition 3.6—Optimistic ε-Achievable Source Coding Rate: Fix $0 < \varepsilon < 1$. $R \geq 0$ is an optimistic ε-achievable rate if, for every $\gamma > 0$, there exists a sequence of (n, M) fixed-length source codes $\mathcal{C}_n$ such that

$$\frac{1}{n}\log M < R + \gamma \quad \text{and} \quad P_e^{(n)} \leq \varepsilon \quad \text{for infinitely many } n.$$

The infimum of all ε-achievable source coding rates for source $\boldsymbol{X}$ is denoted by $\underline{T}_\varepsilon(\boldsymbol{X})$. Also define

$$\underline{T}(\boldsymbol{X}) \triangleq \sup_{0<\varepsilon<1} \underline{T}_\varepsilon(\boldsymbol{X}) = \lim_{\varepsilon \downarrow 0} \underline{T}_\varepsilon(\boldsymbol{X})$$

as the optimistic source coding rate.

Lemma 3.1—[7, Lemma 1.5]: Fix a positive integer n. There exists an (n, M) source block code $\mathcal{C}_n$ for P_{X^n} such that its error probability satisfies

$$P_e^{(n)} \leq \Pr\left[\frac{1}{n} h_{X^n}(X^n) > \frac{1}{n}\log M\right].$$

Lemma 3.2—[7, Lemma 1.6]: Every (n, M) source block code $\mathcal{C}_n$ for P_{X^n} satisfies

$$P_e^{(n)} \geq \Pr\left[\frac{1}{n} h_{X^n}(X^n) > \frac{1}{n}\log M + \gamma\right] - \exp\{-n\gamma\}$$

for every $\gamma > 0$.

We next use Lemmas 3.1 and 3.2 to prove *general* optimistic (fixed-length) source coding theorems.

Theorem 3.1—Optimistic Minimum ε-Achievable Source Coding Rate Formula: Fix $0 < \varepsilon < 1$. For any source $\boldsymbol{X}$

$$\underline{H}_{\varepsilon^-}(\boldsymbol{X}) \leq \underline{T}_{1-\varepsilon}(\boldsymbol{X}) \leq \underline{H}_\varepsilon(\boldsymbol{X}).$$

Note that actually $\underline{T}_{1-\varepsilon}(\boldsymbol{X}) = \underline{H}_\varepsilon(\boldsymbol{X})$, except possibly at the points of discontinuities of $\underline{H}_\varepsilon(\boldsymbol{X})$ (which are countable).

Proof:

1) Forward Part (Achievability)—$\underline{T}_{1-\varepsilon}(X) \leq \underline{H}_\varepsilon(X)$: We need to prove the existence of a sequence of block codes $\{\mathcal{C}_n\}_{n\geq 0}$ such that, for every $\gamma > 0$, $(1/n)\log|\mathcal{C}_n| < \underline{H}_\varepsilon(\boldsymbol{X}) + \gamma$ and $P_e^{(n)} \leq 1-\varepsilon$ for infinitely many n. Lemma 3.1 ensures the existence (for any $\gamma > 0$) of a source block code $\mathcal{C}_n = (n, \exp\{n(\underline{H}_\varepsilon + \gamma/2)\})$ with error probability

$$P_e^{(n)} \leq \Pr\left\{\frac{1}{n} h_{X^n}(X^n) > \underline{H}_\varepsilon + \frac{\gamma}{2}\right\}.$$

Therefore,

$$\begin{aligned}\liminf_{n\to\infty} P_e^{(n)} &\leq \liminf_{n\to\infty} \Pr\left\{\frac{1}{n} h_{X^n}(X^n) > \underline{H}_\varepsilon + \frac{\gamma}{2}\right\} \\ &= 1 - \limsup_{n\to\infty} \Pr\left\{\frac{1}{n} h_{X^n}(X^n) \leq \underline{H}_\varepsilon(\boldsymbol{X}) + \frac{\gamma}{2}\right\} \\ &< 1-\varepsilon \end{aligned} \tag{5}$$

where (5) follows from the definition of $\underline{H}_\varepsilon(\boldsymbol{X})$. Hence, $P_e^{(n)} \leq 1-\varepsilon$ for infinitely many n.

