On Error Exponents for Arbitrarily Varying Channels

Brian L. Hughes, *Member, IEEE*, and Tony G. Thomas, *Member, IEEE*

Abstract— **The minimum probability of error achievable by random codes on the arbitrarily varying channel (AVC) is investigated. New exponential error bounds are found and applied to the AVC with and without input and state constraints. Also considered is a simple subclass of random codes, called *randomly modulated codes*, in which encoding and decoding operations are separate from code randomization. A universal coding theorem is proved which shows the existence of randomly modulated codes that achieve the same error bounds as "fully" random codes for all AVC's.**

Index Terms— **Arbitrarily varying channels, error exponents, random codes, jamming.**

I. Introduction

THE *arbitrarily varying channel* (AVC) models a communication channel with an unknown state that varies with time in an arbitrary way from one symbol transmission to the next. The practical significance of the model lies mainly in its relevance to the problem of communication in the presence of jamming.

The capacity of the AVC depends on whether deterministic codes or random codes are used. The deterministic code capacity further depends on whether the average or the maximum of the error probability over all codewords is used as the measure of performance. For many AVC's of practical interest, random codes can achieve a much larger capacity and reliability function than deterministic codes. Indeed, when the channel is *symmetrizable* [6], a positive rate of transmission is possible only if random coding is used. Consequently, random codes are important not only as tools for proving coding theorems, but also as models for practical communication systems.

A central problem in the information-theoretic study of the AVC is to determine the minimum error probability achievable with random block codes. The first results on this problem are due to Blackwell, Breiman, and Thomasian [3] who, in their pioneering paper introducing the AVC, derived the capacity for random codes. Stiglitz [9] presented a simplified proof of their result which also established an exponential error bound. Ahlswede [1] showed that capacity can be achieved by random codes with an ensemble of size at most n^2, where n is the blocklength of the code. Ericson [7] generalized Stiglitz's error bound to reflect dependence on a quantity he termed the *key rate*, a measure of the size of the random code ensemble. More recently, Csiszár and Narayan [5] extended the coding theorem of Blackwell *et al.* to the situation where the channel input and state symbols are subject to cost constraints.

This paper revisits the problem of exponential error bounds for the minimum error probability achievable on the AVC with random codes. Our main contributions comprise two parts. First, we derive AVC analogs of the random coding, sphere packing, and expurgated bounds for random codes and state sequences of fixed composition. We then use these results to obtain a stronger form of Stiglitz's bound, and to refine the coding theorem of Csiszár and Narayan for the AVC subject to constraints.

Coding theorems for AVC's are typically proved using random codes in which codewords are independent and identically distributed. Because of the obvious practical disadvantages of such schemes, it is of interest to determine whether the same performance can be achieved by less complex codes. One approach to this goal, pursued in [1] and [7], is to look for random codes with a small ensemble size. In this paper, we explore an alternate approach. We consider a simple subclass of random codes in which coding and randomization are separate. More specifically, a *randomly modulated* code consists of a deterministic code combined with a random permutation mapping. Our second contribution is a universal coding theorem which demonstrates the existence of randomly modulated codes that achieve the same error bounds as "fully" random codes for every AVC and for every state sequence.

The rest of the paper is organized as follows. Section II introduces terminology and summarizes our main results. These results are proved in Section III. A simple example is given and conclusions are summarized in Sections IV and V, respectively.

II. Summary of Results

A. Preliminaries

The notation used in this paper is adapted from [5]. Let $\mathcal{X}$, $\mathcal{S}$, and $\mathcal{Y}$ be finite sets. A discrete memoryless channel (DMC) is defined by a stochastic matrix $W\colon \mathcal{X} \to \mathcal{Y}$. For n channel uses, the transition probability is

$$W^n(\boldsymbol{y} \mid \boldsymbol{x}) \triangleq \prod_{i=1}^{n} W(y_i \mid x_i)$$

where $\boldsymbol{x} = (x_1, \cdots, x_n) \in \mathcal{X}^n$ and $\boldsymbol{y} = (y_1, \cdots, y_n) \in \mathcal{Y}^n$.

