Variable-Rate Channel Capacity

Sergio Verdú, *Fellow, IEEE*, and Shlomo Shamai (Shitz), *Fellow, IEEE*

***Abstract*—This paper introduces the notions of variable-to-fixed and fixed-to-variable channel capacity, without feedback. For channels that satisfy the strong converse, these notions coincide with the conventional Shannon capacity. For channels that do not behave ergodically, the conventional fixed-rate Shannon capacity only depends on least-favorable channel conditions, while the variable-rate capacity notions are able to capture the whole range of channel states and their likelihood, even in the absence of any side information about channel state at the transmitter. Particular emphasis is placed on memoryless channels that are governed by finitely valued states. We show that (single-user) variable-to-fixed channel capacity is intimately connected to the capacity region of broadcast channels with degraded message sets, and we give an expression for the fixed-to-variable capacity.**

***Index Terms*—Bayesian modeling, broadcast channels with degraded message sets, channel capacity, fixed-to-variable coding, fountain codes, nonergodic channels, Shannon theory, state-dependent channels, variable-to-fixed coding.**

I. Introduction

THE notion of variable-rate channel coding where the encoder adjusts its rate (and/or other resources such as power and bandwidth) to the actual channel condition (or *state*) is well established in both the research literature and current technology. When the channel state is known at the transmitter, the average of the capacities achievable for the individual channel states is an important fundamental limit. Furthermore, the benefits of variable-length channel coding are well known when feedback is available (e.g., [4], [28], [32], and [22]).

In this paper, we study a different setup: neither feedback nor side information about the channel conditions (other than its statistical description) is available at the encoder. In such a scenario, we develop the notion of "variable-rate" not from the adaptation of the encoding strategy to channel conditions, which is no longer possible, but from the adaptation of the decoder. For channels that behave ergodically (nothing can be learned beyond the already available statistical description), the new notions coincide with the conventional Shannon capacity.

It is interesting to contrast the channel coding setup with lossless data compression where the almost-lossless fixed-rate and the lossless variable-rate approach are well-known. For stationary ergodic sources, the Shannon–MacMillan theorem shows that the entropy rate is the minimum achievable almost-lossless fixed rate. For stationary (not necessarily ergodic) sources, the minimum average rate achievable with lossless fixed-to-variable (or variable-to-fixed) codes is also the entropy rate. For general sources, the almost-lossless minimum achievable fixed rate is the sup-entropy rate[1] while the lossless minimum average rate is $\limsup_{n\to\infty} \frac{1}{n} H(X^n)$ [13, Theorems 1 and 2]. For example, if with probability π_i the source is equal to a stationary ergodic source whose entropy rate is H_i, then the almost-lossless minimum fixed-rate is $\max_i H_i$, while the lossless minimum average rate is $\sum_i \pi_i H_i$. In such a nonergodic case, allowing variable rate has the obvious advantage of letting the code adapt to the actual ergodic mode in effect. Extending this parallel to channels is fairly straightforward if the channel state is known at the encoder; otherwise, it is not at all obvious.

To motivate the new notions of variable-rate channel capacity, consider a binary-input binary-output channel that:

- with probability $1-q$ reproduces the input sequence error-free;
- with probability q introduces errors independently with probability $0.11 = h^{-1}(\frac{1}{2})$, where $h(x)$ is the binary entropy function in bits.[2]

If the encoder knew which of the two states is in effect (through, for example, a feedback link from the decoder, which can learn the channel state with negligible overhead asymptotically), it could adapt its rate and transmission strategy (uncoded for the error-free channel, and rate-$\frac{1}{2}$ code otherwise) resulting in an average rate of reliable information transmission of $1-\frac{q}{2}$. But if the encoder does not know the channel state, no code with rate greater than $\frac{1}{2}$ can lead to vanishing block error probability, and the Shannon capacity is equal to $\frac{1}{2}$, as if the channel were in the error-free state with zero probability. A finer measure of the channel transmission capability is the ϵ-capacity (e.g., [8] and [13]) which gives the maximum rate compatible with block error probability not larger than ϵ. In the current example

$$C_\epsilon = \begin{cases} 1 & q \le \epsilon < 1 \\ \frac{1}{2} & 0 < \epsilon < q \end{cases}. \tag{1}$$

Although traditionally the terms outage capacity and ϵ-capacity are used interchangeably, a different notion of outage capacity is introduced in [11]: the outage-ϵ capacity is the maximum value of $(1-\epsilon)R$ for which there exist codes with rate R, vanishing undetected error probability, and detected error probability not larger than ϵ. It easy to see that in the foregoing example, the

Manuscript received September 01, 2009; revised February 21, 2010. Current version published May 19, 2010. This work was supported in part by the National Science Foundation under Grants CCF-0635154 and CCF-0728445, in part by the US-Israel Binational Science Foundation, and in part by a Cisco Collaborative Research Initiative Grant.

S. Verdú is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: verdu@princeton.edu).

S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: sshlomo@ee.technion.ac.il).

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

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2010.2046220

[1]The fundamental limit is potentially higher in an arbitrarily varying setup where reliability is required in the worst case among a collection of source distributions. For the memoryless arbitrarily varying source the minimum achievable rate is the entropy of the least-favorable mixture of sources [30].

[2]$h(x) = -x\log_2 x - (1-x)\log_2(1-x)$, for $0 < x < 1$.

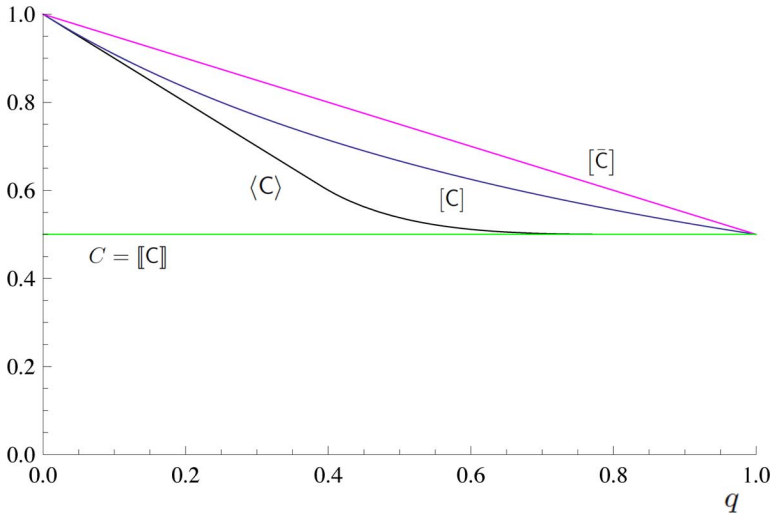

Fig. 1. Comparison of capacities for a binary symmetric channel whose crossover probability is 0.11 with probability q and 0 with probability $1-q$.

outage-ϵ capacity is $(1-\epsilon)/2$ if $0 < \epsilon < q$, and the outage-q capacity is $1-q$.

The result in (1) seems to indicate that there are only two reasonable coding strategies: send the information uncoded if block error probability $\epsilon = q$ is acceptable; otherwise, code at rate $\frac{1}{2}$. But this is only within the paradigm of the following.

- **Fixed-to-fixed coding:** the number of transmitted information bits and the number of observed channel symbols (blocklength) are prespecified. This is the conventional setup.

There are, in fact, other possible channel coding strategies, whose fundamental limits are the object of this paper; the first "variable/fixed" refers to the length of the reliably decoded message, while the second "fixed/variable" refers to the length of the observed channel outputs.

- **Variable-to-fixed coding:** the number of observed channel symbols (blocklength) is prespecified, but the number of reliably recovered information bits depends on channel conditions.
- **Fixed-to-variable coding:** the number of transmitted information bits is prespecified but the number of channel observations required to recover them depends on channel conditions. This is the setup of rateless fountain codes [16], [29].
- **Variable-blocklength encoding:** the number of transmitted information bits is prespecified while the blocklength is allowed to the depend on those information bits, but not on channel conditions (unknown at the encoder). The decoder is assumed to able to obtain the blocklength chosen by the encoder noiselessly.

The organization of the remainder of this paper is the following. Section II gives the definitions of:

- C: conventional Shannon channel capacity;
- $\langle \mathsf{C} \rangle$: variable-to-fixed channel capacity;
- $[\![\mathsf{C}]\!]$: fixed-to-variable channel capacity;
- $[\bar{\mathsf{C}}]$: upper fixed-to-variable channel capacity;
- $\llbracket \mathsf{C} \rrbracket$: variable-blocklength channel capacity.

The difference between $[\![\mathsf{C}]\!]$ and $[\bar{\mathsf{C}}]$ resides in that the average is with respect to the variable number of observations for $[\![\mathsf{C}]\!]$, and with respect to the ensuing rate for $[\bar{\mathsf{C}}]$. General relations between C, C_ϵ, $\langle \mathsf{C} \rangle$, $[\![\mathsf{C}]\!]$ and $[\bar{\mathsf{C}}]$ are shown in Sections III and IV. For state-dependent channels, the variable-to-fixed channel capacity $\langle \mathsf{C} \rangle$ is shown to be intimately related to the capacity region of broadcast channels with degraded message sets; and the upper fixed-to-variable capacity can be obtained as the maximal average mutual information, which often coincides with the average capacity obtained when the encoder knows the channel state. In Section V, we show that $C = \llbracket \mathsf{C} \rrbracket$, for any nonanticipatory channel and, therefore, allowing the blocklength to depend on the message is not advantageous. Section VI applies the concepts introduced in the paper to a number of illustrative examples, one of which is the example above whose various capacities are shown in Fig. 1.

II. Definitions

Denote the input and output alphabets of the channel by $\mathcal{X}$ and $\mathcal{Y}$, respectively. The channel is a sequence of conditional probabilities

$$\{P_{Y^n|X^n} : \mathcal{X}^n \mapsto \mathcal{Y}^n\}_{n=1}^{\infty}. \tag{2}$$

A. C: Fixed-Blocklength Channel Capacity

Definition 1: The conventional Shannon capacity, C, (in bits/channel use) is the largest $R \geq 0$ for which there exists a sequence of encoders/decoders:[3]

$$f^{k_n \to n} : \{0,1\}^{k_n} \mapsto \mathcal{X}^n \tag{3}$$

$$g^{n \to k_n} : \mathcal{Y}^n \mapsto \{0,1\}^{k_n} \tag{4}$$

whose (average) block error probability and rate satisfy

$$\lim_{n\to\infty} \epsilon_n = 0 \tag{5}$$

$$\liminf_{n\to\infty} \frac{k_n}{n} = R \tag{6}$$

[3]It will become evident that the punctilious notation for various forms of encoders/decoders is called for despite its cumbersomeness.

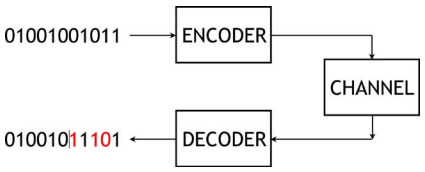

Fig. 2. Example where $m_n = n = 11$ and $L_{11} = 6$.

respectively. If (5) is replaced by

$$\limsup_{n\to\infty} \epsilon_n = \epsilon \tag{7}$$

then, we obtain the notion of ϵ-capacity, C_ϵ.

B. $\langle \mathsf{C} \rangle$: Variable-to-Fixed Channel Capacity

As a shorthand, we denote

$$m_n = \lceil n \log_2 |\mathcal{X}| \rceil. \tag{8}$$

Definition 2: For an encoder-decoder pair

$$\mathsf{f}^{:n} : \{0,1\}^{m_n} \mapsto \mathcal{X}^n \tag{9}$$

$$\mathsf{g}^{n:} : \mathcal{Y}^n \mapsto \{0,1\}^{m_n} \tag{10}$$

with message $(B_1, \ldots, B_{m_n}) \in \{0,1\}^{m_n}$, let Y^n denote the channel response to the input

$$X^n = \mathsf{f}^{:n}(B_1, \ldots, B_{m_n}). \tag{11}$$

The number of (consecutively) recovered bits L_n is defined as the largest integer k such that

$$(B_1, \ldots, B_k) = \mathsf{g}^{n:k}(Y^n) \tag{12}$$

where $\mathsf{g}^{n:k}(y^n)$ denotes bits $(1, \ldots, k)$ in $\mathsf{g}^{n:}(y^n)$. Note that L_n is a random variable that depends on encoder, decoder, message and channel realization. An illustration of L_n is provided in Fig. 2.

Remark 1: When the input alphabet has infinite cardinality, m_n can be substituted by ∞ in Definition 2. However, ordinarily, there is an upper bound $\bar{C}$ to the maximum rate that can be achieved in the most optimistic case. Then, we can safely take $m_n = n\bar{C}$.