2) Converse Part—$\underline{T}_{1-\varepsilon}(\boldsymbol{X}) \geq \underline{H}_{\varepsilon^-}(\boldsymbol{X})$: Assume without loss of generality that $\underline{H}_{\varepsilon^-}(\boldsymbol{X}) > 0$. We will prove the converse by contradiction. Suppose that $\underline{T}_{1-\varepsilon}(\boldsymbol{X}) < \underline{H}_{\varepsilon^-}(\boldsymbol{X})$. Then $(\exists\gamma > 0)$ $\underline{T}_{1-\varepsilon}(\boldsymbol{X}) < \underline{H}_{\varepsilon^-}(\boldsymbol{X}) - 3\gamma$. By definition of $\underline{T}_{1-\varepsilon}(\boldsymbol{X})$, there exists a sequence of codes $\mathcal{C}_n$ such that

$$\frac{1}{n}\log|\mathcal{C}_n| < [\underline{H}_{\varepsilon^-}(\boldsymbol{X}) - 3\gamma] + \gamma$$

and

$$\liminf_{n\to\infty} P_e^{(n)} \leq 1-\varepsilon. \tag{6}$$

By Lemma 3.2

$$\begin{aligned}P_e^{(n)} &\geq \Pr\left[\frac{1}{n} h_{X^n}(X^n) > \frac{1}{n}\log|\mathcal{C}_n| + \gamma\right] - e^{-n\gamma} \\ &\geq \Pr\left[\frac{1}{n} h_{X^n}(X^n) > (\underline{H}_{\varepsilon^-}(\boldsymbol{X}) - 2\gamma) + \gamma\right] - e^{-n\gamma}.\end{aligned}$$

Therefore,

$$\begin{aligned}\liminf_{n\to\infty} P_e^{(n)} &\geq 1 - \limsup_{n\to\infty} \Pr\left[\frac{1}{n} h_{X^n}(X^n) \leq \underline{H}_{\varepsilon^-}(\boldsymbol{X}) - \gamma\right] \\ &> 1-\varepsilon\end{aligned}$$

where the last inequality follows from the definition of $\underline{H}_{\varepsilon^-}(\boldsymbol{X})$. Thus a contradiction to (6) is obtained.

3) Equality: $\underline{H}_\varepsilon(\boldsymbol{X})$ is a nondecreasing function of ε; hence the number of discontinuous points is countable. For any continuous point ε, we have that $\underline{H}_\varepsilon(\boldsymbol{X}) = \underline{H}_{\varepsilon^-}(\boldsymbol{X})$, and thus $\underline{T}_\varepsilon(\boldsymbol{X}) = \underline{H}_\varepsilon(\boldsymbol{X})$. □

Theorem 3.2—Optimistic Minimum Achievable Source Coding Rate Formula: For any source $\boldsymbol{X}$

$$\underline{T}(\boldsymbol{X}) = \underline{H}_{1^-}(\boldsymbol{X}).$$

Proof: By definition,

$$\underline{T}(\boldsymbol{X}) \triangleq \sup_{0<\varepsilon<1} \underline{T}_\varepsilon(\boldsymbol{X}) \geq \sup_{0<\varepsilon<1} \underline{H}_{\varepsilon^-}(\boldsymbol{X}) \geq \underline{H}_{1^-}(\boldsymbol{X}).$$

On the other hand, suppose that $\underline{H}_{1^-}(\boldsymbol{X}) < \underline{T}(\boldsymbol{X})$. Then $\exists\gamma > 0$ such that

$$\underline{H}_{1^-}(\boldsymbol{X}) < \underline{T}(\boldsymbol{X}) - \gamma.$$

But by definition of $\underline{T}(\boldsymbol{X})$, there exists $0 < \varepsilon = \varepsilon(\gamma) < 1$ such that

$$\underline{T}(\boldsymbol{X}) - \gamma < \underline{T}_\varepsilon(\boldsymbol{X}).$$

Therefore,

$$\underline{H}_{1^-}(\boldsymbol{X}) < \underline{T}(\boldsymbol{X}) - \gamma < \underline{T}_{\varepsilon}(\boldsymbol{X}) \le \underline{H}_{1-\varepsilon}(\boldsymbol{X}) \le \underline{H}_{1^-}(\boldsymbol{X})$$

and a contradiction is obtained. □

We conclude this section by examining the expression of $\underline{T}(\boldsymbol{X})$ for information stable sources. It is already known (cf., for example, [13]) that for an information-stable source $\boldsymbol{X}$

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

We herein prove a parallel expression for $\underline{T}(\boldsymbol{X})$.

Definition 3.7—Information-Stable Sources [13]: A source $\boldsymbol{X}$ is said to be information-stable if $H(X^n) > 0$ for n sufficiently large, and $h_{X^n}(X^n)/H(X^n)$ converges in probability to one as $n \to \infty$, i.e.,

$$\limsup_{n\to\infty} \Pr\left[\left|\frac{h_{X^n}(X^n)}{H(X^n)} - 1\right| > \gamma\right] = 0, \qquad \forall\, \gamma > 0$$

where $H(X^n) = E[h_{X^n}(X^n)]$ is the entropy of X^n.