A (discrete memoryless) *arbitrarily varying channel* (AVC), $\mathcal{W} \triangleq \{W(\cdot \mid \cdot, s)\colon s \in \mathcal{S}\}$, is a collection of channels $W\colon \mathcal{X} \to \mathcal{Y}$ indexed by a parameter s called the channel *state*.

Manuscript received September 8, 1994; revised July 15, 1995. This work was supported by the National Science Foundation under Grant NCR-9217457 and by the US Army Research Office and the US Army Research Laboratory under Grant DAAL03-89-K-0130. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, San Diego, CA, January 14–19, 1990.

B. L. Hughes is with the Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD 21218 USA.

T. G. Thomas is with AT&T Bell Laboratories, Holmdel, NJ 07733 USA.

Publisher Item Identifier S 0018-9448(96)00018-1.

We interpret $W(y \mid x, s)$ as the conditional probability that the channel output is $y \in \mathcal{Y}$ when the channel input is $x \in \mathcal{X}$ and the channel state is $s \in \mathcal{S}$. The channel operation on n-tuples $\boldsymbol{x} \in \mathcal{X}^n$, $\boldsymbol{y} \in \mathcal{Y}^n$, $\boldsymbol{s} = (s_1, \cdots, s_n) \in \mathcal{S}^n$, is given by

$$W^n(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{s}) \triangleq \prod_{i=1}^{n} W(y_i \mid x_i, s_i).$$

An (n, M) *code* is a pair (f, φ) consisting of an *encoder* f: $\mathcal{M} \to \mathcal{X}^n$ and a *decoder* φ: $\mathcal{Y}^n \to \mathcal{M}$, where $\mathcal{M} \triangleq \{1, \cdots, M\}$ is the message set. The *rate* of this code is $R \triangleq (1/n) \log M$. (Throughout this paper, all exponents, logarithms, and information measures are to the base 2.) For any channel W_n: $\mathcal{X}^n \to \mathcal{Y}^n$, the probability of error of (f, φ) when message $m \in \mathcal{M}$ is sent is

$$e_m(W_n, f, \varphi) \triangleq 1 - W_n(\varphi^{-1}(m) \mid f(m)). \tag{1}$$

In the particular case $W_n(\cdot \mid \cdot) = W^n(\cdot \mid \cdot, \boldsymbol{s})$, we display the dependence on $\boldsymbol{s}$ of the error probabilities by writing

$$e_m(\boldsymbol{s}, f, \varphi) \triangleq e_m(W^n(\cdot \mid \cdot, \boldsymbol{s}), f, \varphi).$$

An (n, M) *random code*, (F, Φ), is a random variable (RV) that takes values in the set of all (n, M) codes defined on the same alphabets $\mathcal{X}$ and $\mathcal{Y}$.

Following [5], we impose cost constraints on the encoder and the channel state sequences. Let g and l be nonnegative-valued functions defined on $\mathcal{X}$ and $\mathcal{S}$, respectively. For all $\boldsymbol{x} = (x_1, \cdots, x_n) \in \mathcal{X}^n$ and $\boldsymbol{s} = (s_1, \cdots, s_n) \in \mathcal{S}^n$, let

$$g(\boldsymbol{x}) \triangleq \frac{1}{n} \sum_{i=1}^{n} g(x_i)$$

$$l(\boldsymbol{s}) \triangleq \frac{1}{n} \sum_{i=1}^{n} l(s_i).$$

The random code (F, Φ) is said to satisfy the *input constraint* Γ if for all $m \in \mathcal{M}$

$$g(F(m)) \leq \Gamma \quad \text{almost surely (a.s.)} \tag{2}$$

Similarly, a random state sequence $\boldsymbol{S} = (S_1, \cdots, S_n)$ satisfies the *channel constraint* Λ if

$$l(\boldsymbol{S}) \leq \Lambda \quad \text{a.s.} \tag{3}$$

For simplicity, we assume $\min_x g(x) = \min_s l(s) = 0$, $\Gamma > 0$, and $\Lambda > 0$.

