The Source–Channel Separation Theorem Revisited

Sridhar Vembu, Sergio Verdú, *Fellow, IEEE,* and Yossef Steinberg

Abstract— **The single-user separation theorem of joint source–channel coding has been proved previously for wide classes of sources and channels. We find an information-stable source/channel pair which does not satisfy the separation theorem. New necessary and sufficient conditions for the transmissibility of a source through a channel are found, and we characterize the class of channels for which the separation theorem holds regardless of the source statistics.**

Index Terms—**Shannon Theory; channel capacity, source coding; joint source-channel coding; separation theorem.**

I. Introduction

The MEETING point of the two main branches of the Shannon theory is the joint source–channel coding theorem. This theorem has two parts: a direct part that states that if the minimum achievable source coding rate of a given source is strictly below the capacity of a channel, then the source can be reliably transmitted through the channel by appropriate encoding and decoding operations; and a converse part stating that if the source coding rate is strictly greater than capacity, then reliable transmission is impossible. Implicit in the direct source–channel coding theorem is the fact that reliable transmission can be accomplished by separate source and channel coding, where the source (resp., channel) encoder and decoder need not take into account the channel (resp., source) statistics. Because of the converse theorem (and except for the residual uncertainty in the case when the minimum source coding rate is equal to the channel capacity) it follows that either reliable transmission is possible by separate source–channel coding or it is not possible at all. This is the reason why the joint source–channel coding theorem is commonly referred to as the separation theorem.

Ever since Claude Shannon's 1948 paper [1], where the result was stated for stationary memoryless sources and channels, the separation theorem has received considerable attention, with a number of researchers proving versions that apply to more and more general classes of sources and channels. Dobrushin [2] and Hu [4] considered the separation theorem in the context of information stable sources and channels, i.e., situations where, essentially, the Asymptotic Equipartition Property (AEP) is satisfied and minimum achievable source coding rate and channel capacity are equal to the entropy rate and the limit of maximal mutual information rates, respectively. In addition, joint source–channel coding has been a main focus of ergodic–theoretic researchers in Shannon theory, who have obtained general expressions for the maximal entropy rate of those sources that can be transmitted reliably through a given channel. In the foregoing discussion, reliable transmission of the source through the channel means that the probability of correctly decoding a block of n transmitted symbols goes to 1 as $n \to \infty$. Since the works by Kolmogorov [6], Shannon [7], and Dobrushin [2], the separation theorem has also been investigated in the context of transmission with a distortion measure. In this paper we focus exclusively on the aforementioned reliability criterion of "almost noiseless" fixed-length block coding.

Even though most analytically tractable channels and sources are encompassed by previous versions of the separation theorem, it is of considerable theoretical interest to study the validity of this theorem in the context of very general sources and channels. In particular, we do not impose restrictions such as memorylessness, stationarity, ergodicity, causality, information stability, etc. This is motivated by the recent papers [13] and [14], which find general expressions for the minimal source coding rate and channel capacity that apply without those restrictions. A result [14, Theorem 4] which leads to a new general converse to the channel coding theorem in [14], proves to be a key tool in our investigation of the source–channel coding theorem. Despite the generality of [13], [14] and the present paper, the proofs are, in fact, conceptually simple.

After a review of definitions and previous results in Section II, we show an example in Section III where the converse to the separation theorem fails to hold: a memoryless information stable source/channel pair such that the source is transmissible through the channel (with zero error), yet its minimum achievable source coding rate is twice the channel capacity. We note that previous instances where the separation theorem was known to fail were always within the context of multiterminal sources and channels (e.g., [8]). The example in Section III reveals that, in general, the channel capacity and the minimum source coding rate do not provide sufficient knowledge in order to determine whether the source can be transmitted reliably through the channel. A finer look at the statistical structure of the channel and source is necessary. This is done in Section IV where two similar conditions, domination and strict domination, are shown to be necessary and sufficient,

Manuscript received July 20, 1993; revised January 31, 1994. This research has been partially supported by the National Science Foundation under PYI award ECSE-8857689 and by and IBM Graduate Fellowship award. This paper was presented at the 1994 International Symposium on Information Theory, Trondheim, Norway, June 1994.

S. Vembu was with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA. He is now with Qualcom Inc., San Diego, CA 92121 USA.

S. Verdú is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA.

Y. Steinberg was with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA. He is now with the Department of Electrical and Computer Engineering, Ben Gurion University of the Negev, Beer-Sheva 84105, Israel.

IEEE Log Number 9407436.

0018–9448/95$04.00

respectively, for reliable transmissibility. In Section V we characterize those channels for which the classical statement of the separation theorem holds for every source. It turns out that those are the channels whose definition of capacity is insensitive to whether good codes are required for all sufficiently long blocklengths or for only infinitely many blocklengths. We also characterize those sources for which the separation theorem holds for every channel. This class of sources includes but is not restricted to stationary sources. A conclusion to be drawn from our results is that when dealing with nonstationary probabilistic models, care should be exercised before applying the separation theorem.

II. Preliminaries

A. Definitions and Classical Results

Let F, A, and B be finite sets. A source $\boldsymbol{Z}$ with alphabet F is the sequence $\{P_{Z^n}\}_{n\geq 1}$, where P_{Z^n} is a probability distribution on F^n. Similarly, a channel $\boldsymbol{W}$ with input alphabet A and output alphabet[1] B is a sequence of conditional distributions $\{W^n(\cdot|\cdot)\}_{n\geq 1}$ such that $W^n(\cdot|a^n)$ is a probability distribution on B^n for every $a^n \in A^n$. Given an A-valued source $\boldsymbol{X}$, and a channel $\boldsymbol{W}$, we denote the joint source whose finite-dimensional distributions are $W^n P_{X^n}$ by $(\boldsymbol{X}, \boldsymbol{Y})$.

Definition 1: Given a joint distribution $P_{X^nW^n}$ on $A^n \times B^n$ with marginals P_{X^n}, P_{Y^n} the *information density* is the function

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

The distribution of the random variable $(1/n)i_{X^nW^n}(a^n; b^n)$ is referred to as the *information spectrum* of $P_{X^nW^n}$, [13], and the expected value of the information spectrum is the normalized mutual information $(1/n)I(X^n; Y^n)$. The mutual information rate of $(\boldsymbol{X}, \boldsymbol{Y})$ is defined as

$$\boldsymbol{I}(\boldsymbol{X}; \boldsymbol{Y}) = \lim_{n\to\infty} \frac{1}{n} I(X^n; Y^n)$$

provided that the limit exists. In case that $\boldsymbol{X}$ is equal to $\boldsymbol{Y}$ (i.e., $\boldsymbol{W}$ is an identity channel), the information density $i_{X^nW^n}(a^n; b^n)$ is referred to as the *entropy density*, which is given by

$$h_{X^n}(a^n) = \log \frac{1}{P_{X^n}(a^n)}.$$

The expected value of $(1/n)h_{X^n}(X^n)$ is the normalized entropy $(1/n)H(X^n)$, and the entropy rate of $\boldsymbol{X}$ is

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

provided that the limit exists. We proceed now to the classical definition of information stability of sources and channels. The concept of information stability was first introduced by Dobrushin [2].

[1] The assumption that the channel alphabets are finite is made for notational convenience. It can be readily lifted without impacting our results or their proofs.

Definition 2 [2], [3]: A source $\boldsymbol{Z}$ is said to be *information-stable* if $H(Z^n) > 0$ for all sufficiently large n, and $h_{Z^n}(Z^n)/H(Z^n)$ converges in probability to one as $n \to \infty$, i.e.,

$$\lim_{n\to\infty} P\left(\left|\frac{h_{Z^n}(Z^n)}{H(Z^n)} - 1\right| > \epsilon\right) = 0 \tag{1}$$

for every $\epsilon > 0$.