Lemma 3.3: Every information source $\boldsymbol{X}$ satisfies

$$\underline{T}(\boldsymbol{X}) = \liminf_{n\to\infty} \frac{1}{n} H(X^n).$$

Proof:

1) $(\underline{T}(\boldsymbol{X}) \ge \liminf_{n\to\infty}(1/n)H(X^n))$*:* Fix $\varepsilon > 0$ arbitrarily small. Using the fact that $h_{X^n}(X^n)$ is a (finite-alphabet) nonnegative bounded random variable, we can write the normalized block entropy as

$$\begin{aligned}\frac{1}{n}H(X^n) &= E\left[\frac{1}{n}h_{X^n}(X^n)\right]\\ &= E\left[\frac{1}{n}h_{X^n}(X^n)1\left\{0 \le \frac{1}{n}h_{X^n}(X^n) \le \underline{H}_{1^-}(\boldsymbol{X}) + \varepsilon\right\}\right]\\ &\quad + E\left[\frac{1}{n}h_{X^n}(X^n)1\left\{\frac{1}{n}h_{X^n}(X^n) > \underline{H}_{1^-}(\boldsymbol{X}) + \varepsilon\right\}\right].\end{aligned} \tag{7}$$

From the definition of $\underline{H}_{1^-}(\boldsymbol{X})$, it directly follows that the first term in the right-hand side of (7) is upper-bounded by $\underline{H}_{1^-}(\boldsymbol{X}) + \varepsilon$, and that the liminf of the second term is zero. Thus

$$\underline{T}(\boldsymbol{X}) = \underline{H}_{1^-}(\boldsymbol{X}) \ge \liminf_{n\to\infty} \frac{1}{n} H(X^n).$$

2) $(\underline{T}(\boldsymbol{X}) \le \liminf_{n\to\infty}(1/n)H(X^n))$*:* Fix $\varepsilon > 0$. Then for infinitely many n

$$\begin{aligned}\Pr&\left\{\frac{h_{X^n}(X^n)}{H(X^n)} - 1 > \varepsilon\right\}\\ &= \Pr\left\{\frac{1}{n}h_{X^n}(X^n) > (1+\varepsilon)\left(\frac{1}{n}H(X^n)\right)\right\}\\ &\ge \Pr\left\{\frac{1}{n}h_{X^n}(X^n) > (1+\varepsilon)\right.\\ &\qquad \left.\cdot\left(\liminf_{n\to\infty} \frac{1}{n}H(X^n) + \varepsilon\right)\right\}.\end{aligned}$$

Since $\boldsymbol{X}$ is information stable, we obtain that

$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}h_{X^n}(X^n) > (1+\varepsilon)\left(\liminf_{n\to\infty} \frac{1}{n}H(X^n) + \varepsilon\right)\right\} = 0.$$

By the definition of $\underline{H}_{1^-}(\boldsymbol{X})$, the above implies that

$$\underline{T}(\boldsymbol{X}) = \underline{H}_{1^-}(\boldsymbol{X}) \le (1+\varepsilon)\left(\liminf_{n\to\infty} \frac{1}{n}H(X^n) + \varepsilon\right).$$

The proof is completed by noting that ε can be made arbitrarily small. □

Observations:

- If the source $\boldsymbol{X}$ is *both* information-stable *and* stationary, the above Lemma yields

$$T(\boldsymbol{X}) = \underline{T}(\boldsymbol{X}) = \lim_{n\to\infty} \frac{1}{n} H(X^n).$$

  This implies that given a stationary and information-stable source $\boldsymbol{X}$, the classical separation theorem holds for every channel.
- Recall that both Lemmas 3.1 and 3.2 hold not only for arbitrary sources $\boldsymbol{X}$, but also for *arbitrary* fixed block length n. This leads us to conclude that they can analogously be employed to provide a simple proof to the conventional source coding theorems [8]

$$T(\boldsymbol{X}) = \overline{H}(\boldsymbol{X}),$$

  and

$$\overline{H}_{\varepsilon^-}(\boldsymbol{X}) \le T_{1-\varepsilon}(\boldsymbol{X}) \le \overline{H}_{\varepsilon}(\boldsymbol{X}).$$

IV. Optimistic Channel Coding Theorems

In this section, we state without proving the general expressions for the optimistic ε-capacity[2] $(\overline{C}_\varepsilon)$ and for the optimistic capacity $(\overline{C})$ of arbitrary single-user channels. The proofs of these expressions are straightforward once the right definition (of $\overline{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y})$) is made. They employ Feinstein's lemma and the Verdú–Han lower bound [14, Theorem 4], and follow the *same* arguments used in [14] to show the general expressions of the conventional channel capacity

$$C = \sup_{\boldsymbol{X}} \underline{I}_0(\boldsymbol{X};\boldsymbol{Y}) = \sup_{\boldsymbol{X}} \underline{I}(\boldsymbol{X};\boldsymbol{Y}),$$

and the conventional ε-capacity

$$\sup_{\boldsymbol{X}} \underline{I}_{\varepsilon^-}(\boldsymbol{X};\boldsymbol{Y}) \le C_\varepsilon \le \sup_{\boldsymbol{X}} \underline{I}_{\varepsilon}(\boldsymbol{X};\boldsymbol{Y}).$$

We close this section by proving the formula of $\overline{C}$ for information-stable channels.