Definition 3: R (bits/channel use) is a *variable-to-fixed achievable rate* if there exists a sequence of codes $(\mathsf{f}^{:n}, \mathsf{g}^{n:})$ whose expected number of recovered bits satisfies

$$R = \liminf_{n\to\infty} \frac{\mathbb{E}[L_n]}{n} \tag{13}$$

where the expectation is over the channel random transformation and fair bits $(B_1, \ldots, B_{m_n})$. The maximum variable-to-fixed achievable rate is the variable-to-fixed capacity $\langle \mathsf{C} \rangle$.

Remark 2: Note that the notion of block error probability does not play any role in the definition of $\langle \mathsf{C} \rangle$. In the definition of the conventional fixed-to-fixed capacity C, one erroneous bit is as bad as many erroneous bits. In the definition of variable-to-fixed capacity, reliability is gauged by the average length of the initial run of correctly decoded bits. The idea is that we can protect the initial bits in the message with the most redundancy so that they can be correctly decoded regardless of the channel state. As we show below, if the channel has only one state (more precisely, if it satisfies the strong converse) then, there is nothing to be gained by unequal error protection and $\langle \mathsf{C} \rangle = C$.

Remark 3: Error-correcting code technology is easy to adapt so that the decoder can get a fairly reliable estimate of the number of consecutively correctly decoded bits L_n, based on the observation of Y^n. Alternatively, once the decoder learns the state, it can compute an estimate of L_n (or more precisely of L_n/n).

C. $[\mathsf{C}]$: Fixed-to-Variable Channel Capacity

As a shorthand, a sequence of decoder mappings

$$\{\mathsf{g}^{n:k} : \mathcal{Y}^n \mapsto \{0,1\}^k\}_{n=1}^{\infty} \tag{14}$$

is denoted by $\mathsf{g}^{:k}$. Moreover, a rateless encoder is a mapping

$$\mathsf{f}^{k:} : \{0,1\}^k \mapsto \mathcal{X}^\infty. \tag{15}$$

Definition 4: For a $(\mathsf{f}^{k:}, \mathsf{g}^{:k})$ pair, define the number N_k of channel symbols required to recover the k transmitted bits as the smallest integer n such that

$$B^k = (B_1, \ldots, B_k) = \mathsf{g}^{n:k}(Y^n) \tag{16}$$

where Y^n are symbols $(1, \ldots n)$ in the response of the channel to the semi-infinite input sequence $\mathsf{f}^{k:}(B_1, \ldots B_k)$. Note that N_k is a function of the encoder, decoder, message and channel noise.

Definition 5: R (bits/channel use) is a *fixed-to-variable achievable rate* if there is a sequence of pairs $(\mathsf{f}^{k:}, \mathsf{g}^{:k})$ such that

$$R = \liminf_{k\to\infty} \frac{k}{\mathbb{E}[N_k]} \tag{17}$$

$$= \left(\limsup_{k\to\infty} \frac{\mathbb{E}[N_k]}{k} \right)^{-1} \tag{18}$$

where the expectation is over the channel random transformation and fair bits $(B_1, \ldots, B_k)$. The maximum fixed-to-variable achievable rate is the fixed-to-variable capacity $[\mathsf{C}]$.

Remark 4: Definition 5 gives the fundamental limit of rateless codes [16], [29], where the reliably decoded message has a fixed number of bits and the number of channel uses is channel dependent. Note that the definition of fountain capacity given in [24] refers to a different setup where the decoder only observes the output of the channel at times that are unknown to the encoder. The application of rateless codes to arbitrarily-varying channels (where no distribution is placed on the sate of states) is explored in [9].

Remark 5: Variable-to-fixed channel coding is dual to fountain rateless coding: the blocklength (number of channel uses) is fixed, while the number of successfully decoded bits is channel dependent. Note that in contrast to fixed-to-variable coding, in variable-to-fixed coding, the transmitter does not actually know

in general which fraction of the transmitted information is delivered reliably.

Remark 6: Similarly to Remark 3, the decoder can gain a very reliable estimate of N_k at the expense of minimal loss in efficiency (if k is sufficiently high) through, for example, hashing.

D. $[\![\bar{\mathsf{C}}]\!]$: Upper Fixed-to-Variable Channel Capacity

Definition 6: R (bits/channel use) is an *upper fixed-to-variable achievable rate* if there is a sequence of pairs $(\mathsf{f}^{k:}, \mathsf{g}^{:k})$ such that

$$R = \liminf_{k\to\infty} \mathbb{E}\left[\frac{k}{N_k}\right] \tag{19}$$

where the expectation is over the channel random transformation and fair bits $(B_1, \ldots, B_k)$. The maximum upper fixed-to-variable achievable rate is the upper fixed-to-variable capacity $[\![\bar{\mathsf{C}}]\!]$.

Remark 7: According to Jensen's inequality

$$[\![\mathsf{C}]\!] \le [\![\bar{\mathsf{C}}]\!]. \tag{20}$$

Definition 6 is useful for example, when a population of receivers retrieves the same message through statistically different channels. Definition 5 is geared to gauging the expected delay to reliably recover the information, and to the common setting where consecutive equal-size blocks of information are transmitted.[4]

E. $[\![\mathsf{C}]\!]$: Variable-Blocklength Channel Capacity

The set of all nonempty strings drawn from $\mathcal{X}$ is denoted by $\mathcal{X}^+$. The length of $\mathsf{s} \in \mathcal{X}^+$ is denoted by $\ell(\mathsf{s})$.

Definition 7: The variable-blocklength channel capacity, $[\![\mathsf{C}]\!]$, (cf. [8, Problem 2.1.25]) is the largest $R \ge 0$ such that there exists a sequence of encoders/decoders

$$f^{k\to} : \{0,1\}^k \mapsto \mathcal{X}^+ \tag{21}$$

$$g^{\to k} : \mathcal{Y}^+ \mapsto \{0,1\}^k \tag{22}$$

with vanishing error probability and

$$\liminf_{k\to\infty} \frac{k}{\mathbb{E}[\ell(f^{k\to}(B_1, \ldots, B_k))]} = R \tag{23}$$

where the expectation is over independent equiprobable bits $B_1, \ldots, B_k$.

III. Variable-to-Fixed Capacity

A. Relationship to Fixed-to-Fixed Capacity

In this subsection we investigate the relationship between variable-to-fixed capacity and conventional (fixed-rate) Shannon capacity and ϵ-capacity. These notions turn out to be identical for channels that behave ergodically.

Theorem 1: For any channel $\{P_{Y^n|X^n}\}_1^\infty$, the Shannon capacity C and the variable-to-fixed capacity $\langle\mathsf{C}\rangle$, satisfy

$$C \le \langle\mathsf{C}\rangle. \tag{24}$$

[4]See [19] for a similar receiver-centric notion in a broadcast setting.

Proof: To show that the Shannon capacity, C, is an achievable variable-to-fixed achievable rate, take a sequence of (conventional) k-to-n codes $(f^{k\to n}, g^{n\to k})$ $(k \le m_n)$ with block error probability ϵ_n such that

$$\liminf_{n\to\infty} \frac{k}{n} = C \tag{25}$$

$$\lim_{n\to\infty} \epsilon_n = 0. \tag{26}$$

Let

$$\mathsf{f}^{:n}(b_1, \ldots b_{m_n}) = f^{k\to n}(b_1, \ldots b_k) \tag{27}$$

$$\mathsf{g}^{n:}(y^n) = (g^{n\to k}(y^n), 0, \ldots 0) \tag{28}$$

where the number of 0s is equal to $m_n - k$. The expected length of the number of consecutive correctly decoded bits can be lower bounded by $k(1-\epsilon_n)$ (if the conventional code recovers the message, the first k bits in $\mathsf{g}^{n:}(y^n)$ are correctly decoded). Therefore

$$\liminf_{n\to\infty} \frac{\mathbb{E}[L_n]}{n} \ge \liminf_{n\to\infty} \frac{k(1-\epsilon_n)}{n} \tag{29}$$

$$= C \tag{30}$$

where (30) follows from (25)–(26). ∎

Theorem 2: Suppose that the channel input alphabet is finite. Then

$$\langle\mathsf{C}\rangle \le \lim_{\epsilon\uparrow 1} C_\epsilon \tag{31}$$

where C_ϵ is the ϵ-capacity.

Proof: We will argue by contradiction. Let us assume that (31) is false and, therefore

$$\lim_{\epsilon\uparrow 1} C_\epsilon < C_2 < \langle\mathsf{C}\rangle. \tag{32}$$

where

$$C_2 = \frac{1}{2}\langle\mathsf{C}\rangle + \frac{1}{2}\lim_{\epsilon\uparrow 1} C_\epsilon. \tag{33}$$

By definition of $\langle\mathsf{C}\rangle$, since $\frac{L_n}{n} \le \lceil \log_2 |\mathcal{X}| \rceil$, there must exist $\epsilon_0 < 1$, n_0 and a sequence of variable-to-fixed encoders/decoders $(\mathsf{f}^{:n}, \mathsf{g}^{n:})$ such that

$$\mathbb{P}\left[\frac{L_n}{n} \le C_2\right] \le \epsilon_0 \tag{34}$$

for $n \ge n_0$. We now construct conventional k_n-to-n codes with

$$k_n = \lfloor C_2 n \rfloor \tag{35}$$

$$f^{k_n\to n}(b_1, \ldots b_{k_n}) = \mathsf{f}^{:n}(b_1, \ldots b_{k_n}, d_{k_n+1}, \ldots d_{m_n}) \tag{36}$$

$$g^{n\to k_n}(y^n) = \mathsf{g}^{n:k_n}(y^n) \tag{37}$$

where d_j, $j = k_n + 1, \ldots, m_n$ are deterministic symbols (known to the decoder) chosen to maximize the conditional probability of $L_n \ge k_n$ given those symbols for the particular variable-to-fixed code that has been chosen. The asymptotic rate of this sequence of codes is equal to C_2. The conventional k_n-to-n code makes an error if and only if $L_n < k_n$, but the probability of that event is upper bounded by ϵ_0 according to

(34) (where the probability is taken with respect to a larger and less favorable set of messages). Thus

$$C_{\epsilon_0} \geq C_2 > \lim_{\epsilon \uparrow 1} C_\epsilon \tag{38}$$

where we assumed the strict inequality in (32). But (38) is impossible since C_ϵ is monotone nondecreasing. ■

Remark 8: Theorems 1 and 2 imply that for channels that satisfy the strong converse, i.e.,

$$C = \lim_{\epsilon \uparrow 1} C_\epsilon \tag{39}$$

the conventional capacity is equal to the variable-to-fixed capacity. More generally, suppose that conditioned on a random parameter Δ known to both receiver and transmitter, the channel satisfies the strong converse

$$C(\Delta) = \lim_{\epsilon \uparrow 1} C_\epsilon(\Delta). \tag{40}$$

Then, the variable-to-fixed capacity is equal to the average capacity $\mathbb{E}[C(\Delta)]$ when Δ is known at both encoder and decoder.

A simple example for which Theorems 1 and 2 are not tight, but for which it is nevertheless straightforward to compute the variable-to-fixed channel capacity is the case when the channel is useless with some probability (Example 1 in Section VI).

B. State-Dependent Channels

In this section we consider channels that can be expressed as

$$P_{Y^n \mid X^n}(b^n \mid a^n) = \sum_{\ell=1}^{K} \pi_\ell \, P_{Y^n \mid X^n, S}(b^n \mid a^n, \ell) \tag{41}$$

for all n. These channels are also called "averaged" [14] and "mixed" [12]. The important subclass of state-dependent discrete memoryless channels, considered in more detail in Section IV.C, is such that the alphabets are finite and

$$P_{Y^n \mid X^n, S}(b^n \mid a^n, \ell) = \prod_{i=1}^{n} W_\ell(b_i \mid a_i) \tag{42}$$

where $\sum_{b \in \mathcal{Y}} W_\ell(b \mid a) = 1$ for all $a \in \mathcal{X}$, and all $\ell = 1, \ldots, K$.