We will need several definitions from the method of types [4, pp. 29–39]. For any finite sets $\mathcal{X}$ and $\mathcal{Y}$, $\mathcal{D}(\mathcal{X})$ denotes the set of all probability distributions on $\mathcal{X}$. The *type* of a sequence $\boldsymbol{x} \in \mathcal{X}^n$ is the probability distribution $P_{\boldsymbol{x}} \in \mathcal{D}(\mathcal{X})$ given by $P_{\boldsymbol{x}}(x) \triangleq N(x \mid \boldsymbol{x})/n$ for $x \in \mathcal{X}$, where $N(x \mid \boldsymbol{x})$ is the number of occurrences of x in $\boldsymbol{x}$. The set of all types of sequences in $\mathcal{X}^n$ is denoted by $\mathcal{D}_n(\mathcal{X})$, and the set of all n-tuples $\boldsymbol{x} \in \mathcal{X}^n$ of type P is denoted by $\mathcal{T}_P^n$, or simply $\mathcal{T}_P$ when n is understood. Similarly, the *joint type* of a pair $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ is the probability distribution $P_{\boldsymbol{xy}} \in \mathcal{D}(\mathcal{X} \times \mathcal{Y})$ given by

$$P_{\boldsymbol{xy}}(x, y) \triangleq N(x, y \mid \boldsymbol{x}, \boldsymbol{y})/n$$

and the *conditional type of* $\boldsymbol{y}$ *given* $\boldsymbol{x}$ is defined by

$$P_{\boldsymbol{y}|\boldsymbol{x}}(y \mid x) \triangleq P_{\boldsymbol{xy}}(x, y)/P_{\boldsymbol{x}}(x)$$

for $(x, y) \in \mathcal{X} \times \mathcal{Y}$. For $\boldsymbol{x} \in \mathcal{X}^n$ and V: $\mathcal{X} \to \mathcal{Y}$, let $\mathcal{T}_V^n(\boldsymbol{x})$ be the set of all $\boldsymbol{y} \in \mathcal{Y}^n$ of conditional type V given $\boldsymbol{x}$.

For any V: $\mathcal{X} \to \mathcal{Y}$ and $P \in \mathcal{D}(\mathcal{X})$, $I(P, V)$ denotes the *mutual information* between RV's X and Y with joint distribution $P(x)V(y \mid x)$. The mutual information between the sequences $\boldsymbol{x} \in \mathcal{X}^n$ and $\boldsymbol{y} \in \mathcal{Y}^n$ is defined by

$$I(\boldsymbol{x} \wedge \boldsymbol{y}) \triangleq I(P_{\boldsymbol{x}}, P_{\boldsymbol{y}|\boldsymbol{x}}).$$

Given a code (f, φ), we say that φ is a *maximum mutual information* (MMI) decoder for f if for all $m \in \mathcal{M}$ and $\boldsymbol{y} \in \mathcal{Y}^n$

$$\varphi(\boldsymbol{y}) = m \Longrightarrow I(f(m) \wedge \boldsymbol{y}) = \max_{m' \in \mathcal{M}} I(f(m') \wedge \boldsymbol{y}).$$

Similarly, given a random code (F, Φ), we say that Φ is an MMI decoder for F if, for all $\boldsymbol{y} \in \mathcal{Y}^n$ and $m \in \mathcal{M}$, the above condition holds almost surely on the ensemble of (F, Φ).

B. Error Exponents

The coding problem associated with the AVC subject to constraints is to construct random codes satisfying the input constraint Γ such that the *maximum error probability*

$$e(\boldsymbol{s}, F, \Phi) \triangleq \max_{m \in \mathcal{M}} \boldsymbol{E} e_m(\boldsymbol{s}, F, \Phi) \tag{4}$$

is uniformly small for all $\boldsymbol{s}$ satisfying the channel constraint Λ. The main goal of the information-theoretic study of this channel is to determine the most favorable possible relationship among the error probability, blocklength, and code rate of a random code. A first step toward this goal was taken by Csiszár and Narayan [5], who showed that the (random code) capacity is

$$\begin{aligned} C &= \min_{Q \in \mathcal{Q}_\Lambda} \max_{P \in \mathcal{P}_\Gamma} I(P, QW) \\ &= \max_{P \in \mathcal{P}_\Gamma} \min_{Q \in \mathcal{Q}_\Lambda} I(P, QW) \end{aligned} \tag{5}$$

where

$$QW(\cdot \mid \cdot) \triangleq \sum_s Q(s) W(\cdot \mid \cdot, s)$$

$$\mathcal{Q}_\Lambda \triangleq \{Q \in \mathcal{D}(\mathcal{S}) : \sum_s Q(s) l(s) \leq \Lambda\}$$

and where

$$\mathcal{P}_\Gamma \triangleq \{P \in \mathcal{D}(\mathcal{X}) : \sum_s P(s) g(s) \leq \Gamma\}.$$