Definition 4.8—Channel Block Code: An (n, M) code for channel W^n with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ is a pair of mappings

$$f\colon \{1, 2, \cdots, M\} \to \mathcal{X}^n$$

and

$$g\colon \mathcal{Y}^n \to \{1, 2, \cdots, M\}.$$

Its average error probability is given by

$$P_e^{(n)} \triangleq \frac{1}{M}\sum_{m=1}^{M} \sum_{\{y^n:\, g(y^n)\ne m\}} W^n(y^n|f(m)).$$

Definition 4.9—Optimistic ε-Achievable Rate: Fix $0 < \varepsilon < 1$. $R \ge 0$ is an optimistic ε-achievable rate if, for every $\gamma > 0$, there exists a sequence of (n, M) channel block codes such that

$$\frac{\log M}{n} > R - \gamma \quad \text{and} \quad P_e^{(n)} \le \varepsilon \quad \text{for infinitely many } n.$$

Definition 4.10—Optimistic ε-Capacity $\overline{C}_\varepsilon$: Fix $0 < \varepsilon < 1$. The supremum of optimistic ε-achievable rates is called the optimistic ε-capacity, $\overline{C}_\varepsilon$.

Definition 4.11—Optimistic Capacity $\overline{C}$: The optimistic channel capacity $\overline{C}$ is defined as the supremum of the rates that are optimistic ε-achievable for all $0 < \varepsilon < 1$. It follows immediately from the

[2] The authors would like to point out that the expression of $\overline{C}_\varepsilon$ was also separately obtained in [11, Theorem 7].

definition that

$$\overline{C} = \inf_{0<\varepsilon<1} \overline{C}_\varepsilon = \lim_{\varepsilon \downarrow 0} \overline{C}_\varepsilon$$

and that $\overline{C}$ is the supremum of all the rates R for which, for every $\gamma > 0$, there exists a sequence of (n, M) channel block codes such that

$$\frac{1}{n} \log M > R - \gamma \quad \text{and} \quad \liminf_{n\to\infty} P_e^{(n)} = 0.$$

Theorem 4.3—Optimistic ε-Capacity Formula: Fix $0 < \varepsilon < 1$. The optimistic ε-capacity $\overline{C}_\varepsilon$ satisfies

$$\sup_{\boldsymbol{X}} \overline{I}_{\varepsilon-}(\boldsymbol{X}; \boldsymbol{Y}) \le \overline{C}_\varepsilon \le \sup_{\boldsymbol{X}} \overline{I}_\varepsilon(\boldsymbol{X}; \boldsymbol{Y}). \tag{8}$$

Note that actually $\overline{C}_\varepsilon = \sup_{\boldsymbol{X}} \overline{I}_\varepsilon(\boldsymbol{X}; \boldsymbol{Y})$, except possibly at the points of discontinuities of $\sup_{\boldsymbol{X}} \overline{I}_\varepsilon(\boldsymbol{X}; \boldsymbol{Y})$ (which are countable).

Theorem 4.4—Optimistic Capacity Formula: The optimistic capacity $\overline{C}$ satisfies

$$\overline{C} = \sup_{\boldsymbol{X}} \overline{I}_0(\boldsymbol{X}; \boldsymbol{Y}).$$

We next investigate the expression of $\overline{C}$ for information-stable channels. The expression for the capacity of information-stable channels is already known (cf., for example, [13])

$$C = \liminf_{n\to\infty} \sup_{X^n} \frac{1}{n} I(X^n; Y^n)$$

where

$$C_n \triangleq \sup_{X^n} \frac{1}{n} I(X^n; Y^n).$$

We prove a dual formula for $\overline{C}$.

Definition 4.12—Information-Stable Channels [6], [9]: A channel $\boldsymbol{W}$ is said to be information-stable if there exists an input process $\boldsymbol{X}$ such that $0 < C_n < \infty$ for n sufficiently large, and

$$\limsup_{n\to\infty} \Pr\left[\left|\frac{i_{X^nW^n}(X^n; Y^n)}{nC_n} - 1\right| > \gamma\right] = 0, \qquad \forall \gamma > 0.$$

Lemma 4.4: Every information-stable channel $\boldsymbol{W}$ satisfies

$$\overline{C} = \limsup_{n\to\infty} \sup_{X^n} \frac{1}{n} I(X^n; Y^n).$$

Proof:

1) ($\overline{C} \le \limsup_{n\to\infty} \sup_{X^n} (1/n) I(X^n; Y^n)$)*:* By using a similar argument as in the proof of [14, Theorem 8, property h)], we have

$$\overline{I}_0(\boldsymbol{X}; \boldsymbol{Y}) \le \limsup_{n\to\infty} \sup_{X^n} \frac{1}{n} I(X^n; Y^n).$$

Hence

$$\overline{C} = \sup_{\boldsymbol{X}} \overline{I}_0(\boldsymbol{X}; \boldsymbol{Y}) \le \limsup_{n\to\infty} \sup_{X^n} \frac{1}{n} I(X^n; Y^n).$$