This definition is slightly more general than the AEP [16] which requires that $(1/n)h_{Z^n}(Z^n)$ converges in probability to $\lim_{n\to\infty}(1/n)H(Z^n)$. The Shannon–McMillan theorem implies that stationary and ergodic finite alphabet sources are information-stable.

Definition 3 [2], [4]: A channel $\boldsymbol{W}$ is called *information-stable* if there exists an input process $\boldsymbol{X}$ such that

$$\frac{i_{X^nW^n}(X^n; Y^n)}{nC_n(W^n)} \to 1$$

where the convergence is in probability, and

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

The reason for defining information stability as in Definition 2 lies in the fact that information stability is precisely the condition under which C_n has an operational meaning asymptotically. Before stating the classical results on source coding, channel capacity, and source–channel separation theorem, we record the standard definitions of the minimal achievable (fixed-length) source coding rate and channel capacity.

Definition 4: An (n, M, ϵ) fixed-length source code for Z^n is a collection of M n-tuples $\{a_1^n, \cdots, a_M^n\}$ such that

$$P(Z^n \notin \{a_1^n, \cdots, a_M^n\}) \leq \epsilon.$$

R is and ϵ-achievable source coding rate for $\boldsymbol{Z}$ if for every $\gamma > 0$ there exist, for all sufficiently large n, (n, M, ϵ) codes with

$$\frac{1}{n}\log M < R + \gamma.$$

R is an *achievable (fixed-length) source coding rate for* $\boldsymbol{Z}$ if it is ϵ-achievable for all $\epsilon > 0$. The *minimal achievable source coding rate* of $\boldsymbol{Z}$ is denoted by $T(\boldsymbol{Z})$.

Definition 5: An (n, M, ϵ) code for a random transformation W^n with input alphabet A and output alphabet B is a pair of mappings

$$f: \{1, 2, \cdots, M\} \to A^n$$

$$g: B^n \to \{1, 2, \cdots, M\}$$

such that

$$\frac{1}{M}\sum_{m=1}^{M} \sum_{b^n:\, g(b^n)\neq m} W^n(b^n|f[m]) \leq \epsilon.$$

The mappings f and g are referred to as the encoder and decoder, respectively.

Definition 6: R is an ϵ-achievable rate for $\boldsymbol{W}$ if for every $\gamma > 0$ there exists, for all sufficiently large n, an (n, M, ϵ) code for W^n with

$$\frac{1}{n} \log M > R - \gamma.$$

The maximum ϵ-achievable rate for a channel $\boldsymbol{W}$ is called the ϵ-capacity, $C_\epsilon(W)$, of the channel. The *channel capacity*[2], $C(\boldsymbol{W})$, is defined as the maximal rate that is ϵ-achievable for every $\epsilon > 0$. The definition implies that

$$C(\boldsymbol{W}) = \lim_{\epsilon \to 0} C_\epsilon(\boldsymbol{W})$$

and that $C(\boldsymbol{W})$ is the supremum of all the rates R for which there exist a sequence of (n, M, ϵ_n) codes such that

$$\frac{\log M}{n} > R$$

and

$$\lim_{n \to \infty} \epsilon_n = 0.$$

In the definition of capacity we have used the *average probability of error*. However, capacity remains unchanged if we use the *maximal probability of error*. This is due to the fact that the existence of (n, M, ϵ) codes according to the average probability of error criterion implies the existence of $(n, M/2, 2\epsilon)$ codes in the maximal probability of error criterion.

Definition 7: Let $\boldsymbol{Z}$ be a source with alphabet F, and let $\boldsymbol{W}$ be a channel with input alphabet A and output alphabet B. $\boldsymbol{Z}$ is said to be reliably transmissible over $\boldsymbol{W}$ if there exist a sequence of encoders $\{f_n(\cdot)\}_{n \geq 1}$

$$f_n \colon F^n \to A^n$$

and a sequence of decoders $\{g_n(\cdot)\}_{n \geq 1}$

$$g_n \colon B^n \to F^n$$

such that

$$\lim_{n \to \infty} P(Z^n \neq \tilde{Z}^n) = 0$$

where $\tilde{Z}^n$ is the output due to Z^n of the cascade encoder–channel–decoder.

We briefly review the well-known results on source coding and transmission. The following theorems are not the most general statements that can be deduced from the results in [2], [4], but give a flavor of what is implied by the results in these works with regard to a separation theorem.

Theorem 1: Every information stable source $\boldsymbol{Z}$ satisfies

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

Theorem 2: Every information stable channel $\boldsymbol{W}$ satisfies

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

If the liminf in Theorem 2 is actually a limit, the following joint source/channel coding theorem follows.

[2]The explicit dependence on $\boldsymbol{W}$ will be omitted when convenient.

Theorem 3: Let $\boldsymbol{Z}$ and $\boldsymbol{W}$ be an information-stable source and an information-stable channel, respectively. Assume that $\lim_{n \to \infty} C_n(W^n)$ exists (cf. Definition 3). Then

a) $T(\boldsymbol{Z}) < C(\boldsymbol{W})$ implies that $\boldsymbol{Z}$ can be reliably transmitted over $\boldsymbol{W}$.

b) If $\boldsymbol{Z}$ can be reliably transmitted over $\boldsymbol{W}$, then $T(\boldsymbol{Z}) \leq C(\boldsymbol{W})$.

The restriction on the classes of sources and channels in [4] are somewhat more general than those in Theorem 3; however, they are hard to verify. A summary and further discussion of those conditions can be found in [5]. Another well-known body of results on the source–channel transmission problem is ergodic theoretic in nature. Tutorial discussions on the ergodic theoretic approach to the separation theorem can be found in [9], [10].

For every channel we can define

$$C_E(\boldsymbol{W}) \triangleq \sup_{\boldsymbol{Z} \text{ ergodic, block transmissible}} \boldsymbol{H}(\boldsymbol{Z}).$$

It was proved by Kieffer [11] that for a *weakly continuous ergodic* channel (see [11] for definitions) C_E equals the channel capacity C, and that C equals the supremum of mutual information rate over all stationary inputs. This amounts to the separation theorem for this class of channels and ergodic sources. Kieffer [11] also shows that for the class of weakly continuous stationary channels, C_E equals the *information quantile capacity* introduced by Winkelbauer [12].

B. General Sources and Channels

In this subsection we describe the results on source coding and channel capacity obtained in [13], [14] for general sources and channels. We start with a few definitions.

Definition 8 [13]: The *limsup in probability* of a sequence of random variables $\{A_n\}$ is defined as the smallest extended real number β such that for all $\epsilon > 0$

$$\lim_{n \to \infty} P[A_n \geq \beta + \epsilon] = 0.$$

Analogously, the *liminf in probability* is the largest extended real number α such that for all $\epsilon > 0$

$$\lim_{n \to \infty} P[A_n \leq \alpha - \epsilon] = 0.$$

Note that a sequence of random variables converges in probability to a constant if and only if its limsup in probability is equal to its liminf in probability. The *limsup in probability* (resp., *liminf in probability*) of the sequence of random variables $\{(1/n) i_{X^n W^n}(X^n, Y^n)\}_{n=1}^{\infty}$ is referred to as the *sup-information rate* (resp., *inf-information rate*) of the pair $(\boldsymbol{X}, \boldsymbol{Y})$ and is denoted as $\overline{\boldsymbol{I}}(\boldsymbol{X}; \boldsymbol{Y})$ (resp., $\underline{\boldsymbol{I}}(\boldsymbol{X}; \boldsymbol{Y})$). The limsup (resp., liminf) in probability of the sequence of random variables $\{(1/n) h_{X^n}(X^n)\}_{n=1}^{\infty}$ is referred to as the *sup* (resp., *inf*)*-entropy rate* of $\boldsymbol{X}$ and is denoted by $\overline{\boldsymbol{H}}(\boldsymbol{X})$ (resp., $\underline{\boldsymbol{H}}(\boldsymbol{X})$).