This paper investigates the error exponents of the AVC subject to constraints, which give bounds on the exponential rate of decrease of the error probability with respect to the blocklength n, as a function of the rate R of the random code.

Definition 1: A number $E \geq 0$ is an *achievable error exponent at rate* R of $\mathcal{W}$ under random coding if, for every $\delta > 0$ and all sufficiently large n, there exists an (n, M) random code (F, Φ) satisfying the input constraint Γ such that $M \geq \exp\{n(R-\delta)\}$ and

$$\max_{\boldsymbol{s}:\ l(\boldsymbol{s}) \leq \Lambda} e(\boldsymbol{s}, F, \Phi) \leq \exp\{-n(E-\delta)\}.$$

The largest achievable error exponent at rate R of $\mathcal{W}$, considered as a function of R, is called the (random code) reliability function of $\mathcal{W}$ and is denoted by $E(R)$.

In general, we must be content to merely bound $E(R)$, since an explicit characterization of the reliability function is lacking even in the special case of the DMC. Our results rely on the methods used in [4, pp. 161–174] to bound the reliability function of the DMC. In particular, we find it useful to begin by restricting attention to codes and channel state sequences of fixed composition. To this end, we say that a random code (F, Φ) is *of constant type* P if $F(m) \in \mathcal{T}_P$, almost surely, for all $m \in \mathcal{M}$.

Theorem 1 (Random Coding Bound): For all $R > 0$, $\delta > 0$, $M \triangleq \lceil \exp\{n(R-\delta)\} \rceil$, and $P \in \mathcal{P}_n(\mathcal{X})$, let (F, Φ) be an (n, M) random code such that the RV's $F(m), m \in \mathcal{M}$, are independent and uniformly distributed on $\mathcal{T}_P$, and Φ is an MMI decoder for F. Then for all $Q \in \mathcal{P}_n(\mathcal{S})$

$$\max_{\boldsymbol{s} \in \mathcal{T}_Q} e(\boldsymbol{s}, F, \Phi) \leq \exp\{-n[E_r(R, W, P, Q) - \delta]\}$$

whenever $n \geq n_1(|\mathcal{Y}|, |\mathcal{X}|, |\mathcal{S}|, \delta)$, where

$$E_r(R, W, P, Q) \triangleq \min_{U_{YXS}:\ U_X = P, U_S = Q} D(U_{YXS} \,\|\, W \times P \times Q) + |I(P, U_{Y|X}) - R|^+. \tag{6}$$

Here

$$(W \times P \times Q)(y, x, s) \triangleq W(y \mid x, s) P(x) Q(s);$$

D is the divergence (e.g. [4, p. 20]); $|r|^+ \triangleq \max\{r, 0\}$; U_{YXS} denotes a probability distribution on $\mathcal{Y} \times \mathcal{X} \times \mathcal{S}$; and $U_{Y|X}$, U_X, and U_S are the obvious conditional and marginal distributions induced by U_{YXS}.

Theorem 2 (Sphere Packing Bound): For all $R > 0$, $\delta > 0$, $M \geq \exp\{n(R+\delta)\}$, $Q \in \mathcal{P}_n(\mathcal{S})$, and $P \in \mathcal{P}_n(\mathcal{X})$, every (n, M) random code (F, Φ) of constant type P satisfies

$$\max_{\boldsymbol{s} \in \mathcal{T}_Q} e(\boldsymbol{s}, F, \Phi) \geq \max_{m \in \mathcal{M}} \frac{1}{|\mathcal{T}_Q|} \sum_{\boldsymbol{s} \in \mathcal{T}_Q} e_m(\boldsymbol{s}, F, \Phi) \geq \exp\{-n[E_{sp}(R, W, P, Q) + \delta]\} \tag{7}$$