2) ($\overline{C} \ge \limsup_{n\to\infty} \sup_{X^n} (1/n) I(X^n; Y^n)$)*:* Suppose $\tilde{\boldsymbol{X}}$ is the input process that makes the channel information-stable. Fix $\varepsilon > 0$. Then for infinitely many n

$$\begin{aligned} P_{\tilde{X}^nW^n}&\left[\frac{1}{n} i_{\tilde{X}^nW^n}(\tilde{X}^n; Y^n) \le (1-\varepsilon)\left(\limsup_{n\to\infty} C_n - \varepsilon\right)\right] \\ &\le P_{\tilde{X}^nW^n}\left[\frac{i_{\tilde{X}^nW^n}(\tilde{X}^n; Y^n)}{n} < (1-\varepsilon) C_n\right] \\ &= P_{\tilde{X}^nW^n}\left[\frac{i_{\tilde{X}^nW^n}(\tilde{X}^n; Y^n)}{nC_n} - 1 < -\varepsilon\right]. \end{aligned}$$

Since the channel is information-stable, we get that

$$\liminf_{n\to\infty} P_{\tilde{X}^nW^n}\left[\frac{1}{n} i_{\tilde{X}^nW^n}(\tilde{X}^n; Y^n) \le (1-\varepsilon)\left(\limsup_{n\to\infty} C_n - \varepsilon\right)\right] = 0.$$

By the definition of $\overline{C}$, the above immediately implies that

$$\overline{C} = \sup_{\boldsymbol{X}} \overline{I}_0(\boldsymbol{X}; \boldsymbol{Y}) \ge \overline{I}_0(\tilde{\boldsymbol{X}}; \boldsymbol{Y}) \ge (1-\varepsilon)\left(\limsup_{n\to\infty} C_n - \varepsilon\right).$$

Finally, the proof is completed by noting that ε can be made arbitrarily small. □

Observations:

- It is known that for discrete memoryless channels, the optimistic capacity $\overline{C}$ is equal to the (conventional) capacity C [5], [14]. The same result holds for $\text{modulo} - q$ additive noise channels with stationary ergodic noise. However, in general, $\overline{C} \ge C$ since $\overline{I}_0(\boldsymbol{X}; \boldsymbol{Y}) \ge \underline{I}(\boldsymbol{X}; \boldsymbol{Y})$ [3], [4].
- Remark that [13, Theorem 11] holds if and only if

$$\sup_{\boldsymbol{X}} \underline{I}(\boldsymbol{X}; \boldsymbol{Y}) = \sup_{\boldsymbol{X}} \overline{I}_0(\boldsymbol{X}; \boldsymbol{Y}).$$

Furthermore, note that, if $\overline{C} = C$ and there exists an input distribution $P_{\hat{\boldsymbol{X}}}$ that achieves C, then $P_{\hat{\boldsymbol{X}}}$ also achieves $\overline{C}$.

V. Examples

We provide four examples to illustrate the computation of C and $\overline{C}$. The first two examples present information-stable channels for which $\overline{C} > C$. The third example shows an information-unstable channel for which $\overline{C} = C$. These examples indicate that information stability is neither necessary nor sufficient to ensure that $\overline{C} = C$ or thereby the validity of the classical source-channel separation theorem. The last example illustrates the situation where $0 < C < \overline{C} < C_{SC} < \log_2 |\mathcal{Y}|$ where C_{SC} is the channel strong capacity.[3] We assume in this section that all logarithms are in base 2 so that C and $\overline{C}$ are measured in bits.

A. Information-Stable Channels

Example 5.1: Consider a nonstationary channel $\boldsymbol{W}$ such that at odd time instances $n = 1, 3, \cdots, W^n$ is the product of the transition distribution of a binary-symmetric channel with crossover probability $1/8$ (BSC $(1/8)$), and at even time instances $n = 2, 4, 6, \cdots, W^n$ is the product of the distribution of a BSC $(1/4)$. It can be easily verified that this channel is information-stable. Since the channel is symmetric, a Bernoulli $(1/2)$ input achieves $C_n = \sup_{X^n} (1/n) I(X^n; Y^n)$; thus

$$C_n = \begin{cases} 1 - h_b(1/8), & \text{for } n \text{ odd} \\ 1 - h_b(1/4), & \text{for } n \text{ even} \end{cases}$$

where

$$h_b(a) \triangleq -a \log_2 a - (1-a) \log_2 (1-a)$$

is the binary entropy function. Therefore,

$$C = \liminf_{n\to\infty} C_n = 1 - h_b(1/4)$$

and

$$\overline{C} = \limsup_{n\to\infty} C_n = 1 - h_b(1/8) > C.$$

[3] The strong (or strong converse) capacity C_{SC} is defined [2] as the infimum of the numbers R for which there exists $\gamma > 0$ such that for all (n, M) codes with $(1/n) \log M > R - \gamma$, $\liminf_{n\to\infty} P_e^{(n)} = 1$. This definition of C_{SC} implies that for any sequence of (n, M) codes with $\liminf_{n\to\infty} (1/n) \log M > C_{SC}$, $P_e^{(n)} > 1 - \varepsilon$ for every $\varepsilon > 0$ and for n sufficiently large. It is shown in [2] that $C_{SC} = \lim_{\varepsilon\uparrow 1} \overline{C}_\varepsilon = \sup_{\boldsymbol{X}} \overline{I}(\boldsymbol{X}; \boldsymbol{Y})$.