We are in a position to state the general source coding and channel capacity results of [13], [14].

Theorem 4 [13]: For any finite alphabet source $\boldsymbol{Z}$

$$T(\boldsymbol{Z}) = \overline{\boldsymbol{H}}(\boldsymbol{Z}).$$

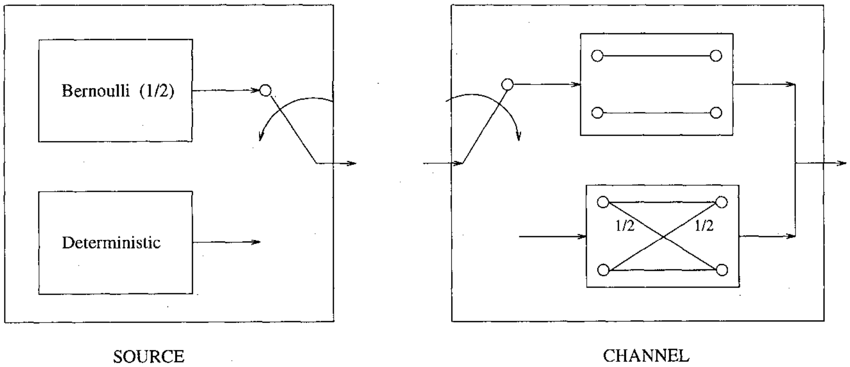

Fig. 1. A source and channel that do not satisfy the separation theorem.

Theorem 5 [14]: The capacity $C(\boldsymbol{W})$ of any finite alphabet channel $\boldsymbol{W}$ is given by

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

where $\boldsymbol{Y}$ is the output of $\boldsymbol{W}$ to $\boldsymbol{X}$.

In view of Theorems 4 and 5, we have the direct part of the separation theorem that holds for general sources and channels:

Theorem 6: If a given source $\boldsymbol{Z}$ and channel $\boldsymbol{W}$ satisfy

$$T(\boldsymbol{Z}) < C(\boldsymbol{W})$$

then $\boldsymbol{Z}$ can be reliably transmitted over $\boldsymbol{W}$.

Proof: Let us assume that $T + \delta < C$ for some $\delta > 0$. From Theorem 4, we can find a sequence of $(n, \exp\{n(T + (\delta/2))\}, \epsilon_n)$ source codes for $\boldsymbol{Z}$ with $\epsilon_n \to 0$. Furthermore, Theorem 5 implies that we can find a sequence of $(n, \exp\{n(T+(\delta/2))\}, \gamma_n)$ channel codes for $\boldsymbol{W}$ with $\gamma_n \to 0$.

We define the encoder and decoder pair needed in Definition 7 as follows: the encoder consists of the cascade of the source encoder and the channel encoder. The decoder is the cascade of the channel decoder and the source decoder. The overall probability of error is upper bounded by $\epsilon_n + \gamma_n$ which goes to zero. Note that we have proved not just that $\boldsymbol{Z}$ is transmissible over $\boldsymbol{W}$, but also that the encoder can be split into a source encoder and a channel encoder and these functions can be performed independently. This is the crux of the separation principle. ∎

A completely general separation theorem would follow if we could prove that

$$T(\boldsymbol{Z}) \le C(\boldsymbol{W})$$

is a necessary condition for reliable transmission of $\boldsymbol{Z}$ over $\boldsymbol{W}$ to be possible. This, however, is not true: in Section III we give an example that disproves such a statement.

III. An Example Where Separation Does Not Hold

In this section we construct an example of a nonstationary source $\boldsymbol{Z}$ and a nonstationary channel $\boldsymbol{W}$ over which $\boldsymbol{Z}$ can be reliably transmitted, yet $\overline{\boldsymbol{H}}(\boldsymbol{Z}) > C(\boldsymbol{W})$.

The source and channel pair is given in Fig. 1. Note that the switches move in synchronism—either both are "up" position or both are "down." The switches deterministically change position right before times 2^i, $i = 1, 2, 3, \cdots$. Note that both the source and channel are memoryless.

Let J denote the set of times at which the switch is in the "up" position. Assuming that the switch is "down" at time $i = 1$, we have

$$J = \{2, 3, 8, 9, 10, 11, 12, 13, 14, 15, 32, 33, \cdots, 62, 63, 128, 129, \cdots\}.$$

At times $i \in J$, the binary source is independently equally likely $\{0, 1\}$, and the channel is noiseless, whereas at times $i \notin J$, the source is deterministic and the channel output is independent of the input. Since the set J is deterministic and known to both transmitter and receiver, $\boldsymbol{Z}$ can be transmitted over $\boldsymbol{W}$ with zero probability of error. However, as we shall show, $T(\boldsymbol{Z}) = 2/3$ and $C(\boldsymbol{W}) = 1/3$.

To evaluate $\overline{\boldsymbol{H}}(\boldsymbol{Z})$, write

$$\frac{1}{n} \log P_{Z^n}(z^n) = \frac{1}{n} \sum_{i=1}^{n} \log P_{Z_i}(z_i)$$

and observe that $\log P_{Z_i}(Z_i)$ is deterministic, attaining the value -1 bit for $i \in J$ and 0 for $i \notin J$. (For convenience, the logarithms in this section have base 2.) Thus it is straightforward to verify that

$$\overline{\boldsymbol{H}}(\boldsymbol{Z}) = \limsup_{n \to \infty} \frac{J(n)}{n} = \frac{2}{3} \tag{2}$$

where $J(n)$ stands for the cardinality of the intersection of J with the set $\{1, 2, \cdots, n\}$, i.e.,

$$J(n) \triangleq |J \cap \{1, 2, \cdots, n\}|.$$

We proceed to the evaluation of the capacity of $\boldsymbol{W}$. It was shown in [14, sec. 7] that the capacity of a memoryless binary symmetric channel is given by

$$C(\boldsymbol{W}) = 1 - \overline{\boldsymbol{H}}(\boldsymbol{N})$$

where N denotes the random process of errors. Following the arguments leading to (2), we conclude that

$$\overline{\boldsymbol{H}}(\boldsymbol{N}) = \limsup_{n \to \infty} 1 - \frac{J(n)}{n}.$$

Thus

$$C(\boldsymbol{W}) = \liminf_{n \to \infty} \frac{J(n)}{n} = \frac{1}{3}. \tag{3}$$

It is interesting to note that both the source and the channel in this example are not only memoryless but information-stable. To check this, note that nonzero probability source strings of a given length all have the same probability; thus (1) follows immediately. Analogously, the information stability of the channel can be checked by noticing that Bernoulli $(\frac{1}{2})$ inputs achieve maximal input–output mutual information, and with that choice, every pair of input–output strings of a given length has the same probability or cannot occur at all. Separation need not hold because $C_n(W^n)$ does not have a limit (cf. Theorem 3).

IV. A General Transmission Theorem

In view of the example above, it is clear that the minimum source-coding rate and the channel capacity are inadequate to completely characterize the problem of reliable transmission of a source over a channel. In this section we present a general transmission theorem that serves to give an answer to this question.

A. Notions of Domination

We will now define two related notions, which we call *strict domination* and *domination.*

Definition 9: A channel $\boldsymbol{W}$ is said to *strictly dominate* a source $\boldsymbol{Z}$ if there exists a $\delta > 0$ and a channel input process $\boldsymbol{X}$ such that

$$\lim_{n\to\infty} \inf_{c_n \in R} \left\{ P\left[\frac{1}{n} h_{Z^n}(Z^n) \geq c_n\right] + P\left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n) \leq c_n + \delta\right]\right\} = 0.$$

The above definition captures the notion that asymptotically the information spectrum of the channel lies entirely above the spectrum of the source. For example, if the sup-entropy rate of the source and the channel capacity satisfy $\overline{\boldsymbol{H}}(\boldsymbol{Z}) < C(\boldsymbol{W})$, strict domination holds. Note that strict domination does not necessarily mean that there exists a fixed scalar such that the channel information spectrum is asymptotically above it while the source spectrum is asymptotically below it. Such a case would be included in this definition but that is not the only way a channel can strictly dominate a source.