for all $n \geq n_2(|\mathcal{Y}|, |\mathcal{X}|, |\mathcal{S}|, \delta, \gamma)$, where γ is the smallest nonzero value of W and

$$E_{sp}(R, W, P, Q) \triangleq \min_{\substack{U_{YXS}:\ U_X = P, \\ U_S = Q, I(P, U_{Y|X}) \leq R}} D(U_{YXS} \,\|\, W \times P \times Q). \tag{8}$$

Remarks: Proceeding as in [4, p. 168], we can show the above exponents enjoy properties similar to their DMC counterparts. In particular, for fixed W, P, and Q, $E_{sp}(R, W, P, Q)$ is convex and decreasing in $R \geq 0$, positive if and only if $R < I(P, QW)$, and

$$E_r(R, W, P, Q) = \begin{cases} E_{sp}(\hat{R}, W, P, Q) + \hat{R} - R, & 0 \leq R \leq \hat{R} \\ E_{sp}(R, W, P, Q), & \hat{R} \leq R \leq C \end{cases}$$

where $\hat{R} \triangleq \hat{R}(W, P, Q)$ is the smallest $R \geq 0$ at which $E_{sp}(R, W, P, Q)$ meets its supporting line of slope -1. Note that $E_{sp}(R, W, P, Q)$ can be infinite; we denote by $R_\infty(W, P, Q)$ the infimum of $R \geq 0$ such that $E_{sp}(R, W, P, Q)$ is finite.

Returning now to the coding problem for the AVC subject to constraints, we observe that a random code of constant type P satisfies the input constraint Γ if and only if $P \in \mathcal{P}_\Gamma$, and $\boldsymbol{s} \in \mathcal{T}_Q$ satisfies the channel constraint Λ if and only if $Q \in \mathcal{Q}_\Lambda$. Recalling Definition 1, we see that Theorem 1 implies a lower bound for $E(R)$, and Theorem 2 implies an upper bound. These bounds are summarized in the following theorem.

Theorem 3: For $R > 0$, the reliability function of the AVC $\mathcal{W}$ subject to constraints satisfies

$$E_r(R) \leq E(R) \leq E_{sp}(R)$$

except possibly at

$$R = R_\infty(W) \triangleq \max_{P \in \mathcal{P}_\Gamma} \min_{Q \in \mathcal{Q}_\Lambda} R_\infty(W, P, Q)$$

where the upper bound need not hold. Here

$$E_r(R) \triangleq \max_{P \in \mathcal{P}_\Gamma} \min_{Q \in \mathcal{Q}_\Lambda} E_r(R, W, P, Q) \tag{9}$$

$$E_{sp}(R) \triangleq \max_{P \in \mathcal{P}_\Gamma} \min_{Q \in \mathcal{Q}_\Lambda} E_{sp}(R, W, P, Q). \tag{10}$$

Remarks: i) As above, $E_{sp}(R)$ is convex and decreasing, positive if and only if $R < C$, and

$$E_r(R) = \begin{cases} E_{sp}(\hat{R}) + \hat{R} - R, & 0 \leq R \leq \hat{R} \\ E_{sp}(R), & \hat{R} \leq R \leq C \end{cases}$$

where $\hat{R}$ is the smallest $R \geq 0$ at which $E_{sp}(R)$ meets its supporting line of slope -1. ii) When the constraints in (9) are absent or inactive (e.g., $g = l \triangleq 0$), it can be shown as in [4, p. 192] that $E_r(R)$ reduces to Stiglitz's exponent [9]. However, Stiglitz's proof used a different decoder for every R and every AVC $\mathcal{W}$, whereas it is evident from the proof of Theorem 3 that MMI decoding suffices to achieve this exponent for all R and $\mathcal{W}$.

It is interesting to compare the above exponents with the capacity (5). Observe that C equals the mutual information optimized over all joint distributions $P(x)Q(s)$ satisfying the constraints. By contrast, the programs in (6) and (8) do not require $U_{XS}(x, s) = P(x)Q(s)$. Indeed, as will become apparent from the proof of Theorem 4, the choice of U_{YXS} achieving the minimum in (6) and (7) will not generally be such that U_{XS} is of product form.