A related, but different, class of channels are the "composite" channels considered in [10], [11] where the receiver observes S in addition to Y^n. Most of our results in this section do not require (42) but they do require that $P_{Y^n \mid X^n, S}$ behaves ergodically under each of the states. Under that condition we have the simple upper bound:

Theorem 3: Suppose that for each state i in (41), the channel $\{P_{Y^n \mid X^n, S=i}\}$ satisfies the strong converse and has capacity C_i. Then

$$\langle \mathsf{C} \rangle \leq \sum_{i=1}^{K} \pi_i C_i. \tag{43}$$

Proof: The upper bound in (43) would be tight were the encoder to know the state before the start of transmission (recall Remark 8). Indeed, we can view the channel as having an initial condition S with distribution $\pi_1, \ldots, \pi_K$; If $S = i$ is revealed to both encoder and decoder, then for any $\delta > 0$, the number of bits that can be reliably transmitted with blocklength n is upper bounded by $nC_i + n\delta$. Averaging over S we get the desired result. ■

The variable-rate capacity of state-dependent channels (41) is intimately connected with the capacity region of a certain broadcast channel. Cover [6], in his seminal paper on broadcast channels, suggested to deal with "compound channels with a prior distribution" [i.e., (41)] by maximizing the average rate achievable for a broadcast channel, where each state is associated with one of the users. This notion has been pursued for various channels such as fading [25], MIMO [26], multiple-access [18], [27] and binary-symmetric channels [11]. The broadcast approach has also been used in lossy source-channel coding in combination with successive refinement [21], [33]. When the state is known at the receiver, the broadcast approach has been shown to provide an achievable *expected rate* [11] in a setup where the error probability averaged with respect to the state is forced to vanish.

In our context, of particular interest are the special case of *broadcast channels with degraded message sets*. Consider the following generalization of the 2-user definition in [15].[5]

Definition 8: K-user broadcast channel with degraded message sets. Consider an ordered collection of K single-user channels with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$

$$\left(\{P_{Y^n \mid X^n, S=i}\}_{n=1}^{\infty}, i = 1, \ldots, K\right)$$

A K-tuple $(R_1, \ldots R_K)$ is achievable if there exists a sequence of codewords

$$\mathbf{c}(m_1, \ldots, m_K) \in \mathcal{X}^n, m_i \in \{1, \ldots, M_i\}, i = 1, \ldots K$$

and disjoint decoding sets[6] such that

$$\lim_{n \to \infty} P_{Y^n \mid X^n, S}(\mathcal{D}_i(m_1, \ldots, m_i) \,|\, \mathbf{c}(m_1, \ldots, m_K), i) = 1 \tag{44}$$

for all $m_i \in \{1, \ldots, M_i\}$, and

$$\liminf_{n \to \infty} \frac{\log M_i}{n} \geq R_i \tag{45}$$

for all $i = 1, \ldots K$. Note that (44) implies that the ith user requires reliable reception of the messages reliably received by users $1, \ldots, i-1$. The closure of the set of achievable K-tuples is the capacity region of the K-user broadcast channel with degraded message sets. Since the ordering of the users matters, it is convenient to specify the ordering in the notation for capacity region: $\mathcal{C}^{\mathsf{p}}$ is the capacity region of the K-user broadcast channel with degraded message sets

$$\left\{\left\{P_{Y^n \mid X^n, S=\mathsf{p}(i)}\right\}_{n=1}^{\infty}, i = 1, \ldots, K\right\} \tag{46}$$

where p is a K-vector obtained as a permutation of $\{1, \ldots, K\}$.

[5] Definition 8 adopts the special case in which the output alphabets of the channels seen by all the users are identical. Dropping this restriction from the definition is straightforward but unduly general for our purposes.

[6] $\mathcal{D}_i(\mathbf{m}) \cap \mathcal{D}_i(\mathbf{m}') = \emptyset$ if $\mathbf{m} \neq m'$, and, thus, $\mathcal{D}_i^{-1}$ is well defined.

Theorem 4: [15], [20]. The capacity region of the two-user discrete memoryless broadcast channel with degraded message sets (where any information destined for user 1 is also destined for user 2) is given by (47), shown at the bottom of the page, where U and (Y_1, Y_2) are independent conditioned on X, and the cardinality of U is upperbounded by $|\mathcal{X}|$ [23].

Theorem 5: The variable-to-fixed channel capacity of (41) satisfies

$$\max_{\mathsf{p}} \max_{(R_1,\ldots,R_K)\in\mathcal{C}^{\mathsf{p}}} \sum_{j=1}^{K} R_j \sum_{i=j}^{K} \pi_{\mathsf{p}^{-1}(i)} \le \langle \mathsf{C} \rangle \tag{48}$$

where $\mathcal{C}^{\mathsf{p}}$ is the capacity region of the broadcast channel with degraded message sets (46) obtained from (41).

Proof: Fix an arbitrary permutation p, choose an arbitrary point $(R_1, \ldots, R_K)$ in the capacity region $\mathcal{C}^{\mathsf{p}}$, and a broadcast code sequence that achieves that point. To simplify notation, we assume in the remainder of the proof that p is the identity. Thus, the goal is to show that

$$\sum_{j=1}^{K} R_j \sum_{i=j}^{K} \pi_i \le \langle \mathsf{C} \rangle \tag{49}$$

Without loss of generality we consider code sizes that are powers of 2 for each user and denote

$$k_i = \log_2 M_i \tag{50}$$

$$m_j = (b_{k_1+\cdots k_{j-1}+1}, \ldots b_{k_1+\cdots+k_j}) \tag{51}$$

where $b_1, \ldots, b_{k_1+\ldots+k_K}$ are the information bits available at the decoder. Note that $k_1 + \ldots + k_K \le m_n$, and we only need to specify the action of the encoder for $k_1 + \ldots + k_K$ bits. The single-user variable-to-fixed encoder/decoder is

$$\mathsf{f}^{:n}(m_1, \ldots, m_K) = \mathbf{c}(m_1, \ldots, m_K) \tag{52}$$

$$\mathsf{g}^{n:}(y^n) = (\mathcal{D}_{\hat{i}}^{-1}(y^n), \mathsf{a}, \ldots, \mathsf{a}) \tag{53}$$

where the tail consists of $k_{\hat{i}+1} + \cdots + k_K$ repetitions of an arbitrary symbol $\mathsf{a} \in \mathcal{Y}$. Since we are not assuming side information at the receiver, the index $\hat{i}$ is the decoder's best guess of the channel state on the basis of y^n. Since there are only a finite number of states, with a negligible loss in rate it is possible to send a training sequence that will render the probability of erroneous state detection a fraction of the probability of message decoding error. Therefore, the construction inherits the vanishing error probability of the broadcast code sequence. To evaluate the achieved variable-to-fixed rate, note that

$$\mathbb{P}[L_n \ge k_1 + \cdots k_i \,|\, S = i] \ge \mathbb{P}[\hat{i} = i \,|\, S = i] \tag{54}$$

which goes to 1 by construction. Therefore

$$\liminf_{n\to\infty} \frac{1}{n}\mathbb{E}[L_n] \ge \liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{K} \pi_i (k_1 + \cdots k_i) \tag{55}$$

$$\ge \sum_{i=1}^{K} \pi_i (R_1 + \cdots + R_i) \tag{56}$$

$$= \sum_{j=1}^{K} R_j \sum_{i=j}^{K} \pi_i \tag{57}$$

where (56) follows from (45). ∎

It is easy to see that (48) is not necessarily tight in the absence of any restrictions on the constituent channels in (41): just take $K = 1$ and the channel in Example 1. The left side of (48) is 0, while the right side is $q\, C_0$. For this reason, the "expected capacity" of [11] differs from the variable-to-fixed capacity whenever there is a state under which the channel does not behave ergodically.

Theorem 6: Suppose that the K-user broadcast channels derived from (41) satisfy the strong converse. Then, (48) holds with equality.

Proof: Suppose that R is an achievable variable-to-fixed rate, and choose a sequence of variable-to-fixed codes $(\mathsf{f}^{:n}, \mathsf{g}^{n:})$ that satisfies (19). Now number the channel states in order of increasing expected number of recovered bits

$$r_i = \liminf_{n\to\infty} \frac{k_i}{n} \tag{58}$$

where

$$k_i = \mathbb{E}[L_n \,|\, S = i] \tag{59}$$

and denote for an arbitrarily small $\delta > 0$

$$R_1 = r_1 - \delta \tag{60}$$

$$R_i = r_i - r_{i-1}. \tag{61}$$

Consider the broadcast channel code

$$\mathbf{c}(m_1, \ldots, m_K) = \mathsf{f}^{:n}(m_1, \ldots, m_K) \tag{62}$$

for $m_i \in \{0,1\}^{\lfloor nR_i \rfloor}$, and

$$\mathcal{D}_i^{-1}(y^n) = \mathsf{g}^{n:(m_1+\cdots+m_i)}(y^n). \tag{63}$$

For each state i, $P_{Y^n \,|\, X^n, S}(\mathcal{D}_i(m_1, \ldots, m_i) \,|\, \mathbf{c}(m_1, \ldots, m_K), i)$ is bounded away from zero, as otherwise it would be impossible to achieve the average lengths in (59). This means that for some ϵ sufficiently close to 1, $(R_1, \ldots, R_K)$ is an ϵ-achievable K-tuple for the broadcast

$$\mathcal{C}^{(1,2)} = \bigcup_{P_{X,U}} \left\{ (R_1, R_2) : \begin{array}{rcl} R_1 & \le & \min\{I(U;Y_1), I(U;Y_2)\} \\ R_2 & \le & I(X;Y_2 \,|\, U) \end{array} \right\} \tag{47}$$

channel with degraded message sets (defined as in Definition 8 except that the liminf of the left side of (44) exceeds $1-\epsilon$. Finally, under the assumption of the theorem, the broadcast channel satisfies the strong converse; therefore, $(R_1,\ldots,R_K)$ is an achievable K-tuple for the broadcast channel, such that

$$R = \sum_{i=1}^{K} \pi_i \liminf_{n\to\infty} \frac{k_i}{n} \tag{64}$$

$$= \delta + \sum_{i=1}^{K} \pi_i (R_1 + \cdots + R_i) \tag{65}$$

$$= \delta + \sum_{j=1}^{K} R_j \sum_{i=j}^{K} \pi_i \tag{66}$$

where (65) follows from (58)–(61). But since δ is arbitrarily small we conclude that that there exists a permutation of the users of the broadcast channel such that

$$R \le \max_{(R_1,\ldots,R_K)\in\mathcal{C}^{\mathsf{p}}} \sum_{j=1}^{K} R_j \sum_{i=j}^{K} \pi_{\mathsf{p}^{-1}(i)}. \tag{67}$$

■

Theorem 7: Suppose that the channel has finite input/output alphabets $\mathcal{X}$ and $\mathcal{Y}$ and

$$P_{Y^n\,|\,X^n}(b^n\,|\,a^n) = \pi_1 \prod_{i=1}^{n} W_1(b_i\,|\,a_i) + \pi_2 \prod_{i=1}^{n} W_2(b_i\,|\,a_i) \tag{68}$$

where $\pi_1+\pi_2=1$ and $\sum_{b\in\mathcal{Y}} W_1(b\,|\,a) = \sum_{b\in\mathcal{Y}} W_2(b\,|\,a) = 1$ for all $a\in\mathcal{X}$. The variable-to-fixed capacity of the channel is equal to[7]

$$\langle \mathsf{C} \rangle = \max_{P_{XU}} \{ \min\{ I(U,W_1), I(U,W_2) \} + \max\{ \pi_1 I(X,W_1\,|\,U), \pi_2 I(X,W_2\,|\,U) \} \} \tag{69}$$

where the cardinality of U does not exceed $|\mathcal{X}|$, and conditioned on X the outputs are independent of U.

Proof: It is shown in [15] that when the constituent channels are discrete memoryless, the broadcast channel with degraded message sets satisfies the strong converse. Thus, (69) follows from Theorems 4 and 6. ■

In the most important special case, the constituent channels are memoryless and degraded versions of each other, i.e., with probability π_k the receiver observes Y_k, where $X \leftrightarrow Y_1 \leftrightarrow Y_2 \leftrightarrow \cdots \leftrightarrow Y_K$ form a Markov chain. Then, (48) is satisfied with equality and

$$\langle \mathsf{C} \rangle = \max_{U_1,\ldots,U_K} I(U_K;Y_K) + \sum_{k=1}^{K-1} (\pi_1 + \cdots + \pi_k) I(U_k;Y_k\,|\,U_{k+1}) \tag{70}$$

[7]As in [8], $I(U,W) = I(U;Y)$ where $P_{UY}(u,y) = P_U(u)W(y\,|\,u)$.

where the optimization is with respect to $U_1,\ldots,U_K$ such that $U_1 = X$ and

$$U_K \leftrightarrow \cdots \leftrightarrow U_1 \leftrightarrow Y_1 \leftrightarrow \cdots \leftrightarrow Y_K.$$

IV. Fixed-to-Variable Capacity

A. Relationship to Fixed-to-Fixed Capacity

For the rateless setup of fixed-to-variable capacity it makes sense to restrict attention to nonanticipatory channels defined by a sequence of conditional probability distributions

$$\left\{ P_{Y_i\,|\,X_1^i, Y_1^{i-1}} \right\}_{i=1}^{\infty}. \tag{71}$$

This is the most general class of channels considered when dynamical issues such as feedback are considered (e.g., [31]).