We will define a closely related notion called *domination.*

Definition 10: A channel $\boldsymbol{W}$ is said to *dominate* a source $\boldsymbol{Z}$ if for any $\delta > 0$ and any sequence of nonnegative numbers $\{c_n\}_{n=1}^{\infty}$, there exists $\boldsymbol{X}$ such that

$$\lim_{n\to\infty} P\left[\frac{1}{n} h_{Z^n}(Z^n) \geq c_n\right] \cdot P\left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n) \leq c_n - \delta\right] = 0.$$

The gist of both definitions is that the upper tail of the source information spectrum (i.e., the distribution of the normalized entropy density) has vanishing overlap with the lower tail of the channel information spectrum (distribution of the normalized information density) evaluated with a favorable input process. Domination compares the tails in nonoverlapping intervals $[c_n, +\infty)$ and $(-\infty, c_n - \delta]$, and forbids that both tails be nonvanishing for any sequence c_n. Strict domination compares the tails in overlapping intervals $[c_n, +\infty)$ and $(-\infty, c_n + \delta]$, and dictates that both tails vanish provided $\{c_n\}$ is suitably selected.

As we will see, domination (resp., strict domination) will take the role of the condition that the minimum source coding rate is *greater than or equal to* (resp., *strictly greater than*) the capacity.

B. Transmission Theorem

The following result[3] is a generalization of the well-known source–channel separation theorem. In the classical theory, the case when capacity equals the minimum source coding rate is left unresolved. We have the same ambiguity here and that is the reason for two definitions in Section IV-A.

Theorem 7:

i) Reliable transmission for a source–channel pair is possible if the channel strictly dominates the source.

ii) If reliable transmission is possible for a source–channel pair, then the channel dominates the source.

Proof:

i) Let us assume that channel $\boldsymbol{W}$ strictly dominates source $\boldsymbol{Z}$. From the definition of strict domination, we can find a $\delta > 0$ and a channel input process $\boldsymbol{X}$ such that

$$\lim_{n\to\infty} \inf_{c_n \in R} \left\{ P\left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n) < c_n + \delta\right] + P\left[\frac{1}{n} h_{Z^n}(Z^n) \geq c_n\right]\right\} = 0.$$

This means that there exists a sequence $\{c_n\}_{n=1}^{\infty}$ such that

$$P\left[\frac{1}{N} i_{X^n W^n}(X^n; Y^n) < c_n + \delta\right] \leq \tau_n$$

and

$$P\left[\frac{1}{n} h_{Z^n}(Z^n) \geq c_n\right] \leq \tau_n$$

where τ_n goes to zero as n goes to infinity.

Now we will apply Feinstein's Lemma (see, for example, [17]) to claim that there exists a sequence of (n, M, ϵ_n) codes for the channel where $M = \exp\{n(c_n + (\delta/2))\}$ and ϵ_n is upper bounded by $\tau_n + \exp(-n\delta/2)$ which goes to zero with n.

We will only encode the source words that fall within the set

$$E_n = \left\{z^n \colon \frac{1}{n} h_{Z^n}(z^n) < c_n\right\}$$

whose probability goes to 1 with n. When an outcome in E_n^c occurs we declare error.

Clearly, E_n contains fewer than $\exp(nc_n)$ elements and they can be transmitted with probability of error at most ϵ_n. Hence the overall probability of error is upper-bounded by $\epsilon_n + \tau_n$ which goes to zero. This establishes the direct part.

[3] With suitable modifications to the definitions of strict domination and domination, it is possible to prove a straightforward extension of Theorem 7 to the case where the source words and the channel codewords do not have the same blocklength; see [5].

ii) Recall from Definition 7 that reliable transmission of $\boldsymbol{Z}$ over $\boldsymbol{W}$ implies the existence of a sequence of encoders $\{f_n\}$ and decoders $\{g_n\}$ such that

$$\sum_{z^n} P_{Z^n}(z^n)V^n(g_n^{-1}(z^n)|z^n) \to 1$$

where we have used the notation

$$V^n(y^n|z^n) = W^n(y^n|f_n(z^n)).$$

Furthermore, given a channel and a sequence of encoders, we will denote the cascade encoder channel by $\boldsymbol{V}$. The following result gives a very simple property satisfied by the information spectrum of the cascade of a deterministic transformation and a random transformation.

Lemma 1: Fix n. For any P_{Z^n}, deterministic transformation f_n and random transformation W^n, the information densities $i_{Z^nV^n}(Z^n; Y^n)$ and $i_{X^nW^n}(X^n; Y^n)$ are identical, where $X^n = f_n(Z^n)$ and $V^n(\cdot|z^n) = W^n(\cdot|f_n(z^n))$.

Proof of Lemma 1: Let $\mathbb{1}$ stand for the indicator function. Note that

$$\begin{aligned} &P[i_{Z^nV^n}(Z^n; Y^n) \le c] \\ &= \sum_{z^n}\sum_{y^n} P_{Z^n}(z^n)V(y^n|z^n) \\ &\quad \cdot \mathbb{1}\left\{\log\frac{V(y^n|z^n)}{P_{Y^n}(y^n)} \le c\right\} \\ &= \sum_{z^n}\sum_{y^n}\sum_{x^n} P_{Z^n}(z^n)\mathbb{1}\{x^n = f(z^n)\}W(y^n|x^n) \\ &\quad \cdot \mathbb{1}\left\{\log\frac{W(y^n|x^n)}{P_{Y^n}(y^n)} \le c\right\} \\ &= P[i_{X^nW^n}(X^n; Y^n) \le c] \end{aligned}$$

where the second equation is due to the fact that $V(y^n|z^n) = W(y^n|f_n(z^n))$. ■

Now we will show that the reliable transmission of $\boldsymbol{Z}$ implies that $\boldsymbol{V}$ dominates $\boldsymbol{Z}$. Due to Lemma 1 it follows immediately that $\boldsymbol{Z}$ is dominated by $\boldsymbol{W}$ as well, which is the desired result.

Lemma 2: Reliable transmissibility of $\boldsymbol{Z}$ over $\boldsymbol{W}$ implies that the cascade channel $\boldsymbol{V}$ dominates $\boldsymbol{Z}$.

Proof of Lemma 2: We will verify the condition of domination in Definition 10. Let us fix an arbitrary subsequence K of the sequence of positive integers. Let us consider a sequence $\{c_n\}_{n\in K}$ of nonnegative real numbers such that the sequence of sets $\{B_n\}_{n\in K}$

$$B_n = \left\{z^n\colon \frac{1}{n}h_{Z^n}(z^n) \ge c_n\right\}$$

satisfies the condition

$$\lim_{n\to\infty,\, n\in K} P_{Z^n}(B_n) > 0. \tag{4}$$

Hence we can find some $\alpha > 0$ such that for all large $n \in K$

$$P_{Z^n}(B_n) > \alpha. \tag{5}$$

Henceforth we will concentrate on those $n \in K$ for which (5) is true.

We know that there exists a sequence of encoder–decoder pairs such that the probability of error

$$\epsilon_n = \sum_{z^n} e(z^n)P_{Z^n}(z^n)$$

goes to zero with increasing $n \in K$. We have used $e(z^n)$ to denote the probability of error when the input is z^n, i.e.,

$$e(z^n) = P[\tilde{Z}^n \ne z^n | Z^n = z^n]$$

where $\tilde{Z}^n$ is the decoder output.