From (5), the capacity of the AVC can be interpreted as the minimum capacity of the DMC QW for all $Q \in \mathcal{Q}_\Lambda$. It is natural to ask whether the error exponents of the AVC admit

a similar interpretation. In particular, what is the relationship between $E_r(R, W, P, Q)$ and $\bar{E}_r(R, QW, P)$, where

$$\bar{E}_r(R, \hat{W}, P) \triangleq \min_{V:\ \mathcal{X}\to\mathcal{Y}} D(V \,\|\, \hat{W} \mid P) + |I(P,V) - R|^+ \tag{11}$$

is the random coding exponent of the DMC $\hat{W}\colon \mathcal{X} \to \mathcal{Y}$ for codes of constant type P [4, p. 165]?

Theorem 4: For all R, W, P, and Q

$$E_r(R, W, P, Q) \geq \bar{E}_r(R, QW, P)$$

with equality if and only if

$$\bar{E}_r(R, QW, P) = \min_{Q'\in\mathcal{D}(\mathcal{S}):\ Q'\ll Q} \bar{E}_r(R, Q'W, P). \qquad \square$$

One consequence of Theorem 4 is that

$$\min_{Q\in\mathcal{Q}_\Lambda} E_r(R, W, P, Q) > \min_{Q\in\mathcal{Q}_\Lambda} \bar{E}_r(R, QW, P)$$

unless there exists a $Q \in \mathcal{Q}_\Lambda$ that achieves the minimum on the left which also achieves

$$\min_{Q'\in\mathcal{D}(\mathcal{S}):\ Q'\ll Q} \bar{E}_r(R, Q'W, P).$$

Roughly speaking, this occurs when the channel constraint is inactive on the support of Q. In particular, this occurs when

$$\min_{Q\in\mathcal{Q}_\Lambda} \bar{E}_r(R, QW, P) = \min_{Q\in\mathcal{D}(\mathcal{S})} \bar{E}_r(R, QW, P).$$

However, except in these rare circumstances, the random coding exponent of the AVC is strictly larger than the corresponding exponent of the DMC QW for every $Q \in \mathcal{Q}_\Lambda$. It follows that there exist codes (f, φ) such that

$$\max_{Q\in\mathcal{Q}_\Lambda} e((QW)^n, f, \varphi)$$

is substantially larger than

$$\max_{\boldsymbol{s}:\ l(\boldsymbol{s})\leq\Lambda} e(\boldsymbol{s}, f, \varphi).$$

To see why this can occur, observe that

$$e_m((QW)^n, f, \varphi) = \sum_{\boldsymbol{s}\in\mathcal{S}^n} Q^n(\boldsymbol{s}) e_m(\boldsymbol{s}, f, \varphi).$$

When this sum is dominated by terms $\boldsymbol{s}$ satisfying $l(\boldsymbol{s}) > \Lambda$

$$\max_{\boldsymbol{s}:\ l(\boldsymbol{s})\leq\Lambda} e(\boldsymbol{s}, f, \varphi)$$

can be much smaller because such terms are not permitted by the channel constraint Λ.

C. A Restricted Class of Codes

Theorems 1 and 3 imply that $E_r(R)$ can be achieved by random codes in which codewords are chosen independently and uniformly on $\mathcal{T}_P$ for some $P \in \mathcal{P}_\Gamma$. Because of the obvious practical disadvantages of such codes, it is of interest to determine whether similar performance can be achieved with less complex codes. In this paper, we consider random codes in which encoding and decoding operations are separate from code randomization. Specifically, we define a *randomly modulated* (RM) code to be a random code (F, Φ) of the form

$$F(m) = \boldsymbol{T} f(m), m \in \mathcal{M}, \Phi(\boldsymbol{y}) = \varphi(\boldsymbol{T}^{-1}\boldsymbol{y}), \quad \boldsymbol{y} \in \mathcal{Y}^n \tag{12}$$

where (f, φ) is a deterministic code of blocklength n and where $\boldsymbol{T}$ is a *random permutation mapping*. By this we mean that $\boldsymbol{T}$ maps any n-tuple $x = (x_1, \cdots, x_n)$ into a randomly selected permutation $(x_{\pi_1}, \cdots, x_{\pi_n})$, with all such permutations equally likely. Hereafter, we write $(F, \Phi) = (\boldsymbol{T}f, \varphi\boldsymbol{T}^{-1})$ as a shorthand for (12).