Theorem 8: The Shannon capacity C and the upper fixed-to-variable capacity $[\bar{\mathsf{C}}]$, satisfy

$$C \le [\bar{\mathsf{C}}]. \tag{72}$$

Proof: Suppose $(f^{k\to n_k}, g^{n_k\to k})$ is a sequence of k-to-n_k codes with error probability ϵ_k such that

$$\liminf_{k\to\infty} \frac{k}{n_k} = C \tag{73}$$

$$\lim_{k\to\infty} \epsilon_k = 0. \tag{74}$$

We will now show that C is an upper fixed-to-variable achievable rate. To that end, define a fixed-to-variable code such that for each k

$$\mathsf{f}^{k:}(b_1,\ldots,b_k) = [f^{k\to n_k}(b_1,\ldots,b_k), \mathsf{a}, \mathsf{a}, \ldots] \tag{75}$$

where $\mathsf{a}\in\mathcal{X}$, and $\mathsf{g}^{:k}$ is such that

$$\mathsf{g}^{n_k:k} = g^{n_k\to k} \tag{76}$$

and is immaterial how $\mathsf{g}^{n:k}$ is defined for $n \ne n_k$. The nonanticipatory assumption ensures that inputs beyond n_k do not affect the conditional probabilities of Y^{n_k}. If the $(f^{k\to n_k}, g^{n_k\to k})$ code succeeds in recovering the message, then the random variable k/N_k is lower bounded by k/n_k; otherwise we simply lower bound k/N_k by 0. Thus

$$\liminf_{n\to\infty} \mathbb{E}\left[\frac{k}{N_k}\right] \ge \liminf_{n\to\infty} \frac{k(1-\epsilon_k)}{n_k} \tag{77}$$

$$= C \tag{78}$$

where (78) follows from (73) and (74). Thus, C is an upper fixed-to-variable achievable rate. ■

Under a mild technical assumption, we can also prove that the fixed-to-variable capacity is lower bounded by Shannon capacity.

Theorem 9: Suppose that the channel has finite memory in the sense that there is an integer ℓ such that given $X_{i-\ell}^{i}$ and $Y_{i-\ell}^{i-1}$, Y_i is independent of $X_1^{i-\ell-1}$ and $Y_1^{i-\ell-1}$. Then

$$C \le [\mathsf{C}]. \tag{79}$$

Proof: The proof of Theorem 8 is modified so that N_k is finite with probability 1. Now, the rateless encoder sends repeatedly the same codeword interspersed with ℓ dummy symbols which ensure that previous codewords do not affect the reception of future codewords

$$\begin{aligned}&\mathsf{f}^{k:}(b_1,\ldots,b_k)\\ &\quad=[f^{k\to n_k}(b_1,\ldots,b_k),\mathsf{a},\ldots,\mathsf{a},f^{k\to n_k}(b_1,\ldots,b_k)\\ &\qquad\mathsf{a},\ldots,\mathsf{a},f^{k\to n_k}(b_1,\ldots,b_k),\ldots]\end{aligned} \tag{80}$$

If the decoder is successful in decoding $(b_1,\ldots,b_k)$ for the first time at the jth attempt, then

$$N_k = j(n_k+\ell) \tag{81}$$

But this happens with probability $(1-\epsilon_k)\epsilon_k^{j-1}$. Therefore

$$\frac{k}{\mathbb{E}[N_k]} = \frac{k}{(n_k+\ell)(1-\epsilon_k)\sum_{j=1}^{\infty} j\epsilon_k^{j-1}} \tag{82}$$

$$= \frac{k(1-\epsilon_k)}{n_k+\ell} \tag{83}$$

and the $\liminf$ of (83) is C because of (73) and (74). ∎

The counterpart of Theorem 2 for fixed-to-variable capacity is

Theorem 10: Assume that either the input or output alphabets are finite. Then

$$[\mathsf{C}] \le [\bar{\mathsf{C}}] \le \lim_{\epsilon\uparrow 1} C_\epsilon. \tag{84}$$

Proof: The left inequality in (84) was given in (20).

For convenience, we prevent the decoder from starting to make guesses about the transmitted k bits until $\lceil k/\bar{R}\rceil$ channel symbols have been received. In view of the assumption of finite input or output alphabet, this entails no loss of asymptotic fixed-to-variable rate provided sufficiently large $\bar{R} > 0$ is chosen. This is because when only a tiny fraction of channel symbols have been observed, the decoder is forced to make essentially uninformed guesses about the transmitted information. The beneficial effect of such crippling is purely technical: it enables the upper bound

$$\mathbb{E}\left[\frac{k}{N_k}\right] \le \bar{R}. \tag{85}$$

The goal is to show that for some $\epsilon_0 < 1$, we can construct a conventional code sequence whose error probability is not worse that ϵ_0, and whose rate is $[\bar{\mathsf{C}}]-\delta$, regardless how small $\delta > 0$ is selected. To that end, select a fixed-to-variable encoder/decoder sequence $(\mathsf{f}^{k:},\mathsf{g}^{:k})$ that achieves $[\bar{\mathsf{C}}]$. Denote the channel blocklength

$$n_k = \frac{k}{[\bar{\mathsf{C}}]-\delta} \tag{86}$$

and select the fixed-to-fixed encoder/decoder

$$f^{k\to n_k} = \mathsf{f}^{k:n_k} \tag{87}$$

$$g^{n_k\to k} = \mathsf{g}^{n_k:k}. \tag{88}$$

Arguing by contradiction, assume that there is a subsequence of k, along which the error probability of $(f^{k\to n_k}, g^{n_k\to k})$ is not bounded away from 1. Then, along that subsequence

$$\lim_{i\to\infty}\mathbb{E}\left[\frac{k_i}{N_{k_i}}\right] < \lim_{i\to\infty}\frac{k_i}{n_{k_i}} \tag{89}$$

$$= [\bar{\mathsf{C}}]-\delta \tag{90}$$

where (89) follows because

$$\lim_{i\to\infty}\mathbb{P}[N_{k_i} > n_{k_i}] = 1 \tag{91}$$

and (85). But (90) contradicts the assumption that

$$\liminf_{k\to\infty}\mathbb{E}\left[\frac{k}{N_k}\right] = [\bar{\mathsf{C}}]. \tag{92}$$

∎

Remark 9: As in Remark 8, if the channel satisfies the strong converse (40) for all Δ known to both receiver and transmitter, then

$$[\mathsf{C}] = [\bar{\mathsf{C}}] = \mathbb{E}[C(\Delta)] \tag{93}$$

when Δ is known at both encoder and decoder.

B. Relationship to Variable-to-Fixed Capacity

For channels that satisfy the strong converse, Theorems 1, 2, 8, 10 imply that

$$C = \langle\mathsf{C}\rangle = [\bar{\mathsf{C}}]. \tag{94}$$

Under a very mild assumption we have the following result.

Theorem 11: Suppose the channel is such that the upper fixed-to-variable capacity in Definition 6 coincides with the same concept defined with $\limsup$ in lieu of $\liminf$ in (19). Then

$$\langle\mathsf{C}\rangle \le [\bar{\mathsf{C}}]. \tag{95}$$

Proof: Suppose that $0 < R \le \langle\mathsf{C}\rangle$, and we are given a sequence $\{\mathsf{f}^{:n},\mathsf{g}^{:n}\}$ of rate-R variable-to-fixed codes (Section II.B). In the sequential nonanticipatory setting (71) of this section, it makes sense to focus attention without loss of generality on compatible variable-to-fixed encoder sequences

where $\mathsf{f}^{:n}(b_1,\ldots,b_{m_n})$ consists of all but the last component in $\mathsf{f}^{:(n+1)}(b_1,\ldots,b_{m_n})$.

We construct a sequence of rateless encoders (15) by letting the ith component of $\mathsf{f}^{k:}(b_1,\ldots,b_k)$ be equal to the ith component of $\mathsf{f}^{:n}(b_1,\ldots,b_{m_n})$ for n sufficiently large (well defined because of the compatibility assumption). Moreover, $\mathsf{g}^{n:k}(y^n)$ is equal to the first k bits in the variable-to-fixed decoder $\mathsf{g}^{n:}(y^n)$. To relate the performance of the constructed fixed-to-variable code to that of the original variable-to-fixed code, note that according to Definitions 2 and 4

$$N_{L_i} \le i. \tag{96}$$

for every positive integer i. Therefore

$$\frac{L_n}{n} \le \frac{L_n}{N_{L_n}} \tag{97}$$

but since L_n grows without bound we must have

$$R = \liminf_{n\to\infty} \mathbb{E}\left[\frac{L_n}{n}\right] \tag{98}$$

$$\le \limsup_{n\to\infty} \mathbb{E}\left[\frac{k}{N_k}\right] \tag{99}$$

and we conclude that according to the assumption of Theorem 11, R is shown to be an achievable fixed-to-variable rate. ∎

Remark 10: The proof actually shows more than the statement of Theorem 11 since in those rare cases that do not satisfy the assumption, $\langle \mathsf{C} \rangle$ is still upper bounded by the "optimistic" definition of fixed-to-variable capacity. See [5] (and [35]) regarding the contrast between "optimistic" and conventional definitions.

Remark 11: It is easy to find channels for which $[\mathsf{C}] < \langle \mathsf{C} \rangle$: if with nonnegligible probability the channel conditions are very bad, then the expectation in (18) is dominated by very large values. (See Examples 1, 4, and 5.)

C. State-Dependent Discrete Memoryless Channels

As mentioned in Section I, the conventional fixed-to-fixed setting is unable to capitalize on the prior distribution $\pi_1,\ldots\pi_K$ and the capacity is equal to that found in [1] for the compound channel $\mathcal{W} = \{W_1,\ldots,W_K\}$:

Theorem 12: The conventional capacity of (41) is

$$C = \max_P \min_\ell I(P, W_\ell). \tag{100}$$

Proof: It follows specializing the general formula in [35].

Theorem 13: The fixed-to-variable capacity of the channel in (41) satisfies

$$[\mathsf{C}] \ge \left(\min_{P_X} \sum_{\ell=1}^{K} \frac{\pi_\ell}{I(P_X, W_\ell)}\right)^{-1} \tag{101}$$

Proof: With negligible loss of efficiency we can send a training sequence that will enable the decoder to identify $S = 1,\ldots K$, the actual channel in effect with vanishing probability of error. Thus, for simplicity we assume that the decoder knows S. Fix an input distribution P_X and $\delta > 0$. For convenience assume that the channels are labeled such that

$$I(P_X, W_K) \le I(P_X, W_{K-1}) \le \cdots \le I(P_X, W_1). \tag{102}$$

Using a random coding argument where symbols are chosen independently with distribution P_X, we can a find a code that maps k information symbols to $\frac{k}{I(P_X,W_K)-\delta}$ channel symbols such that for all $\ell = 1,\ldots K$, if $S = \ell$ then after

$$n_{k,\ell} = \frac{k}{I(P_X, W_\ell) - \delta} \tag{103}$$

channel symbols a decoder (that has side information about S) outputs the k information bits with probability of error not greater than ϵ_k, where $\epsilon_k \to 0$. After the transmission of this code, the encoder transmits repeatedly a code with blocklength $\frac{2k}{C}$ where C is given in (100). Since this code operates at a fraction of the compound capacity there exists $\rho < 1$ such that we can find a code with error probability lower than ρ^k for sufficiently large k [8, p. 173]. We proceed to upper bound $\mathbb{E}[N_k]$ for this code construction. For every $i = 0, 1, \ldots$

$$\mathbb{P}\left[N_k > n_{k,\ell} + \frac{2ki}{C} \middle| S = \ell\right] \le \epsilon_k \rho^{ik}. \tag{104}$$

Using (104), we get

$$\mathbb{E}[N_k] = \sum_{j=0}^{\infty} \mathbb{P}[N_k > j] \tag{105}$$

$$= \sum_{\ell=1}^{K} \pi_\ell \sum_{j=0}^{\infty} \mathbb{P}[N_k > j \,|\, S = \ell] \tag{106}$$

$$\le \sum_{\ell=1}^{K} \pi_\ell \left(n_{k,\ell} + \frac{2k\epsilon_k}{C} \sum_{i=0}^{\infty} \rho^{ik}\right) \tag{107}$$

$$= \frac{2k\epsilon_k}{C(1-\rho^k)} + \sum_{\ell=1}^{K} \pi_\ell \frac{k}{I(P_X, W_\ell) - \delta} \tag{108}$$

which implies that

$$\liminf_{k\to\infty} \frac{k}{\mathbb{E}[N_k]} \ge \left(\sum_{\ell=1}^{K} \frac{\pi_\ell}{I(P_X, W_\ell) - \delta}\right)^{-1}. \tag{109}$$

Finally, since $\delta > 0$ and P_X were chosen arbitrarily, (109) indicates that $\ge$ holds in (101). ∎

The bound in Theorem 13 is not tight for some channels of the form (41) as illustrated in Example 3. However, it is indeed tight for a sub-class of those channels as the following result shows.