Let us define the following set:

$$D_n = \left\{z^n \in B_n\colon e(z^n) > \frac{2\epsilon_n}{\alpha}\right\}.$$

Since $E[e(Z^n)\mathbb{1}\{Z^n \in B_n\}] \le \epsilon_n$

$$P_{Z^n}(D_n) \le \frac{\alpha}{2}. \tag{6}$$

It follows from (5) and (6) that the set $G_n = B_n \backslash D_n$ has probability at least $\alpha/2$. Moreover, for each $z^n \in G_n$, $e(z^n) \le 2\epsilon_n/\alpha$.

Let us fix an arbitrary $\delta > 0$. The cardinality of G_n is bounded as

$$\begin{aligned} |G_n| &> \frac{\alpha}{2}\exp\{nc_n\} \\ &> \exp\{n(c_n - \delta)\} \end{aligned}$$

provided n is large.

Therefore, we have found a set G_n which has no fewer than $\exp\{n(c_n - \delta)\}$ elements and the probability of error when any of these elements is used as a codeword is at most $2\epsilon_n/\alpha$, which goes to zero as n increases in K. In other words, for the cascade channel $\boldsymbol{V}$, we have found a sequence of $(n, \exp\{c(c_n - \delta)\}, 2\epsilon_n/\alpha)$ codes for large $n \in K$.

Now we are in a position to invoke the central result in the new converse to the coding theorem proved in [14]. This result gives a simple lower bound to the average probability of error of any (n, M, ϵ) code.

Lemma 3 [14]: Every (n, M, ϵ) code for a conditional distribution $\hat{W}^n$ satisfies

$$\epsilon \ge P\left[\frac{1}{n}i_{\hat{X}^n\hat{W}^n}(\hat{X}^n; \hat{Y}^n) \le \frac{1}{n}\log M - \gamma\right] - \exp(-\gamma n)$$

for every $\gamma > 0$, where $\hat{X}^n$ places probability mass $1/M$ on each codeword.

Taking $M = \exp\{n(c_n - \delta)\}$, we know that there exist a sequence of $(n, M, 2\epsilon_n/\alpha)$ codebooks for large $n \in K$ for the channel $\boldsymbol{V}$. Let us denote the distribution that puts equal probability mass on each of these codewords as Z_c^n. Applying Lemma 3 with $\gamma = \delta$ to this sequence of codebooks, we get

$$\frac{2\epsilon_n}{\alpha} \ge P\left[\frac{1}{n}i_{Z_c^nV^n}(Z_c^n; Y_c^n) \le c_n - 2\delta\right] - \exp(-n\delta). \tag{7}$$

where Y_c^n is the output of $\boldsymbol{V}$ due to input Z_c^n. Since ϵ_n goes to zero as n goes to infinity in K, we get

$$\lim_{n\to\infty,\, n\in K} P\left[\frac{1}{n}i_{Z_c^nV^n}(Z_c^n; Y_c^n) \le c_n - 2\delta\right] = 0. \tag{8}$$

Since the choice of $\delta > 0$ and the subsequence K satisfying (4) are arbitrary we have shown that the channel $\boldsymbol{V}$ dominates $\boldsymbol{Z}$. This concludes the proof of Lemma 2 and Theorem 7. ∎

Remark: Theorem 7 can be strengthened by replacing δ in the definition of strict domination and domination by a vanishing sequence $\{\delta_n\}$ such that $n\delta_n \to \infty$.

To illustrate an application of Theorem 7, we now state the following lemma, which immediately leads to Theorem 3 whose proof is in the Appendix. In Section V, we derive more general results, dispensing with the information stability requirement.

Lemma 4: Under the conditions of Theorem 3

i) If $T(\boldsymbol{Z}) < C(\boldsymbol{W})$, channel $\boldsymbol{W}$ strictly dominates source $\boldsymbol{Z}$.

ii) If channel $\boldsymbol{W}$ dominates source $\boldsymbol{Z}$, then $T(\boldsymbol{Z}) \leq C(\boldsymbol{W})$.

To conclude this section we point out that for the example in Section III the channel dominates the source, as it should from the converse part of Theorem 7. As we noted earlier the source $\boldsymbol{Z}$ and the channel with Bernoulli $(\frac{1}{2})$ input process $\boldsymbol{X}$, for each n, have their respective information spectra concentrated at $J(n)/n$. Hence for any n, δ, and c_n, at least one of the probabilities in Definition 10 is identically zero.

We now change the example in Section III slightly to illustrate a case of strict domination of a source by a channel whose capacity is strictly smaller than the minimum source-coding rate. The channel remains as before. But the source $\boldsymbol{Z}$ is defined slightly differently: The Bernoulli subsource now has probability $p < 1/2$. The minimum source coding rate of $\boldsymbol{Z}$ is $(2/3)h(p)$, whereas the channel capacity is $1/3$. Here $h(p)$ denotes the binary entropy function in bits. To see that the channel strictly dominates $\boldsymbol{Z}$, we will select the sequence

$$c_n = \frac{J(n)}{2n}[1 + h(p)].$$

From the law of large numbers it is easy to check that

$$\lim_{n\to\infty} P\left[\frac{1}{n} h_{Z^n}(z^n) \geq c_n\right] = 0.$$

On the other hand, if we let $\boldsymbol{X}$ be Bernoulli $(\frac{1}{2})$, we know that, with probability 1,

$$\frac{1}{n} i_{X^n W^n}(X^n; Y^n) = \frac{J(n)}{n}$$

which implies that as long as

$$\delta < \frac{1 - h(p)}{6}$$

$$\lim_{n\to\infty} P\left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n) \leq c_n + \delta\right] = 0.$$

In the special case of discrete memoryless channels and sources with a finite number of states, such as the example in Section III, the notions of strict domination and domination boil down to the asymptotic comparison of the deterministic sequences

$$\frac{1}{n}\sum_{i=1}^{n} H(Z_i) \tag{9}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} I(\overline{X}_i; \overline{Y}_i) = \frac{1}{n}\sum_{i=1}^{n} \max_{X_i} I(X_i; Y_i) \tag{10}$$

It can be checked that in that special case, strict domination is equivalent to the existence of $\delta > 0$ such that

$$\frac{1}{n}\sum_{i=1}^{n} I(\overline{X}_i; Y_i) - \sum_{i=1}^{n} H(Z_i) > \delta$$

for all sufficiently large n, whereas domination is equivalent to

$$\liminf_{n\to\infty} \sum_{i=1}^{n} I(\overline{X}_i; \overline{Y}_i) - \sum_{i=1}^{n} H(Z_i) \geq 0.$$

Note that in this special case of nonstationary discrete memoryless channels, the question of transmissibility depends only on the behavior of the deterministic sequences (9) and (10). Thus in the case, it is sensible to describe the source and the channel by those respective sequences, instead of the conventional scalar definitions of maximum source coding rate and capacity, which as we showed in Section III fail to predict whether transmissibility is possible. However, in the context of general (possibly information-unstable) sources and channels, we do not believe that the notions of domination can be substituted by simpler tests based on the comparison of a pair of deterministic scalar sequences characterizing the source and the channel, respectively.

V. Optimistic Capacity and the Separation Theorem

We saw in Theorem 6 that the direct part of the separation theorem holds in complete generality, namely, if the minimum achievable source coding rate is strictly less than the channel capacity, then the source is reliably transmissible through the channel. Furthermore, we saw in Section III that the converse part of the separation theorem fails to hold for some source–channel pairs. In this section, we characterize those channels (and sources) for which the converse separation theorem always holds.