Example 5.2: Here we use the information-stable channel provided in [13, Sec. III] to show that $\overline{C} > C$. Let $\mathcal{N}$ be the set of all positive integers. Define the set $\mathcal{J}$ as

$$\begin{aligned}\mathcal{J} &\triangleq \{n \in \mathcal{N}\colon 2^{2i+1} \le n < 2^{2i+2},\ i = 0, 1, 2, \cdots\}\\ &= \{2, 3, 8, 9, 10, 11, 12, 13, 14, 15, 32, 33, \cdots, 63,\\ &\qquad 128, 129, \cdots, 255, \cdots\}.\end{aligned}$$

Consider the following nonstationary symmetric channel $\boldsymbol{W}$. At times $n \in \mathcal{J}$, W_n is a BSC(0), whereas at times $n \notin \mathcal{J}$, W_n is a BSC$(1/2)$. Put $W^n = W_1 \times W_2 \times \cdots \times W_n$. Here again C_n is achieved by a Bernoulli$(1/2)$ input $\hat{X}^n$. We then obtain

$$\begin{aligned}C_n &= \frac{1}{n}\sum_{i=1}^{n} I(\hat{X}_i; Y_i) = \frac{1}{n}[J(n)\cdot(1) + (n - J(n))\cdot(0)]\\ &= \frac{J(n)}{n}\end{aligned}$$

where $J(n) \triangleq |\mathcal{J} \cap \{1, 2, \cdots, n\}|$. It can be shown that

$$\frac{J(n)}{n} = \begin{cases} 1 - \dfrac{2}{3} \times \dfrac{2^{\lfloor \log_2 n \rfloor}}{n} + \dfrac{1}{3n}, & \text{for } \lfloor \log_2 n \rfloor \text{ odd}\\ \dfrac{2}{3} \times \dfrac{2^{\lfloor \log_2 n \rfloor}}{n} - \dfrac{2}{3n}, & \text{for } \lfloor \log_2 n \rfloor \text{ even.}\end{cases}$$

Consequently,

$$C = \lim\inf_{n\to\infty} C_n = 1/3$$

and

$$\overline{C} = \lim\sup_{n\to\infty} C_n = 2/3.$$

B. Information-Unstable Channels

Example 5.3—The Polya-Contagion Channel: Consider a discrete additive channel with binary input and output alphabet $\{0, 1\}$ described by

$$Y_i = X_i \oplus Z_i, \qquad i = 1, 2, \cdots$$

where X_i, Y_i, and Z_i are, respectively, the ith input, ith output, and ith noise, and $\oplus$ represents modulo-2 addition. Suppose that the input process is independent of the noise process. Also assume that the noise sequence $\{Z_n\}_{n\ge1}$ is drawn according to the Polya contagion urn scheme [1], [10], as follows: an urn originally contains R red balls and B black balls with $R < B$; we make successive draws from the urn; after each draw, we return to the urn $1 + \Delta$ balls of the same color as was just drawn $(\Delta > 0)$. The noise sequence $\{Z_i\}$ corresponds to the outcomes of the draws from the Polya urn: $Z_i = 1$ if ith ball drawn is red and $Z_i = 0$, otherwise. Let $\rho \triangleq R/(R+B)$ and $\delta \triangleq \Delta/(R+B)$. It is shown in [1] that the noise process $\{Z_i\}$ is stationary and nonergodic; thus the channel is information-unstable.

From Lemma 2 and [4, Sec. IV, Pt. I], we obtain

$$1 - \overline{H}_{1-\varepsilon}(\boldsymbol{Z}) \le C_\varepsilon \le 1 - \overline{H}_{(1-\varepsilon)^-}(\boldsymbol{Z})$$

and

$$1 - \underline{H}_{1-\varepsilon}(\boldsymbol{Z}) \le \overline{C}_\varepsilon \le 1 - \underline{H}_{(1-\varepsilon)^-}(\boldsymbol{Z}).$$