Theorem 14: Suppose that in (41), the constituent channels $W_1,\ldots,W_K$ are such that there is an input distribution P_X^* that maximizes mutual information simultaneously, i.e.,

$$\max_{P_X} I(P_X, W_\ell) = I(P_X^*, W_\ell) = C_\ell. \tag{110}$$

Then, (101) holds with equality and the fixed-to-variable capacity is equal to the harmonic mean of the capacities

$$[\mathsf{C}] = \left(\sum_{\ell=1}^{K} \frac{\pi_\ell}{C_\ell} \right)^{-1}. \tag{111}$$

Proof: To prove this converse result, we assume that the decoder knows the channel state S, fix k, and fix an arbitrary $(\mathsf{f}^{k:}, \mathsf{g}^{:k})$ encoder/decoder pair. As in Definition 4, the message is denoted by $B^k \in \{0,1\}^k$. The data processing theorem implies that for any of the possible channel states $s \in \{1, \ldots, K\}$

$$I(B^k; \mathsf{g}^{n:k}(Y^n) \,|\, S = s) \le I(\mathsf{f}^{k:}(B^k); Y^n \,|\, S = s) \tag{112}$$

$$= I(\mathsf{f}^{k:n}(B^k); Y^n \,|\, S = s) \tag{113}$$

$$\le nI(P_{\hat{X}_{k:n}}, W_s) \tag{114}$$

$$\le nI(P_X^*, W_s) \tag{115}$$

$$= nC_s \tag{116}$$

where (113) follows from the memorylessness of channel W_s; the empirical marginal distribution of the first n channel input symbols is denoted by

$$P_{\hat{X}_{k:n}} = \frac{1}{n} \sum_{i=1}^{n} P_{\mathsf{f}^{k:i}(B^k)} \tag{117}$$

and again (114) follows from the memorylessness of channel W_s and the concavity of mutual information in the input distribution. We now proceed to average both sides of (112)–(116)

$$C_s \mathbb{E}[N_k \,|\, S = s]$$

$$= \sum_{n=1}^{\infty} \mathbb{P}[N_k = n \,|\, S = s] nC_s \tag{118}$$

$$\ge \sum_{n=1}^{\infty} \mathbb{P}[N_k = n \,|\, S = s] I(B^k; \mathsf{g}^{n:k}(Y^n) \,|\, S = s) \tag{119}$$

$$= k - \sum_{n=1}^{\infty} \mathbb{P}[N_k = n \,|\, S = s] \times H(B^k \,|\, \mathsf{g}^{n:k}(Y^n), S = s) \tag{120}$$

$$= k - \sum_{n=1}^{\infty} \mathbb{E}[1\{N_k = n\} H(B^k \,|\, \mathsf{g}^{N_k:k}(Y^{N_k}) \, S = s) \,|\, S = s] \tag{121}$$

$$= k \tag{122}$$

where (119) follows from (112)–(116) and (122) follows from the definition of N_k (Definition 4). Therefore

$$\frac{\mathbb{E}[N_k]}{k} = \sum_{s=1}^{K} \pi_s \frac{\mathbb{E}[N_k \,|\, S = s]}{k} \tag{123}$$

$$\ge \sum_{s=1}^{K} \frac{\pi_s}{C_s} \tag{124}$$

and the result follows from the definition of $[\mathsf{C}]$ and Theorem 13. ∎

Theorem 15: The upper fixed-to-variable capacity of (41) is equal to

$$[\bar{\mathsf{C}}] = \max_{P_X} \sum_{\ell=1}^{K} \pi_\ell I(P_X, W_\ell). \tag{125}$$

Proof: Achievability: The proof of achievability is easier than the achievability proof in Theorem 13 since atypically large values of N_k do not pose a challenge in this case. Again, we sidestep the analysis of the negligible loss incurred by using a training sequence and we assume the decoder has knowledge of the ergodic mode in effect. A standard random coding argument together with the type of bounding of the ratio k/N_k we used in the proof of Theorem 8 yields that $I(P_X, W_\ell)$ is an asymptotically achievable $\mathbb{E}[\frac{k}{N_k} \,|\, S = \ell]$.

Converse: Fix k and a rateless encoder $\mathsf{f}^{k:}$. Suppose that the decoder is informed by a genie that the DMC actually in effect is W_ℓ. If the decoder outputs $\hat{B}_1, \ldots, \hat{B}_k$ upon examination of the first n channel outputs

$$\frac{1}{n} I(B_1, \ldots B_k; \hat{B}_1, \ldots, \hat{B}_k)$$

$$\le \frac{1}{n} I(\mathsf{f}^{k:}(B_1, \ldots, B_k); Y^n) \tag{126}$$

$$\le I(Q_X, W_\ell) \tag{127}$$

where the mutual informations assume that the channel is W_ℓ, (126) follows from the data processing theorem and (127) follows from the memorylessness of the channel where Q_X is the mixture of the n first marginals of $\mathsf{f}^{k:}(B_1, \ldots, B_k)$. The ratio in the left side of (126) is such that when $n \ge N_k$, the numerator is equal to k. Therefore

$$\mathbb{E}\left[\frac{k}{N_k} \middle| S = \ell \right] \le I(Q_X, W_\ell). \tag{128}$$

Averaging with respect to the state of the channel, and further upper bounding the expression by choosing the best possible input distribution, we obtain the right side of (101). ∎

Another conclusion from (20), and Theorems 13 and 15 is that whenever (125) equals the right side of (101), then $[\bar{\mathsf{C}}] = [\mathsf{C}]$ (e.g., Example 7).

We omit the straightforward technicalities required for the generalization of the results in this subsection to the case of countably or uncountably infinite channels.

V. Variable-Blocklength Capacity

In this section we consider the variable-blocklength capacity (Definition 7) where the encoder chooses the blocklength as a function of the message, and the decoder is able to obtain the value of the blocklength noiselessly. This is an idealization of, for example, a channel where the presence of a transmitted signal is detected by the presence of a carrier at a certain frequency. The rate is gauged by the ratio of transmitted bits to average blocklength.

Theorem 16: For any nonanticipatory channel (71)

$$C = [\![\mathsf{C}]\!]. \tag{129}$$

Proof: By definition, $C \le [\![\mathsf{C}]\!]$. The essence of the idea why $C < [\![\mathsf{C}]\!]$ is impossible can be best appreciated in the simplest channel: a binary noiseless channel with $C = 1$ bit. The codebook consists of M binary strings, each chosen equiprobably. It is wasteful to include a string of length ℓ unless all the binary strings of length $\ell - 1$ have been used. Therefore[8]

$$M = 2^1 + 2^2 + \cdots + 2^J \tag{130}$$

or

$$J = \log_2\left(1 + \frac{M}{2}\right). \tag{131}$$

The rate (bits to average blocklength) is

$$\frac{\log_2 M}{\frac{1}{M}\sum_{j=1}^{L} j2^j} = \frac{M\log_2 M}{2 + (J-1)2^{J+1}} \tag{132}$$

$$= \frac{M\log_2 M}{2 + \left(\log_2\left(1+\frac{M}{2}\right) - 1\right)(M+2)}. \tag{133}$$

If $M = 10$, the average rate is 1.58. However, if we are forced, as we are in the definition of both C and $[\![\mathsf{C}]\!]$, to let $M \to \infty$, the asymptotic rate $[\![\mathsf{C}]\!] = 1$. Even though we can communicate the length for free, the amount of information that it provides pales in comparison with that provided by the payload in the asymptotic regime. It is easy to extend this analysis to any noisy nonanticipatory channel. We can take the optimistic view that if there are $2^{\ell C}$ codewords of length ℓ, they can be decoded noiselessly by the decoder, and since we are interested in vanishing error probability, it is futile to use $2^{\ell(C+\delta)}$ codewords of length ℓ with nonnegligible probability. While this must hold only for asymptotically long blocklength, it is safe to assume it for all blocklengths since in the limit, short blocklengths do not contribute to the variable-length rate because of the nonanticipatory assumption. Following the same steps above, now

$$M = \sum_{\ell=1}^{J} 2^{\ell C} \tag{134}$$

and the variable-length rate of the idealized error-free scheme is

$$\frac{\log_2 M}{\frac{1}{M}\sum_{\ell=1}^{L} \ell 2^{\ell C}}$$

$$= \frac{(2^C-1)(2^{JC}-1)\log_2\left((2^{JC}-1)\frac{2^C}{2^C-1}\right)}{1 + 2^{JC}((2^C-1)J - 1)} \tag{135}$$

$$\to C \tag{136}$$

as M (and, thus, J) goes to ∞. ∎

[8]At the expense of some cumbersomeness, we can deal with any integer M by taking only a subset of the longest strings.

Note that in the general setup of (2), if the input/output alphabets are infinite, we can have $C = 0$ and $[\![C]\!] = \infty$ without the sufficient condition in Theorem 16: consider a channel with alphabets $\mathcal{X} = \mathcal{Y} = \{0, 1, 2, \ldots\}$, such that if a string of length 1 is sent, the output is equal to the input, while for any other string length, the output is equal to $(0, \ldots, 0)$. As it depends only on the tail conditional probabilities in (2), the capacity is zero, but the message (regardless of how many bits it contains) can be delivered noiselessly with a unit-length string.

The nonvanishing error probability version of variable-length channel capacity $[\![\mathsf{C}_\epsilon]\!]$ is more interesting: it is equal to the Shannon capacity C divided by $1 - \epsilon$ as long as the channel satisfies the strong converse and is nonanticipatory [22], a result which was shown in the discrete memoryless case in [2] (see [8, Problem 2.1.25]).

VI. Examples

Example 1: Suppose that with probability q the channel is a discrete memoryless channel with capacity C_0, and with probability $1 - q$, the output is independent of the input. Then, it follows easily from the various definitions that

$$C = [\mathsf{C}] = 0 \tag{137}$$

$$\lim_{\epsilon \uparrow 1} C_\epsilon = C_0 \tag{138}$$

$$\langle \mathsf{C} \rangle = [\bar{\mathsf{C}}] = q\, C_0. \tag{139}$$

Example 2: For the whole duration of the codeword, the channel is:

- with probability $0 < \pi_0 < 1$: a binary symmetric channel with crossover probability $\delta_0 \in [0, \frac{1}{2}]$;
- with probability $\pi_1 = 1 - \pi_0$: a binary symmetric channel with crossover probability $\delta_1 \in (\delta_0, \frac{1}{2}]$.

Therefore

$$P_{Y^n \mid X^n}(b^n \mid a^n) = \pi_0 \delta_0^{w_H(a_n,b_n)}(1-\delta_0)^{n - w_H(a_n,b_n)} + \pi_1 \delta_1^{w_H(a_n,b_n)}(1-\delta_1)^{n-w_H(a_n,b_n)} \tag{140}$$

where $w_H(a_n, b_n)$ is the Hamming distance between a_n and b_n. While the conventional fixed-to-fixed capacity is

$$C = 1 - h(\delta_1). \tag{141}$$

Theorem 7 specializes to

$$\langle \mathsf{C} \rangle = 1 - \pi_0 h(\delta_0) - \min_{0 \le \gamma \le 1} \{h(\gamma * \delta_1) - \pi_0 h(\gamma * \delta_0)\} \tag{142}$$

where

$$\gamma * \delta = (1-\delta)\gamma + (1-\gamma)\delta = \delta + \gamma(1 - 2\delta). \tag{143}$$

Theorem 14 yields

$$[\mathsf{C}] = \left(\frac{\pi_0}{1 - h(\delta_0)} + \frac{\pi_1}{1 - h(\delta_1)}\right)^{-1} \tag{144}$$

while Theorem 15 yields

$$[\bar{\mathsf{C}}] = 1 - \pi_0 h(\delta_0) - \pi_1 h(\delta_1). \tag{145}$$

Fig. 1 compares (141), (142), (144), and (145), for $\delta_0 = 0, \delta_1 = 0.11$, and $\pi_0 = 1 - q$.