In this respect, it is of interest to define for any channel the *source capacity* C_S, as the supremum of the minimum achievable source coding rates of the sources that can be reliably transmitted through the channel.[4] It follows from Theorem 4 that

$$C_S(\boldsymbol{W}) = \sup_{\boldsymbol{S} \text{ transmissible over } \boldsymbol{W}} \overline{\boldsymbol{H}}(\boldsymbol{S}). \tag{11}$$

The next result is an immediate consequence of the definitions and reveals the key role of the source capacity in checking the validity of the converse separation theorem.

Theorem 8: For any channel $\boldsymbol{W}$ the following are equivalent:

i) $C_S(\boldsymbol{W}) = C(\boldsymbol{W})$.

ii) For every source $\boldsymbol{Z}$ transmissible through $\boldsymbol{W}$, we have $T(\boldsymbol{Z}) \leq C(\boldsymbol{W})$.

[4] A similar quantity, C_E, where the supremum is taken over the subset of transmissible ergodic sources was mentioned in Section II.

Proof: First we show $C_S \geq C$ for every channel. According to Definition 6, for all $\gamma > 0$ there exists a sequence of (n, M, ϵ_n) codes such that $\epsilon_n \to 0$ and

$$C - \gamma < \frac{\log M}{n}.$$

Let X^n be the distribution that puts mass $2/M$ on each of the $M/2$ codewords with the lowest conditional probability of error. Clearly, all those codewords are different (for large enough n), for otherwise, ϵ_n does not vanish. The process $\boldsymbol{X}$ is reliably transmissible through the channel (with an *identity* encoder and the decoder of the corresponding channel code). Its minimum achievable source coding rate satisfies,

$$\overline{\boldsymbol{H}}(\boldsymbol{X}) = \limsup_{n\to\infty} \log \frac{M/2}{n} \geq C - \gamma$$

which enables us to conclude that $C_S \geq C$.

Finally, it follows from Theorem 4 and the definition of C_S that property ii) is equivalent to $C_S \leq C$. ■

We would like to give a characterization of those channels whose capacity is equal to the source–capacity, without recourse to source–channel transmission properties.[5] The main result of this section is that C_S is always equal to the so-called *optimistic* channel capacity, a concept which is closely connected with the conventional definition of channel capacity (Definition 6). The conventional definition [15] requires that good codes exist *for all sufficiently large blocklengths*; alternatively, we could require that good codes exist *for infinitely many blocklengths* (cf. Definition 6):

Definition 11: The *optimistic capacity* of channel $\boldsymbol{W}$, $\overline{C}(\boldsymbol{W})$, is the supremum of all the rates R for which there exist a sequence of (n, M, ϵ_n) codes such that

$$\frac{\log M}{n} > R$$

and

$$\liminf_{n\to\infty} \epsilon_n = 0.$$

Unlike C, $\overline{C}$ does not admit a simple expression such as that in Theorem 5 (cf. [14]). For further discussion on the optimistic definition versus the conventional one see [14], [15].

The next result is a consequence of our main result in Section IV (Theorem 7).

Theorem 9: For any channel $\boldsymbol{W}$, the source capacity is equal to the optimistic capacity $C_S(\boldsymbol{W}) = \overline{C}(\boldsymbol{W})$.

Proof: We first prove that $C_S \geq \overline{C}$. To prove this inequality, we have to show that for every $\delta > 0$ there exists a transmissible source $\tilde{\boldsymbol{X}}$ with $\overline{\boldsymbol{H}}(\tilde{\boldsymbol{X}}) \geq \overline{C} - \delta$.

By definition of optimistic capacity, for every $\delta > 0$ there exists a subsequence $K \subset N$ and a sequence of (n, M, ϵ_n) channel codes for $\boldsymbol{W}$ such that

$$\frac{\log M}{n} > \overline{C} - \frac{\delta}{4}$$

and

$$\lim_{n\to\infty,\, n\in K} \epsilon_n = 0.$$

We construct $\tilde{\boldsymbol{X}}$ as follows: for $n \in K$, $\tilde{X}^n$ is uniformly distributed over the codewords of the corresponding (n, M, ϵ_n) code, and for $n \notin K$, $\tilde{X}^n$ is deterministic. Clearly, $\tilde{\boldsymbol{X}}$ can be reliably transmitted over $\boldsymbol{W}$ and $\overline{H}(\tilde{\boldsymbol{X}}) \geq \overline{C} - \delta$. This shows that $C_S \geq \overline{C}$.

We proceed now to show that $C_S \leq \overline{C}$. Pick $\delta > 0$ and let $\boldsymbol{Z}$ be a transmissible source with $\overline{\boldsymbol{H}}(\boldsymbol{Z}) \geq C_S - (\delta/4)$. Such a source always exists by definition of C_S. Thus there exists a subsequence $K \subset N$ such that

$$\lim_{n\to\infty,\, n\in K} P\left(\frac{1}{n} h_{Z^n}(Z^n) \geq C_S - \frac{\delta}{2}\right) > 0$$

and since $\boldsymbol{Z}$ is transmissible, there exists a channel input process $\boldsymbol{X}$ satisfying

$$\lim_{n\to\infty,\, n\in K} P\left(\frac{1}{n} i_{X^n W^n}(X^n; Y^n) \leq C_S - \delta\right) = 0.$$

In particular, there exists a sequence $\tau_n \to 0$ such that

$$P\left(\frac{1}{n} i_{X^n W^n}(X^n; Y^n) \leq C_S - \delta\right) \leq \tau_n, \qquad \text{for all } n \in K.$$

We now apply Feinstein's Lemma to conclude that there exists a sequence of (n, M, ϵ_n) channel code, such that for $n \in K$

$$M = \exp\{n(C_S - 2\delta)\}$$

and

$$\epsilon_n \leq \tau_n + \exp(-n\delta).$$

This proves that $\overline{C} \geq C_S - 2\delta$. Since δ is arbitrary, $\overline{C} \geq C_S$, and the proof is complete. ■

Theorem 9 reveals an important operational characterization for the optimistic capacity which was unknown up to now. It implies that the classical statement of the separation theorem holds for a given channel (and every source) if and only if its optimistic capacity is equal to its conventional capacity. This property is indeed satisfied for most channels of interest. It is tantamount to requiring that no matter which subsequence of blocklengths we concentrate on, we cannot achieve a higher capacity. Another way to characterize this property in terms of the channel statistical description is given by Theorem 11 below.

Theorems 6 and 9 allow us to state the following generalized form of the joint source–channel coding theorem.

Theorem 10:

i) A source $\boldsymbol{Z}$ is reliably transmissible through channel $\boldsymbol{W}$ if $T(\boldsymbol{Z}) < C(\boldsymbol{W})$.

ii) A source $\boldsymbol{Z}$ is not reliably transmissible through $\boldsymbol{W}$ if $T(\boldsymbol{Z}) > \overline{C}(\boldsymbol{W})$.

Thus it is only in the case where the source-coding rate lies between the conventional and the optimistic channel capacities that recourse need be made to the conditions derived in Section IV. Note that it is not possible to strengthen the direct part by replacing C by $\overline{C}$. To check that, the reader may easily

[5] Note that if $C_S > C$, then there may exist a source that cannot be transmitted reliably through the channel but whose sup-entropy rate is less than or equal C_S.

construct a suitable source which is not reliably transmissible through the channel of Section III.

We now give an equivalent characterization of the condition $C = \overline{C}$ based on the information spectrum of the channel. The proof is relegated to the Appendix.

Theorem 11: For any channel $\boldsymbol{W}$, $\overline{C}(\boldsymbol{W}) = C(\boldsymbol{W})$ if an only if for all $\delta > 0$, and for all channel input processes $\boldsymbol{X}$ the following holds:

$$\liminf_{n\to\infty} P\left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n) \le C(\boldsymbol{W}) + \delta\right] > 0. \quad (12)$$

To conclude this section, we consider the dual problem; namely, under what condition on a given source can we guarantee that the converse to the separation theorem holds for any channel? Not surprisingly, the sought-after condition is that the optimistic and pessimistic (i.e., conventional) definitions of minimum achievable source coding rate coincide for that source.