It has been shown [1] that $-(1/n)\log P_{Z^n}(Z^n)$ converges in distribution to the continuous random variable $V \triangleq h_b(U)$, where U is beta-distributed $(\rho/\delta, (1-\rho)/\delta)$, and $h_b(\cdot)$ is the binary entropy function. Thus

$$\begin{aligned}\overline{H}_{1-\varepsilon}(\boldsymbol{Z}) = \overline{H}_{(1-\varepsilon)^-}(\boldsymbol{Z}) = \underline{H}_{1-\varepsilon}(\boldsymbol{Z}) = \underline{H}_{(1-\varepsilon)^-}(\boldsymbol{Z})\\ = F_V^{-1}(1-\varepsilon)\end{aligned}$$

where $F_V(a) \triangleq \Pr\{V \le a\}$ is the cumulative distribution function of V, and $F_V^{-1}(\cdot)$ is its inverse [1]. Consequently,

$$C_\varepsilon = \overline{C}_\varepsilon = 1 - F_V^{-1}(1-\varepsilon)$$

and

$$C = \overline{C} = \lim_{\varepsilon\downarrow 0} 1 - F_V^{-1}(1-\varepsilon) = 0.$$

Example 5.4: Let $\tilde{W}_1, \tilde{W}_2, \cdots$ consist of the channel in Example 5.2, and let $\hat{W}_1, \hat{W}_2, \cdots$ consist of the channel in Example 5.3. Define a new channel $\boldsymbol{W}$ as follows:

$$W_{2i} = \tilde{W}_i \quad \text{and} \quad W_{2i-1} = \hat{W}_i, \qquad \text{for } i = 1, 2, \cdots.$$

As in the previous examples, the channel is symmetric, and a Bernoulli$(1/2)$ input maximizes the inf/sup-information rates. Therefore, for a Bernoulli$(1/2)$ input $\boldsymbol{X}$, we have the equations at the bottom of this page. The fact that $-(1/i)\log[P_{Z^i}(Z^i)]$ converges in distribution to the continuous random variable $V \triangleq h_b(U)$, where U is beta-distributed $(\rho/\delta, (1-\rho)/\delta)$, and the fact that

$$\lim\inf_{n\to\infty}(1/n)J(n) = 1/3 \quad \text{and} \quad \lim\sup_{n\to\infty}(1/n)J(n) = 2/3$$

imply that

$$\begin{aligned}\underline{i}_{\boldsymbol{XW}}(\theta) &\triangleq \lim\inf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{P_{W^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \theta\right\}\\ &= 1 - F_V\left(\tfrac{5}{3} - 2\theta\right)\end{aligned}$$

and

$$\begin{aligned}\bar{i}_{\boldsymbol{XW}}(\theta) &\triangleq \lim\sup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{P_{W^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \theta\right\}\\ &= 1 - F_V\left(\tfrac{4}{3} - 2\theta\right).\end{aligned}$$

Consequently,

$$\overline{C}_\varepsilon = \frac{5}{6} - \frac{1}{2}F_V^{-1}(1-\varepsilon) \quad \text{and} \quad C_\varepsilon = \frac{2}{3} - \frac{1}{2}F_V^{-1}(1-\varepsilon).$$

$$\begin{aligned}\Pr\left\{\frac{1}{n}\log\frac{P_{W^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \theta\right\} &= \begin{cases}\Pr\left\{\dfrac{1}{2i}\left[\log\dfrac{P_{\tilde{W}^i}(Y^i|X^i)}{P_{Y^i}(Y^i)} + \log\dfrac{P_{\hat{W}^i}(Y^i|X^i)}{P_{Y^i}(Y^i)}\right] \le \theta\right\}, & \text{if } n = 2i\\ \Pr\left\{\dfrac{1}{2i+1}\left[\log\dfrac{P_{\tilde{W}^i}(Y^i|X^i)}{P_{Y^i}(Y^i)} + \log\dfrac{P_{\hat{W}^{i+1}}(Y^{i+1}|X^{i+1})}{P_{Y^{i+1}}(Y^{i+1})}\right] \le \theta\right\}, & \text{if } n = 2i+1\end{cases}\\ &= \begin{cases}1 - \Pr\left\{-\dfrac{1}{i}\log P_{Z^i}(Z^i) < 1 - 2\theta + \dfrac{1}{i}J(i)\right\}, & \text{if } n = 2i\\ 1 - \Pr\left\{-\dfrac{1}{i+1}\log P_{Z^{i+1}}(Z^{i+1}) < 1 - \left(2 - \dfrac{1}{i+1}\right)\theta + \dfrac{1}{i+1}J(i)\right\}, & \text{if } n = 2i+1.\end{cases}\end{aligned}$$

Thus

$$0 < C = \frac{1}{6} < \overline{C} = \frac{1}{3} < C_{SC} = \frac{5}{6} < \log_2 |\mathcal{Y}| = 1.$$

References

[1] F. Alajaji and T. Fuja, "A communication channel modeled on contagion," *IEEE Trans. Inform. Theory*, vol. 40, pp. 2035–2041, Nov. 1994.

[2] P.-N. Chen and F. Alajaji, "Strong converse, feedback capacity and hypothesis testing," in *Proc. CISS* (John Hopkins Univ., Baltimore, MD, Mar. 1995).