Example 3: Consider a channel with input/output alphabet equal to $\{0, \dots 1023\}$. With probability 0.9 the output is always equal to the input; with probability 0.1 the channel is such that all input symbols different from 0 are mapped to 1, and 0 is mapped to 0. The right side of (101) is equal to

$$\max_{0<\alpha<1} \left(\frac{0.9}{h(\alpha) + (1-\alpha)\log_2(1023)} + \frac{0.1}{h(\alpha)} \right)^{-1} = 4.4. \tag{146}$$

However, consider the following suboptimal scheme: The information bits are grouped in blocks of 10 bits which address the inputs of the channel; then, the raw bits are sent through the channel. Therefore, $N_k = \frac{k}{10}$ with probability 0.9, while $N_k = \frac{11k}{10}$ with probability 0.1. Accordingly, this scheme achieves

$$\frac{\mathbb{E}[N_k]}{k} = \frac{1}{5} \tag{147}$$

and, consequently, $[\mathsf{C}] \geq 5$.

Example 4: Binary erasure channel. We now consider a binary erasure channel whose erasure probability $0 \leq E \leq 1$ stays constant during the duration of the codeword and is a random variable with

$$F(x) = \mathbb{P}[1 - E \leq x]. \tag{148}$$

Note that if the erasure probability were known at the transmitter, the variable-rate capacity would equal

$$\mathbb{E}[C(E)] = 1 - \mathbb{E}[E] = 1 - \int_0^1 F(x)\, dx. \tag{149}$$

Since equiprobable inputs achieve the maximal mutual information regardless of the erasure probability

$$[\mathsf{C}] = \left(\mathbb{E}\left[\frac{1}{1-E} \right] \right)^{-1} \tag{150}$$

$$= \left(\int_1^\infty F\left(\frac{1}{\alpha}\right)\, d\alpha \right)^{-1} \tag{151}$$

$$[\bar{\mathsf{C}}] = 1 - \mathbb{E}[E] \tag{152}$$

$$= 1 - \int_0^1 F(x)\, dx. \tag{153}$$

To find $\langle \mathsf{C} \rangle$ note that the capacity region of the degraded erasure broadcast channel is achieved by TDMA (see, e.g., [3]); we apportion a fraction of the blocklength equal to $r(s)\, ds$ devoted to a channel whose capacity is s (and consequently has erasure probability $1 - s$. Thus, the actual rate achievable when the channel has capacity s is

$$R(s) = \int_0^s x r(x)\, dx. \tag{154}$$

Therefore, the goal is to choose the function $r(\,\cdot\,)$ to maximize the expected rate

$$\langle \mathsf{C} \rangle = \int_0^1 R(s) dF(s) \tag{155}$$

$$= \int_0^1 (1 - F(s)) dR(s) \tag{156}$$

$$= \int_0^1 (1 - F(s)) s r(s)\, ds \tag{157}$$

subject to

$$\int_0^1 r(x)\, dx = 1. \tag{158}$$

Therefore, the solution is

$$\langle \mathsf{C} \rangle = \max_{0 \leq s \leq 1} s - s\, F(s). \tag{159}$$

If the erasure probability is uniformly distributed between 0 and 1, then

$$C = [\mathsf{C}] = 0 \tag{160}$$

$$\langle \mathsf{C} \rangle = \frac{1}{4} \tag{161}$$

$$\mathbb{E}[C(E)] = [\bar{\mathsf{C}}] = \frac{1}{2}. \tag{162}$$

The TDMA-based approach used above can be applied to any state dependent channel. However, as we illustrate in Examples 5 and 6, the TDMA approach is, in general, suboptimal.

Example 5: Binary Symmetric Channel. We treat here a generalization of Example 2 where the crossover probability is a random variable $0 \leq \Delta \leq \frac{1}{2}$ with cumulative distribution function F_Δ. If Δ were known at the transmitter, the variable-rate capacities would coincide with the average capacity

$$\mathbb{E}[C(\Delta)] = 1 - \mathbb{E}[h(\Delta)]. \tag{163}$$

Since, equiprobable inputs maximize mutual information, we have

$$[\mathsf{C}] = \left(\mathbb{E}\left[\frac{1}{1 - h(\Delta)} \right] \right)^{-1} \tag{164}$$

$$[\bar{\mathsf{C}}] = 1 - \mathbb{E}[h(\Delta)] \tag{165}$$

which for crossover probability uniformly distributed on $[0, 1/2]$ becomes

$$C = [\mathsf{C}] = 0 \tag{166}$$

$$\mathbb{E}[C(\Delta)] = [\bar{\mathsf{C}}] = 1 - \frac{1}{4}\log_2 e = 0.639. \tag{167}$$

According to Theorem 6, solving for $\langle \mathsf{C} \rangle$, entails finding the average rate achieved in a degraded broadcast BSC-channel.[9]

[9]An equivalent optimization following the approach of [26], [27] is solved in [11].

If the encoder were to assume that there are K BSCs with crossover probabilities $0 \le \delta_1 < \cdots < \delta_K < 1/2$, the average rate would be according to (70)

$$\begin{aligned} R(\delta_1, \ldots, \delta_K) &= \max_{\beta_2,\ldots,\beta_K} F_\Delta(\delta_K) I(U_K; Y_K) \\ &\quad + \sum_{k=1}^{K-1} F_\Delta(\delta_k) I(U_k; Y_k \,|\, U_{k+1}) \qquad (168) \\ &= \max_{\beta_2,\ldots,\beta_K} F_\Delta(\delta_K)(1 - h(\delta_K * \mu_K)) \\ &\quad + \sum_{k=1}^{K-1} F_\Delta(\delta_k)[h(\delta_k * \mu_{k+1}) - h(\delta_k * \mu_k)] \qquad (169) \end{aligned}$$

where Y_k is the response to U_1 of a BSC with crossover probability δ_k, U_{k-1} is the response to U_k of a BSC with crossover probability δ_k, $a * b = (1-a)b + (1-b)a$ and

$$\mu_k = \beta_2 * \ldots * \beta_k. \qquad (170)$$

In the continuous limit, with $\delta_1 = \delta_{\min}$ and $\delta_K = \delta_{\max}$, (169) becomes

$$\begin{aligned} \langle \mathsf{C} \rangle &= \max F_\Delta(\delta_{\max})(1 - h(W(\delta_{\max})) \\ &\quad + \int_{\delta_{\min}}^{\delta_{\max}} F_\Delta(x) \left(\dot{W}(x) - \frac{1 - 2W(x)}{1 - 2x} \right) \\ &\quad \times \log_2 \frac{1 - W(x)}{W(x)}\, dx \qquad (171) \end{aligned}$$

where the maximization is over the function $W(x)$, which plays the role of $\delta_k * \mu_k$ for the channel with crossover probability $\delta_k = x$. Therefore

$$0 \le x \le W(x) \le 1/2 \qquad (172)$$

is monotonically nondecreasing. A feasible choice is $W(x) = x$, for which we can optimize (171) to yield

$$\langle \mathsf{C} \rangle \ge \max_{0 \le \delta_{\max}} F_\Delta(\delta_{\max})(1 - h(\delta_{\max})) \qquad (173)$$

which is what we would obtain by simply coding for a single BSC. To maximize the integral in (171) for a given choice of the extreme points $(\delta_{\min}, \delta_{\max})$ we apply the Euler–Lagrange formula (e.g., [17])

$$\frac{\partial \mathcal{F}}{\partial y}(x, W(x), \dot{W}(x)) = \frac{d}{dx} \frac{\partial \mathcal{F}}{\partial z}(x, W(x), \dot{W}(x)) \qquad (174)$$

to the functional

$$\mathcal{F}(x, y, z) = F_\Delta(x) \left(z - \frac{1 - 2y}{1 - 2x} \right) \log_2 \frac{1 - y}{y} \qquad (175)$$

with

$$\begin{aligned} &\frac{\partial \mathcal{F}}{\partial y}(x, y, z) \\ &\quad = \frac{2F_\Delta(x)}{1 - 2x} \left(\log_2 \frac{1 - y}{y} - \frac{1}{1 - y} \log_2 e \right) \\ &\qquad - \frac{F_\Delta(x)}{y(1 - y)} \left(z - \frac{1}{1 - 2x} \right) \log_2 e \qquad (176) \\ &\frac{\partial \mathcal{F}}{\partial z}(x, y, z) \\ &\quad = F_\Delta(x) \log_2 \frac{1 - y}{y} \qquad (177) \\ &\frac{d}{dx} \frac{\partial \mathcal{F}}{\partial z}(x, W(x), \dot{W}(x)) \\ &\quad = f_\Delta(x) \log_2 \frac{1 - W(x)}{W(x)} \\ &\qquad - \frac{\dot{W}(x) F_{\Delta \cdot}(x)}{W(x)(1 - W(x))} \log_2 e \qquad (178) \end{aligned}$$

Assembling (176)–(178), (174) becomes

$$\begin{aligned} &\frac{W(x)(1 - W(x))}{1 - 2W(x)} \log \frac{1 - W(x)}{W(x)} \\ &\qquad = \frac{F_\Delta(x)}{f_\Delta(x)(1 - 2x) - 2F_\Delta(x)}. \qquad (179) \end{aligned}$$

Particularized to the uniform case $F_\Delta(x) = 2x$, $0 \le x \le \frac{1}{2}$, the right side of (179) is equal to $\frac{x}{1-4x}$. The left side of (179) is monotonically increasing with $W(x)$ with a limit of $1/2$ as $W(x) \to 1/2$. Therefore, the maximum value of x for which we can find a valid solution is $1/6$, since at that point $\frac{x}{1-4x} = \frac{1}{2}$. Furthermore, the solution of (179) satisfies $W(x) \ge x$ for $x \ge 0.13605$. Differentiating (179), we obtain the differential equation

$$\dot{W}(x) = \frac{W(x)(1 - W(x))(1 - 2W(x))}{(1 - 4x)((W^2(x) - W(x))(1 - 2x) + x)} \qquad (180)$$

with initial condition $W(0.13605) = 0.13605$. The integrand in (171) is positive for all $0.09228 \le x \le 1/6$. Therefore, we let $\delta_{\min} = 0.13605$, and $\delta_{\max} \le 1/6$. Optimizing (171), we obtain $\delta_{\max} = 1/6$ and

$$\langle \mathsf{C} \rangle = 0.118 \qquad (181)$$

achieved with a "last" layer which sends no information since $W(1/6) = 1/2$. The lower bound in (173) is achieved with a simple fixed-length coding scheme tuned to a BSC with crossover probability 0.1545, yielding an average rate equal to 0.117. Therefore, in practice, the additional complexity required to obtain (181) would hardly be justified.

Example 6: Z-channel. In this example, we consider a Z-channel where 1 is received error-free and 0 is received as 1 with probability $0 \le A \le 1$. Conditioned on A the channel is memoryless, and A has cumulative distribution function F_A. If A were known at the transmitter, the variable-to-fixed capacity

and the fixed-to-variable capacity would equal the expected value of the capacity [36]

$$\mathbb{E}[C(A)] = \mathbb{E}\left[\log\left(1 - A^{\frac{1}{1-A}} + A^{\frac{A}{1-A}}\right)\right] \tag{182}$$

which is equal to 0.368 if A is uniformly distributed on $[0,1]$.