Analogously to $C_S(\boldsymbol{W})$, define for any source $\boldsymbol{Z}$

$$T_C(\boldsymbol{Z}) = \inf_{\boldsymbol{V}} C(\boldsymbol{V})$$

where the infimum is over all channels over which $\boldsymbol{Z}$ can be reliably transmitted. Analogously to Theorem 8 we have,

Theorem 12: For any source $\boldsymbol{Z}$, the following are equivalent:

i) $T_C(\boldsymbol{Z}) = T(\boldsymbol{Z})$.

ii) For every channel $\boldsymbol{W}$ over which the source is transmissible, we have $T(\boldsymbol{Z}) \le C(\boldsymbol{W})$.

Proof: We first show that $T_C(\boldsymbol{Z}) \le T(\boldsymbol{Z})$. Select $\delta > 0$. There exists a sequence of (n, M, ϵ_n) source codes for $\boldsymbol{Z}$ such that

$$\frac{\log M}{n} \le T(\boldsymbol{Z}) + \delta$$

and

$$\lim_{n\to\infty} \epsilon_n = 0.$$

Let $\boldsymbol{W}$ be a channel with the same input and output alphabets with cardinality $\exp\{T(\boldsymbol{Z}) + \delta\}$. Let W^n be such that each of the elements of a subset of size M of A^n is mapped to itself with probability 1, and every other element is mapped to a common element. Clearly, $C(\boldsymbol{W}) \le T(\boldsymbol{Z}) + \delta$ and $\boldsymbol{Z}$ can be reliably transmitted over $\boldsymbol{W}$. Since δ is arbitrary, $T_C(\boldsymbol{Z}) \le T(\boldsymbol{Z})$.

Finally, it is obvious that property ii) is equivalent to $T_C(\boldsymbol{Z}) \ge T(\boldsymbol{Z})$. ■

Analogously to the phenomenon we saw for channels, $T_C(\boldsymbol{Z})$ is equal to $\underline{T}(\boldsymbol{Z})$, the optimistic minimum achievable source-coding rate defined as in Definition 4, substituting *for all sufficiently large* n by *for infinitely many* n. The proof of the following result can be found in the Appendix.

Theorem 13: For any source $\boldsymbol{Z}$; $T_C(\boldsymbol{Z}) = \underline{T}(\boldsymbol{Z})$.

As an example, if the source $\boldsymbol{Z}$ is stationary, then it satisfies the property that the optimistic and pessimistic definitions of the minimum source-coding rate coincide. To prove this, note that the minimum source-coding rate of any stationary source will be the essential infimum of the minimum source-coding rate over all the ergodic modes that comprise this source—this is a simple consequence of the ergodic decomposition of the limit of $\{(1/n)h_{Z^n}(Z^n)\}$. Hence for any real number α,

$$\lim_{n\to\infty} P\left[\frac{1}{n} h_{Z^n}(Z^n) \le \alpha\right]$$

exists. The equality of optimistic and pessimistic definitions of the minimum source-coding rates follows immediately and hence the separation theorem is true whenever the source is stationary regardless of the channel. This result cannot be shown from the classical results of [2], [4].

APPENDIX

Proof of Lemma 4

Using the fact that the channel is information stable and

$$\lim_{n\to\infty} C_n(W^n) = C(\boldsymbol{W})$$

we will show that $(1/n)i_{X^n W^n}(X^n; Y^n)$ converges in probability to $C(\boldsymbol{W})$ for some input process $\boldsymbol{X}$. Indeed, information stability of the channel (Definition 3) implies that there exists an input process $\boldsymbol{X}$ such that for any $\lambda > 0$ and $\tau > 0$, and for sufficiently large n

$$P\left[(1-\tau)C_n(W^n) \le \frac{1}{n} i_{X^n W^n}(X^n; Y^n) \le (1+\tau)C_n(W^n)\right] \ge 1 - \lambda. \quad (13)$$

We also know that for any $\theta > 0$ and for all sufficiently large n, $C(\boldsymbol{W}) - \theta \le C_n(W^n) \le C(\boldsymbol{W}) + \theta$. Using this and (13), and taking n suitably large we conclude that

$$P\left[(1-\tau)(C(\boldsymbol{W}) - \theta) \le \frac{1}{n} i_{X^n W^n}(X^n; Y^n) \le (1+\tau)(C(\boldsymbol{W}) + \theta)\right] \ge 1 - \lambda.$$

Since λ, τ, and θ are all arbitrary positive numbers, we have proved the convergence in probability of $(1/n)i_{X^n W^n}(X^n; Y^n)$ to $C(\boldsymbol{W})$.

From [14], any pair $(\boldsymbol{X}, \boldsymbol{Y})$ satisfies

$$\underline{\boldsymbol{I}}(\boldsymbol{X}; \boldsymbol{Y}) \le \liminf_{n\to\infty} \frac{1}{n} I(X^n; Y^n). \quad (14)$$

This property can be applied for a particular subsequence of N too; the inf-information rate defined over any subsequence is greater than or equal to the inf-information rate defined over the whole sequence.

i) Let $\delta = C - T$, and $c_n = C - 2\delta$ for all n. We see that the definition of strict domination (Definition 9) is satisfied with this choice, using Theorem 4.

ii) Fix $\gamma > 0$. In the definition of domination (Definition 10), choose $c_n = \overline{\boldsymbol{H}}(\boldsymbol{Z}) - \gamma$, for all n. It follows by definition of $\overline{\boldsymbol{H}}(\boldsymbol{Z})$ that there exists a subsequence $K \subset N$ such that

$$\lim_{n\to\infty,\, n\in K} P\left[\frac{1}{n} h_{Z^n}(Z^n) \ge c_n\right] > 0.$$

Hence, for any $\delta > 0$ there exists a channel input process $\hat{\boldsymbol{X}}$ such that

$$\lim_{n\to\infty,\, n\in K} P\left[\frac{1}{n} i_{\hat{X}^n W^n}(\hat{X}^n;\, \hat{Y}^n) \le c_n - \delta\right] = 0.$$

Hence, the inf-information rate of $(\hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}})$, defined over the subsequence K is greater than or equal to $\overline{\boldsymbol{H}}(\boldsymbol{Z}) - \gamma$. We also have, using

$$\lim_{n\to\infty} C_n = C$$

that

$$\limsup_{n\to\infty} \frac{1}{n} I(\hat{X}^n;\, \hat{Y}^n) \le C.$$

Applying (14) over subsequence K, we see that inf-information rate defined over K is less than or equal to C, and hence $C \ge \overline{\boldsymbol{H}}(\boldsymbol{Z}) - \gamma$, where the choice is γ is arbitrary. This concludes the proof. ■

Proof of Theorem 11

Observe from Theorem 5 that if a channel has capacity C, then for any $\delta > 0$ there exists a channel input process $\boldsymbol{X}$ such that

$$\lim_{n\to\infty} P\left[\frac{1}{n} i_{X^n W^n}(X^n;\, Y^n) \le C(\boldsymbol{W}) - \delta\right] = 0. \tag{15}$$

We will use a similar characterization of $\overline{C}$ (replacing the limit in (15) by liminf).