[3] ——, "Generalization of information measures," in *Proc. Int. Symp. Information Theory and Applications* (Victoria, BC, Canada, Sept. 1996).

[4] ——, "Generalized source coding theorems and hypothesis testing," *J. Chinese Inst. Eng.*, vol. 21, no. 3, pp. 283–303, May 1998.

[5] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic, 1981.

[6] R. L. Dobrushin, "General formulation of Shannon's basic theorems of information theory," *AMS Transl.* (AMS, Providence, RI), vol. 33, pp. 323–438, 1963.

[7] T. S. Han, *Information-Spectrum Methods in Information Theory.* Tokyo, Japan: Baifukan, 1998, in Japanese.

[8] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. 39, pp. 752–772, May 1993.

[9] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes.* San Francisco, CA: Holden-Day, 1964.

[10] G. Polya, "Sur quelques points de la théorie des probabilités," *Ann. Inst. H. Poincarré*, vol. 1, pp. 117–161, 1931.

[11] Y. Steinberg, "New converses in the theory of identification via channels," *IEEE Trans. Inform. Theory*, vol. 44, pp. 984–998, May 1998.

[12] Y. Steinberg and S. Verdú, "Simulation of random processes and rate-distortion theory," *IEEE Trans. Inform. Theory*, vol. 42, pp. 63–86, Jan. 1996.

[13] S. Vembu, S. Verdú, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inform. Theory*, vol. 41, pp. 44–54, Jan. 1995.

[14] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.

Gaussian Codes and Shannon Bounds for Multiple Descriptions

Ram Zamir, *Senior Member, IEEE*

***Abstract*— A pair of well-known inequalities due to Shannon upper/lowerbound the rate-distortion function of a real source by the rate-distortion function of the Gaussian source with the same variance/entropy. We extend these bounds to multiple descriptions, a problem for which a general "single-letter" solution is not known. We show that the set $\mathcal{D}_X(R_1, R_2)$ of achievable marginal (d_1, d_2) and central (d_0) mean-squared errors in decoding X from two descriptions at rates R_1 and R_2 satisfies**

$$\mathcal{D}^*(\sigma_x^2, R_1, R_2) \subseteq \mathcal{D}_X(R_1, R_2) \subseteq \mathcal{D}^*(P_x, R_1, R_2)$$

where σ_x^2 and P_x are the variance and the entropy-power of X, respectively, and $\mathcal{D}^*(\sigma^2, R_1, R_2)$ is the multiple description distortion region for a Gaussian source with variance σ^2 found by Ozarow. We further show that like in the single description case, a Gaussian random code achieves the *outer* bound in the limit as $d_1, d_2 \to 0$, thus the outer bound is asymptotically tight at high resolution conditions.

***Index Terms*—Gaussian codes, high resolution, multiple descriptions, Shannon lower bound.**

I. Introduction

The multiple description problem [6] arises in communicating analog source information (speech, image, video) via lossy packet networks. In this increasingly frequent scenario, a source code is broken into a few packets, some of which may not arrive at the destination. Suppose the network does not support retransmission of lost packets (due, e.g., to restrictions on delay, loading, or capacity of feedback channel). In such a case, the decoder wishes to achieve a certain basic reproduction quality if a small subset of the packets arrives, and an improved quality if more packets or the whole source code arrives. Thus portions of various size of the code should contain individually good, complementary descriptions of the source.

The basic formulation of the multiple description problem in the information-theoretic literature involves two (noiseless) subchannels of rates R_1 and R_2, corresponding to two "packets," and three receivers. Each receiver corresponds to a possible case of packet arrival, the first arrived, the second arrived, or both arrived, as depicted in Fig. 1. In response to a source block $\boldsymbol{x} = (x_1, \cdots, x_n)$, the encoder generates two codewords (indices) $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ at rates

$$R_i = \frac{1}{n} \log |f_i(\cdot)|, \qquad i = 1, 2$$

where $|f_i|$ denotes the size of code $f_i(\cdot)$, and transmits codeword f_i through subchannel i, $i = 1, 2$. The two individual ("marginal") receivers and the combined ("central") receiver then generate reconstructions $\hat{\boldsymbol{x}}_1$, $\hat{\boldsymbol{x}}_2$, and $\hat{\boldsymbol{x}}_0$, respectively, using the decoding functions

Manuscript received August 1, 1997; revised April 2, 1999. This work was supported by the ISF, administered by the Israel Academy of Science and Humanities. The material in this correspondence was presented in part at the Maine Conference, Portland, ME, July 1997, and at the International Symposium on Information Theory, Cambridge, MA, August 1998.

The author is with the Department of Electrical Engineering–Systems, Tel-Aviv University, Tel-Aviv 69978, Israel.

Communicated by R. Laroia, Associate Editor for Source Coding.
Publisher Item Identifier S 0018-9448(99)07676-2.