Denote the mutual information achieved by the input distribution $\mathbb{P}[X=1]=q$ when $A=\alpha$ by

$$\mathsf{I}(q,\alpha) = I(X;Y \mid A=\alpha) \tag{183}$$
$$= h((1-q)(1-\alpha)) - (1-q)h(\alpha) \tag{184}$$

Following the lower bound on $[\mathsf{C}]$ given in Theorem 13, we obtain

$$[\mathsf{C}] \geq \left(\min_q \int_0^1 \frac{dF_A(\alpha)}{\mathsf{I}(q,\alpha)}\right)^{-1} \tag{185}$$

while Theorem 15 gives

$$[\bar{\mathsf{C}}] = \max_q \int_0^1 \mathsf{I}(q,\alpha) dF_A(\alpha). \tag{186}$$

If A is uniformly distributed between 0 and 1, we get $C=0$ and

$$0.125 \leq [\mathsf{C}] \leq 0.366 = [\bar{\mathsf{C}}]. \tag{187}$$

In order to find $\langle \mathsf{C} \rangle$, we note that the Z-channel with crossover probability $\alpha_2 > \alpha_1$ is the cascade of two Z-channels with crossover probabilities α_1 and $\frac{\alpha_2-\alpha_1}{1-\alpha_1}$ (see also [37]). According to Theorem 6, we need to analyze the average rate achieved in a degraded broadcast Z-channel. For ease of exposition, we analyze first a suboptimal strategy where encoder and decoder postulate K different values $\alpha_1 < \cdots < \alpha_K$ for the channel crossover probability. The maximum average variable-to-fixed rate achieved by such a scheme is [via (70)]

$$R(\alpha_1,\ldots,\alpha_K)$$
$$= \max_{q,\beta_2,\ldots,\beta_K} F_A(\alpha_K) I(U_K;Y_K) + \sum_{k=1}^{K-1} F_A(\alpha_k) I(U_k;Y_k \mid U_{k+1}) \tag{188}$$
$$= \max_{q,\beta_2,\ldots,\beta_K} F_A(\alpha_K) I(U_K;Y_K) + \sum_{k=1}^{K-1} F_A(\alpha_k)\left(H(Y_k \mid U_{k+1}) - H(Y_k \mid U_k)\right) \tag{189}$$
$$= \max_{q,\beta_2,\ldots,\beta_K} F_A(\alpha_K)\big(h((1-q)\upsilon_K(1-\alpha_K)) - (1-q)h(\upsilon_K(1-\alpha_K))\big) + \upsilon_K(1-q)\sum_{k=1}^{K-1} F_A(\alpha_k)\left(\frac{h(\upsilon_{k+1}(1-\alpha_k))}{\upsilon_{k+1}} - \frac{h(\upsilon_k(1-\alpha_k))}{\upsilon_k}\right) \tag{190}$$

where Y_k is the response to U_1 of a Z-channel with crossover probability α_k, U_{k-1} is the response to U_k of a Z-channel with crossover probability β_k, $\mathbb{P}[U_K=1]=q$, $\upsilon_1=1$ and if $k>1$, then

$$\upsilon_k = \prod_{i=2}^{k}(1-\beta_k). \tag{191}$$

Then, for each value of K, one can proceed to optimize the $2K$ parameters $\alpha_1,\ldots\alpha_K,\beta_2,\ldots,\beta_K,q$. A more general approach is to set up the continuous optimization version of (190), where a monotone nonincreasing function $0 \leq V(\alpha) \leq 1$ plays the role that (191) plays with α_k. Letting $\alpha_1=\alpha_{\min}$ and $\alpha_K=\alpha_{\max}$ be parameters to be optimized, the summation in (190) becomes

$$\int_{\alpha_{\min}}^{\alpha_{\max}} F_A(\alpha)\frac{\dot{V}(\alpha)}{V^2(\alpha)}\big((1-\alpha)V(\alpha)\dot{h}((1-\alpha)V(\alpha)) - h((1-\alpha)V(\alpha))\big)\,d\alpha$$
$$= \int_{\alpha_{\min}}^{\alpha_{\max}} F_A(\alpha)\frac{\dot{V}(\alpha)}{V^2(\alpha)}\left((1-\alpha)V(\alpha) \times \log_2\left(\frac{1-(1-\alpha)V(\alpha)}{(1-\alpha)V(\alpha)}\right) - h((1-\alpha)V(\alpha))\right)d\alpha \tag{192}$$
$$= \int_{\alpha_{\min}}^{\alpha_{\max}} F_A(\alpha)\frac{\dot{V}(\alpha)}{V^2(\alpha)}\log_2\left(1-(1-\alpha)V(\alpha)\right)d\alpha. \tag{193}$$

In order to optimize the functional in (193) with respect to the function $\{V(\alpha), \alpha_{\min} \leq \alpha \leq \alpha_{\max}\}$ we again appeal to the Euler–Lagrange formalism [17] which requires the optimum $V(\cdot)$ to satisfy

$$\frac{\partial \mathcal{G}}{\partial y}(\alpha, V(\alpha), \dot{V}(\alpha)) = \frac{d}{d\alpha}\frac{\partial \mathcal{G}}{\partial z}(\alpha, V(\alpha), \dot{V}(\alpha)) \tag{194}$$

where

$$\mathcal{G}(\alpha,y,z) = F_A(\alpha)\frac{z}{y^2}\log_2(1-(1-\alpha)y). \tag{195}$$

Letting the density function of A be f_A, we have

$$\frac{\partial \mathcal{G}}{\partial z}(\alpha,y,z) = F_A(\alpha)\frac{1}{y^2}\log_2(1-(1-\alpha)y) \tag{196}$$

$$\frac{\partial \mathcal{G}}{\partial y}(\alpha,y,z) = F_A(\alpha)\frac{z}{y^4}\left(\frac{y^2(\alpha-1)}{1-(1-\alpha)y}\log_2 e - 2y\log_2(1-(1-\alpha)y)\right) \tag{197}$$

$$\frac{d}{d\alpha}F_A(\alpha)\frac{1}{V^2(\alpha)}\log_2(1-(1-\alpha)V(\alpha)) = f_A(\alpha)\frac{1}{V^2(\alpha)}\log_2(1-(1-\alpha)V(\alpha)) - F_A(\alpha)\frac{2V(\alpha)\dot{V}(\alpha)}{V^4(\alpha)}\log_2(1-(1-\alpha)V(\alpha)) + F_A(\alpha)\frac{1}{V^2(\alpha)}\frac{V(\alpha)+(\alpha-1)\dot{V}(\alpha)}{1-(1-\alpha)V(\alpha)}\log_2 e. \tag{198}$$

Then, the Euler–Lagrange condition (194) becomes

$$f_A(\alpha)(1-(1-\alpha)V(\alpha))\log_e(1-(1-\alpha)V(\alpha)) + F_A(\alpha)V(\alpha) = 0 \quad (199)$$

Taking the derivative of (199) leads to an ordinary differential equation

$$\dot{V}(\alpha) = V(\alpha)\frac{2-(2-\alpha)V(\alpha)}{1-2\alpha-(1-\alpha)^2V(\alpha)} \quad (200)$$

where we have particularized the solution to A uniformly distributed on $[0,1]$, i.e., $f_A(\alpha)=1$, $F_A(\alpha)=\alpha$. According to (199), $V(\alpha)=0$ is the only possible solution for $\alpha \geq 1/2$. Furthermore, $V(e^{-1})=1$. Therefore, the optimal $\alpha_{\min}=e^{-1}$ since smaller values would lead to $V(\alpha)>1$ for $\alpha<e^{-1}$ and larger values can only lead to a lower value of (193). Then, using (190), $\langle \mathsf{C} \rangle$ is obtained by optimizing

$$\begin{aligned}\langle \mathsf{C} \rangle = \max_{\alpha_{\max},q} \; & \alpha_{\max} h((1-q)V(\alpha_{\max})(1-\alpha_{\max})) \\ & - \alpha_{\max}(1-q)h(V(\alpha_{\max})(1-\alpha_{\max})) \\ & + (1-q)V(\alpha_{\max})\mathcal{I}(e^{-1},\alpha_{\max}) \end{aligned} \quad (201)$$

where $\mathcal{I}(\alpha_{\min},\alpha_{\max})$ stands for (193) evaluated with the optimal V obtained as the solution of the differential equation (200), or equivalently, (199). The optimal value in (201) is

$$\langle \mathsf{C} \rangle = 0.165 \quad (202)$$

and is attained at $q \to 0$ and $\alpha_{\max}=0.47$. So, again, the "last" layer is silent. As a point of comparison, a fixed-rate code designed for a Z-channel with crossover probability 0.43543 achieves average rate 0.1635 when the crossover probability is, in fact, uniformly distributed between 0 and 1.

Example 7: Symmetric Z-channel (known crossover probability). In this case, we have a two-state binary-input binary-output channel (42), where

$$W_0(0\,|\,0)=1 \quad (203)$$
$$W_0(0\,|\,1)=\delta \quad (204)$$
$$W_1(1\,|\,0)=\delta \quad (205)$$
$$W_1(1\,|\,1)=1 \quad (206)$$

and both states are equiprobable. It is easy to see that if error probability $\frac{1}{2} \leq \epsilon < 1$ is tolerated, the encoder/decoder can agree on pretending that state 0 is in effect and

$$C_\epsilon = \log\left(1-\delta^{\frac{1}{1-\delta}}+\delta^{\frac{\delta}{1-\delta}}\right) \quad (207)$$

Particularizing Theorems 12 and 15, we obtain

$$C = [\bar{\mathsf{C}}] = 1-\frac{1+\delta}{2}h\left(\frac{1}{1+\delta}\right). \quad (208)$$

Furthermore, the lower bound in the right side of (101) is also equal to the right side of (208); therefore, $[\mathsf{C}]=[\bar{\mathsf{C}}]$. Particularizing, Theorem 7 we obtain that $\langle \mathsf{C} \rangle$ is also equal to (208). The conclusion is that for this channel, variable-rate does not help in any of its guises even though the channel does not satisfy the strong converse.

Example 8: Symmetric Z-channel (unknown crossover probability). We now generalize Example 7, replacing the deterministic parameter δ by a random variable. Suppose we have $2n$ Z-channels where

$$W_{2i}(0\,|\,0)=1 \quad (209)$$
$$W_{2i}(0\,|\,1)=\delta_i \quad (210)$$
$$W_{2i+1}(1\,|\,0)=\delta_i \quad (211)$$
$$W_{2i+1}(1\,|\,1)=1 \quad (212)$$

for $i=1,\ldots,n$ where W_{2i} and W_{2i+1} have probability $p_i/2$ each and

$$\delta_1 \geq \delta_2 \geq \cdots \geq \delta_n. \quad (213)$$

Due to the concavity of mutual information, it is optimal to assign equiprobable input distributions. In principle, it is not necessarily optimal to assign a layer for each possible channel state. If a layer is associated with channel realization δ_i, we let $a_i=1$; otherwise $a_i=0$. Furthermore, denote the number of active layers up to and including state i

$$\mu_i = \sum_{k=1}^{i} a_k \leq i. \quad (214)$$

Let β_i stand for the effective crossover probability of a Z-channel with crossover δ_i and all undetected layers $i+1,\ldots,n$

$$(1-\beta_i) = (1-\delta_i)2^{\mu_i-\mu_n}. \quad (215)$$

The incremental rate decoded at level i is [37]

$$\Delta R_i = \begin{cases} \tau(\beta_i)2^{1-\mu_i}, & \mu_i=\mu_{i-1}+1 \\ 0, & \mu_i=\mu_{i-1} \end{cases} \quad (216)$$

where

$$\tau(\delta) = h\left(\frac{1-\delta}{2}\right)-\frac{1}{2}h(\delta) \quad (217)$$

with $h(\cdot)$ is the binary entropy function. The expected rate achieved with a given choice of active layers is

$$C(a_1,\ldots,a_n) = \sum_{k=1}^{n} p_k \sum_{i=1}^{k} \Delta R_i \quad (218)$$

Optimizing over the binary parameters $a_1,\ldots,a_n$, we obtain $\langle \mathsf{C} \rangle$.

Even if we have an uncountable number of crossover probabilities governed by a continuous distribution F_Δ, the number of active layers is finite as otherwise μ_i would become infinite and ΔR in (216) would be zero. Now, to obtain $\langle \mathsf{C} \rangle$, we maximize (218) with respect to both n and the collection of $\{\delta_i\}$ satisfying (213), with $(a_1,\ldots,a_n)=(1,\ldots,1)$ and

$$p_i = F_\Delta(\delta_i)-F_\Delta(\delta_{i-1}). \quad (219)$$

Example 9: Gaussian channel with nonergodic fading. Consider the complex-valued channel

$$\mathsf{y}_i = \mathsf{H}\,\mathsf{x}_i + \mathsf{n}_i \tag{220}$$

where $\{\mathsf{n}_i\}$ is a Gaussian memoryless random process with zero mean and $\mathbb{E}[|\mathsf{n}_i|^2] = 1$, $\mathsf{H} \in [H_{\min}, H_{\max}]$ is independent of $\{\mathsf{x}_i\}$ and $\{\mathsf{n}_i\}$, and has cumulative distribution function F_{H} such that $\inf\{x : F_{\mathsf{H}}(x) > 0\} = H_{\min}$ and $\inf\{x : F_{\mathsf{H}}(x) = 1\} = H_{\max}$. The encoder is constrained to satisfy a power constraint

$$\frac{1}{n}\sum_{i=1}^{n} |\mathsf{x}_i|^2 \le P. \tag{221}$$

It is straightforward to extend the foregoing results to channels with input constraints to show the following formulas:

$$C = \log(1 + H_{\min} P) \tag{222}$$

$$\lim_{\epsilon \uparrow 1} C_\epsilon = \log(1 + H_{\max} P) \tag{223}$$

$$[\![\mathsf{C}]\!] = \left(\mathbb{E}\left[\frac{1}{\log(1+\mathsf{H}P)}\right]\right)^{-1} \tag{224}$$

$$[\![\bar{\mathsf{C}}]\!] = \mathbb{E}[\log(1+\mathsf{H}P)] \tag{225}$$

$$\langle\!\langle \mathsf{C} \rangle\!\rangle = \int_{x_0}^{x_1} (1 - F_{\mathsf{H}}(x)) \left(\frac{\dot{f}_{\mathsf{H}}(x)}{f_{\mathsf{H}}(x)} + \frac{2}{x}\right) dx \log e \tag{226}$$

where $f_{\mathsf{H}}(x) = \dot{F}_{\mathsf{H}}(x)$ and the integration interval is defined by (see [26] for details)

$$P = \frac{1 - F_{\mathsf{H}}(x_0) - x_0 f_{\mathsf{H}}(x_0)}{x_0^2 f_{\mathsf{H}}(x_0)} \tag{227}$$

$$1 - F_{\mathsf{H}}(x_1) = x_1 f_{\mathsf{H}}(x_1). \tag{228}$$

Further elaboration of (226) can be found in the case of Rayleigh fading in [26]. In that case, $C = [\![\mathsf{C}]\!] = 0$, while $[\![\bar{\mathsf{C}}]\!]$ coincides with the capacity had the fading been ergodic and known to the receiver. A fading distribution for which $C = 0 < [\![\mathsf{C}]\!]$ is the χ^2-distribution rising from two-antenna diversity.