Lemma 5: For any $\delta > 0$, there exists a channel input process $\boldsymbol{X}$ such that

$$\liminf_{n\to\infty} P\left[\frac{1}{n} i_{X^n W^n}(X^n;\, Y^n) \le \overline{C}(\boldsymbol{W}) - \delta\right] = 0. \tag{16}$$

Proof: Construct the process $\tilde{\boldsymbol{X}}$ which is transmissible over $\boldsymbol{W}$ exactly as in the proof of Theorem 9. This process satisfies (cf. proof of Theorem 8),

$$\lim_{n\to\infty,\, n\in K} P\left(\frac{1}{n} h_{\tilde{X}^n}(\tilde{X}^n) \ge \overline{C} - \frac{\delta}{2}\right) > 0. \tag{17}$$

By Theorem 7, $\boldsymbol{W}$ dominates $\tilde{\boldsymbol{X}}$ and hence from Definition 10, there exists an input process $\boldsymbol{X}$ such that

$$\lim_{n\to\infty,\, n\in K} P\left(\frac{1}{n} i_{X^n W^n}(X^n;\, Y^n) \le \overline{C} - \delta\right) = 0. \quad ■$$

We now proceed with the proof of Theorem 11. Observe that by definition, $\overline{C} \ge C$. Hence to show that condition (12) implies $\overline{C} = C$, it is enough to show that it implies $\overline{C} \le C$. By way of contradiction, assume that (12) is satisfied for every $\delta > 0$ and every $\boldsymbol{X}$, yet $\overline{C} > C + \gamma$ for some $\gamma > 0$. Now from Lemma 5, we know that there exists an input process $\tilde{\boldsymbol{X}}$ such that

$$\liminf_{n\to\infty} P\left(\frac{1}{n} i_{\tilde{X}^n W^n}(\tilde{X}^n;\, \tilde{Y}^n) \le \overline{C} - \frac{\gamma}{2}\right) = 0. \tag{18}$$

Since by assumption $\overline{C} \ge C + \gamma$, (18) implies that

$$\liminf_{n\to\infty} P\left(\frac{1}{n} i_{\tilde{X}^n W^n}(\tilde{X}^n;\, \tilde{Y}^n) \le C + \frac{\gamma}{2}\right) = 0$$

which contradicts our assumption that for all $\delta \ge 0$ and $\boldsymbol{X}$ (12) is satisfied. This establishes the if part.

We proceed now to prove the other direction, again by a contradiction argument. Thus assume that

$$\overline{C} = C \tag{19}$$

yet there exist $\delta > 0$, an input process $\boldsymbol{X}$, a subsequence $K \subset N$ and a sequence $\tau_n \to 0$ such that

$$P\left(\frac{1}{n} i_{X^n W^n}(X^n;\, Y^n) < C + \delta\right) \le \tau_n, \qquad \text{for all } n \in K.$$

Using Feinstein's Lemma we conclude that there exists a sequence of (n, M, ϵ_n) codes for the channel $\boldsymbol{W}$ satisfying, on the subsequence K

$$M = \exp\left\{n\left(C + \frac{\delta}{2}\right)\right\}$$

and

$$\epsilon_n \le \tau_n + \exp\left(-n\frac{\delta}{2}\right)$$

implying that $\overline{C} \ge C + (\delta/2)$, in contradiction with (19). ■

Proof of Theorem 13

We start by proving that

$$T_C(\boldsymbol{Z}) \le \underline{T}(\boldsymbol{Z}). \tag{20}$$

To prove this inequality we have to show that for every $\delta > 0$ there exists a channel $\boldsymbol{W}$ over which $\boldsymbol{Z}$ is transmissible, with $C(\boldsymbol{W}) \le \underline{T}(\boldsymbol{Z}) + \delta$. By definition of $\underline{T}(\boldsymbol{Z})$, there exist a subsequence $K \subset N$ and a sequence of (n, M, ϵ_n) source codes for $\boldsymbol{Z}$ such that

$$\frac{\log M}{n} \le \underline{T}(\boldsymbol{Z}) + \frac{\delta}{4}$$

and

$$\lim_{n\to\infty,\, n\in K} \epsilon_n = 0.$$

Let G_n be any subset of F^n (where F is the alphabet of $\boldsymbol{Z}$) which includes the (n, M, ϵ) source codebook, and has cardinality

$$|G_n| = \exp\{n[\underline{T}(\boldsymbol{Z}) + \delta]\}.$$

Let $\boldsymbol{W}$ be a channel with input and output alphabets both equal to F satisfying

$$W^n(\tilde{z}^n|z^n) = \begin{cases} 1, & \text{for } z^n \in G_n, \quad \tilde{z}^n = z^n \\ \frac{1}{|F|^n}, & \text{for } z^n \notin G_n \end{cases} \qquad \text{for } n \in K$$

and

$$W^n = \text{identity channel on } F^n \text{ for } n \notin K.$$

Clearly $\boldsymbol{Z}$ can be reliably transmitted over $\boldsymbol{W}$, and $C(\boldsymbol{W}) = \underline{T}(\boldsymbol{Z}) + \delta$. Since δ is arbitrary, we have proved (20).

We proceed to prove that $T_C(\boldsymbol{Z}) \geq \underline{T}(\boldsymbol{Z})$. We have to show that the capacity of every channel over which $\boldsymbol{Z}$ is transmissible, is at least $\underline{T}(\boldsymbol{Z})$. By definition of $\underline{T}(\boldsymbol{Z})$ and the source coding results of [13], for every $\delta > 0$ we have

$$\liminf_{n\to\infty} P\left(\frac{1}{n} h_{Z^n}(Z^n) \geq \underline{T}(Z) - \frac{\delta}{4}\right) > 0. \tag{21}$$

Now let $\boldsymbol{W}$ be a channel over which $\boldsymbol{Z}$ is transmissible. Since $\boldsymbol{W}$ dominates $\boldsymbol{Z}$, (21) implies that there exists an input process $\boldsymbol{X}$ such that

$$\lim_{n\to\infty} P\left(\frac{1}{n} i_{X^nW^n}(X^n; Y^n) \leq \underline{T}(Z) - \frac{\delta}{2}\right) = 0. \tag{22}$$

Using again Feinstein's Lemma as in the proofs of Theorem 9 and 11, we conclude that $\underline{T}(\boldsymbol{Z}) - \delta$ is an achievable rate for the channel $\boldsymbol{W}$, whenever $\boldsymbol{Z}$ is transmissible over $\boldsymbol{W}$. Since δ is arbitrary, the proof is complete. ■

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, July 1948.

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

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

[4] G. D. Hu, "On Shannon theorem and its converse for sequence of communication schemes in the case of abstract random variables," in *Trans. 3rd Prague Conf. on Information Theory, Statistical Decision Functions, Random Processes* (Czechoslovak Academy of Sciences, Prague, 1964), pp. 285–333.

[5] S. Vembu, "Information theory without ergodicity assumptions: Some new results," Ph.D. dissertation, Princeton University, Princeton, NJ, June 1994.

[6] A. N. Kolmogorov, "On the Shannon theory of information in the case of continuous signals," *IEEE Trans. Inform. Theory*, vol. IT-2, pp. 102–108, Sept. 1956.

[7] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *IRE Nat. Conv. Rec.*, pt. 4, pp. 142–163, Mar. 1959.

[8] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrary correlated sources," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 6, pp. 648–657, Nov. 1980.

[9] R. M. Gray and L. D. Davisson, Eds., *Ergodic and Information Theory* (Benchmark Papers in Electrical Engineering and Computer Science), Dowden, Hutchinson and Ross, Inc. 1977.

[10] I. Csiszár, "Information theory and ergodic theory," *Probl. Contr. Inform. Theory*, vol. 16, no. 1, pp. 3–27, 1978.

[11] J. C. Kieffer, "Block coding for weakly continuous channels," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 721–727, 1981.

[12] K. Winkelbauer, "On the asymptotic rate of non-ergodic information sources," *Kybernetika*, vol. 6, pp. 127–148, 1970.

[13] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[14] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.

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

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[17] R. B. Ash, *Information Theory.* New York: Interscience, 1965; Dover, 1990.