VII. Summary and Conclusions

We have defined several new variable-rate notions of channel capacity, and proved various expressions and relationships among them. The setup of fixed-to-variable channel coding is inspired by current fountain coding technology, while the setup of variable-to-fixed channel coding is reminiscent of unequal protection coding technology where the fidelity with which a source is decoded depends on channel conditions (e.g., [21], [33], and [34]). Motivated by [6], the literature has followed the "broadcast approach" in a pragmatic way for scalarly parametrized single-user channels that can be viewed as degraded versions of the same channel. In contrast, thanks to our definition of variable-to-fixed capacity, the relationship with broadcast channels with degraded message sets is essential, rather than ad-hoc.

An interesting feature of our variable-rate capacity definitions is that they do not involve the notion of probability of decoding the wrong codeword, unlike the conventional definition of channel capacity. Instead, the new notions rely on the expectation of random variables such as the number of consecutive bits correctly decoded and the number of observations required to select the correct message. A pleasing feature of the various notions of variable-rate channel capacity is that they equal the conventional channel capacity for channels that satisfy the strong converse.

Summarizing the relationships that hold in wide generality

$$[\![\mathsf{C}]\!] = C \tag{229}$$

$$\le \min\{[\mathsf{C}], \langle\mathsf{C}\rangle\} \tag{230}$$

$$\le \max\{[\mathsf{C}], \langle\mathsf{C}\rangle\} \tag{231}$$

$$\le [\bar{\mathsf{C}}] \tag{232}$$

$$\le \lim_{\epsilon \uparrow 1} C_\epsilon \tag{233}$$

while we have shown simple examples where $[\mathsf{C}] < \langle\mathsf{C}\rangle$, and others for which $\langle\mathsf{C}\rangle < [\mathsf{C}]$.

We have also found several formulas and bounds for variable-to-fixed and fixed-to-variable capacity for state-dependent channels. Those formulas enable the analysis of various Bayesian setups. For example (all values in bits):

- a BEC with erasure probability uniformly distributed on $[0,1]$ has $C = [\mathsf{C}] = 0$, $\langle\mathsf{C}\rangle = \frac{1}{4}$, $[\bar{\mathsf{C}}] = \frac{1}{2}$;
- a BSC with crossover probability uniformly distributed on $[0, \frac{1}{2}]$ has $C = [\mathsf{C}] = 0$, $\langle\mathsf{C}\rangle = 0.118$, $[\bar{\mathsf{C}}] = 0.639$;
- a Gaussian-noise channel with Rayleigh fading (constant throughout the codeword duration) and $\mathrm{SNR} = 0$ dB has $C = [\mathsf{C}] = 0$, $\langle\mathsf{C}\rangle = 0.385$, $[\bar{\mathsf{C}}] = 0.860$.

In those cases, $[\bar{\mathsf{C}}]$ is equal to the average capacity that would be obtained if the transmitter knew the channel (but in the case of the fading channel would not be allowed to control the output power). Those examples illustrate the shortcomings of the conventional fixed-rate setup in the absence of channel state information at the transmitter.

While it is fairly straightforward to incorporate cost constraints, other generalizations such as source-channel coding, feedback and multiuser channels are interesting and not necessarily easy.

Acknowledgment

Comments by Y. Polyanskiy are gratefully acknowledged.

References

[1] R. Ahlswede, "The weak capacity of averaged channels," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, vol. 11, pp. 61–73, 1968.

[2] R. Ahlswede and P. Gács, "Two contributions to information theory," in *Colloq. Mathematica Societatis János Bolyai: 16. Topics in Information Theory*, Keszthely, Hungary, 1975, pp. 17–40.

[3] S. Boucheron and M. R. Salamatian, "About priority encoding transmission," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 699–705, Mar. 2000.

[4] M. V. Burnashev, "Data transmission over discrete channel with feedback: Random transmission time," *Prob. Peredac. Inf.*, vol. 12, no. 4, pp. 10–30, 1976.

[5] P.-N. Chen and F. Alajaji, "Optimistic Shannon coding theorems for arbitrary single-user systems," *IEEE Trans. Inf. Theory*, vol. 45, no. 11, pp. 2623–2629, Nov. 1999.

[6] T. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.

[7] T. Cover and J. Thomas, *Elements of Information Theory*. Hoboken, NJ: Wiley-Interscience, 2006.

[8] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
[9] S. C. Draper, F. R. Kschischang, and B. Frey, "Rateless coding for arbitrary channel mixtures with decoder channel state information," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4119–4133, Sep. 2009.
[10] M. Effros and A. Goldsmith, "Capacity definitions and coding strategies for general channels with receiver side information," in *Proc. 1998 IEEE Int. Symp. Information Theory*, Cambridge, MA, Aug. 1998, p. 39.
[11] M. Effros, A. Goldsmith, and Y. Liang, "Capacity definitions of general channels with receiver side information," in *Proc. 2007 IEEE Int. Symp. Information Theory*, Nice, France, Jun. 24–29, 2007, pp. 2226–2230.
[12] T. S. Han, *Information-Spectrum Methods in Information Theory*. New York: Springer, 2003.
[13] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.
[14] K. Jacobs, "Almost periodic channels," in *Colloq. Combinatorial Methods in Probability Theory*, Aarhus, 1962, pp. 118–126.
[15] J. Körner and K. Marton, "General broadcast channels with degraded messages sets," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 1, pp. 60–64, Jan. 1977.
[16] M. G. Luby, "LT codes," in *Proc. 43rd IEEE Symp. Foundations of Computer Science*, 2002, pp. 271–280.
[17] D. Luenberger, *Optimization by Vector Space Methods*. Hoboken, NJ: Wiley, 1969.
[18] P. Minero and D. Tse, "A broadcast approach to multiple access with random states," in *Proc. 2007 IEEE Int. Symp. Information Theory*, Nice, France, Jun. 24–29, 2007, pp. 2566–2570, see also 2010 IEEE Information Theory Workshop, Taormina, Sicily, Oct. 11–16, 2009.
[19] S. Musy, "Variable length codes for degraded broadcast channels," in *Proc. 2007 IEEE Int. Symp. Information Theory*, Nice, France, Jun. 24–29, 2007, pp. 2576–2580.
[20] C. Nair and A. El Gamal, "Capacity region of a class of 3-receiver broadcast channels with degraded message sets," in *Proc. 2008 IEEE Symp. Information Theory*, Toronto, ON, Canada, Jul. 2008, pp. 1706–1710.
[21] C. Ng, D. Gündüz, A. Goldsmith, and E. Erkip, "Minimum expected distortion in gaussian layered broadcast coding with successive refinement," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5074–5086, Nov. 2009.
[22] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Variable-length coding with feedback in the non-asymptotic regime," in *Proc. 2010 IEEE Int. Symp. Information Theory*, Austin, TX, Jun. 2010, to appear.
[23] M. Salehi, "Cardinality Bounds on Auxiliary Variables in Multiple-User Theory via the Method of Ahlswede and Körner," Dept. Statistics, Stanford Univ., Stanford, CA, Tech. Rep. no. 33, Aug. 1978.
[24] S. Shamai (Shitz), E. Telatar, and S. Verdú, "Fountain capacity," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4372–4377, Nov. 2007.
[25] S. Shamai (Shitz), "A broadcast strategy for the gaussian slowly fading channel," in *Proc. 1997 IEEE Int. Symp. Information Theory*, Ulm, Germany, Jun. 29–Jul. 4, 1997, p. 150.
[26] S. Shamai (Shitz) and A. Steiner, "A broadcast approach for a single-user slowly fading MIMO channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2617–2635, Oct. 2003.
[27] S. Shamai (Shitz), "A broadcast approach for the multiple-access slow fading channel," in *Proc. 2000 IEEE Int. Symp. Information Theory*, Sorrento, Italy, Jun. 25–30, 2000, p. 128.
[28] O. Shayevitz and M. Feder, "Achieving the empirical capacity using feedback. Memoryless additive models," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1269–1295, Mar. 2009.
[29] N. Shulman and M. Feder, "Static broadcasting," in *Proc. 2000 IEEE Int. Symp. Information Theory*, Sorrento, Italy, Jun. 25–30, 2000, p. 23.
[30] V. Strassen, "Messfehler und information," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 2, no. 4, pp. 273–305, Jan. 1964.
[31] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 323–349, Jan. 2009.
[32] A. Tchamkerten and E. Telatar, "Variable length coding over an unknown channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2126–2145, May 2006.
[33] C. Tian, A. Steiner, S. Shamai (Shitz), and S. Diggavi, "Successive refinement via broadcast: Optimizing expected distortion of a Gaussian source over a Gaussian fading channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 2903–2918, Jul. 2008.
[34] M. D. Trott, "Unequal error protection codes: Theory and practice," in *Proc. 1996 IEEE Information Theory Workshop*, Haifa, Israel, Jun. 9–13, 1996, p. 11.
[35] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.
[36] S. Verdú, "Channel Capacity," in *The Electrical Engineering Handbook*. Boca Raton, FL: CRC, 1997, pp. 1671–1678.
[37] B. Xie, M. Griot, A. I. V. Casado, and R. D. Wesel, "Optimal transmission strategy and explicit capacity region for broadcast Z channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4296–4304, Sep. 2008.

Sergio Verdú (S'80–M'84–SM'88–F'93) received the Telecommunications Engineering degree from the Universitat Politècnica de Barcelona, Barcelona, Spain, in 1980 and the Ph.D. degree in Electrical Engineering from the University of Illinois at Urbana-Champaign, Urbana, in 1984.

Since 1984, he has been a member of the faculty of Princeton University, Princeton, NJ, where he is the Eugene Higgins Professor of Electrical Engineering.

Dr. Verdú is the recipient of the 2007 Claude E. Shannon Award and the 2008 IEEE Richard W. Hamming Medal. He is a member of the National Academy of Engineering and was awarded a Doctorate Honoris Causa from the Universitat Politècnica de Catalunya in 2005. He is a recipient of several paper awards from the IEEE: the 1992 Donald Fink Paper Award, the 1998 Information Theory Outstanding Paper Award, an Information Theory Golden Jubilee Paper Award, the 2002 Leonard Abraham Prize Award, the 2006 Joint Communications/ Information Theory Paper Award, and the 2009 Stephen O. Rice Prize from IEEE Communications Society. He has also received paper awards from the Japanese Telecommunications Advancement Foundation and from Eurasip. He received the 2000 Frederick E. Terman Award from the American Society for Engineering Education for his book *Multiuser Detection* (Cambridge, U.K.: Cambridge Univ. Press, 1998). He served as President of the IEEE Information Theory Society in 1997. He is currently Editor-in-Chief of *Foundations and Trends in Communications and Information Theory*.

Shlomo Shamai (Shitz) (S'80–M'82–SM'88–F'94) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 1975, 1981, and 1986, respectively.

From 1975–1985, he was with the Communications Research Labs in the capacity of a Senior Research Engineer. Since 1986, he has been with the Department of Electrical Engineering, Technion, where he is now the William Fondiller Professor of Telecommunications. His research interests encompass a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai (Shitz) is a member of the Union Radio Scientifique Internationale (URSI). He is the recipient of the 1999 van der Pol Gold Medal of URSI and a corecipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003 and the 2004 IEEE Communications Society and Information Theory Society Joint Paper Award, and the 2007 IEEE Information Theory Society Paper Award. He is also the recipient of the 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY and also served on the Board of Governors of the IEEE Information Theory Society.