We may define the *RM code capacity* C^* and the *RM code reliability function* $E^*(R)$ of the AVC subject to constraints by restricting the codes in the definitions of C and $E(R)$, respectively, to the form (12). Our objective is to determine whether the class of RM codes can achieve the same performance as random codes with no structural restrictions. Of course, the upper bound in Theorem 3 applies to RM codes, as does the converse proof of (5); hence $E^*(R) \leq E_{sp}(R)$ and $C^* \leq C$.

In (12), $\boldsymbol{T}$ plays a role equivalent to Ahlswede's "robustification" technique [2, sec. IV-B]. Using this technique, it is easily shown for $Q \in \mathcal{D}_n(\mathcal{S})$ and $\boldsymbol{s} \in \mathcal{T}_Q$

$$e(\boldsymbol{s}, \boldsymbol{T}f, \varphi\boldsymbol{T}^{-1}) \leq (n+1)^{|\mathcal{S}|} e((QW)^n, f, \varphi). \tag{13}$$

It is immediate from (13) that any rate achievable for the compound DMC $\bar{\mathcal{W}} = \{QW(\cdot\,|\,\cdot)\colon Q \in \mathcal{Q}_\Lambda\}$ subject to the input constraint Γ is also achievable for the AVC $\mathcal{W}$ under RM coding. A minor modification to the proof of [4, p. 173, Cor. 5.10] shows that the capacity of this compound DMC is equal to C; thus $C^* \geq C$ and hence $C^* = C$. Similarly, it follows from [4, p. 165, Theorem 5.2] that

$$\bar{E}_r(R) \triangleq \max_{P\in\mathcal{P}_\Gamma} \min_{Q\in\mathcal{Q}_\Lambda} \bar{E}_r(R, QW, P)$$

is an achievable error exponent for the compound DMC; hence (13) implies $E^*(R) \geq \bar{E}_r(R)$. From Theorem 4, it is apparent that this lower bound usually falls short of $E_r(R)$. To improve upon it, we need a refinement of (13).

Theorem 5: Let $Q \in \mathcal{D}_n(\mathcal{S})$ and $\hat{Q} \in \mathcal{D}(\mathcal{S})$ be such that $Q \ll \hat{Q}$. Then for all $\boldsymbol{x} \in \mathcal{X}^n$, $\boldsymbol{s} \in \mathcal{T}_Q$, and $\mathcal{A} \subset \mathcal{Y}^n$

$$\boldsymbol{E}W^n(\boldsymbol{T}\mathcal{A} \mid \boldsymbol{T}\boldsymbol{x}, \boldsymbol{s}) \leq [\hat{Q}^n(\mathcal{T}_Q)]^{-1} (\hat{Q}W)^n(\mathcal{A} \mid \boldsymbol{x}). \tag{14}$$

In the particular case $\mathcal{A} = \mathcal{T}_V(\boldsymbol{x})$ for some $V\colon \mathcal{X} \to \mathcal{Y}$

$$\boldsymbol{E}W^n(\boldsymbol{T}\mathcal{T}_V(\boldsymbol{x}) \mid \boldsymbol{T}\boldsymbol{x}, \boldsymbol{s}) \leq (n+1)^{|\mathcal{S}|} \exp\{-nG(V, W, P, Q)\} \tag{15}$$

where P is the type of $\boldsymbol{x}$ and

$$\begin{aligned} G(V, W, P, Q) &\triangleq \max_{\hat{Q}:\ Q\ll\hat{Q}} [D(V \,\|\, \hat{Q}W \mid P) - D(Q \,\|\, \hat{Q})] \\ &= \min_{U_{YXS}:\ U_{YX}=P\times V, U_S=Q} D(U_{YXS} \,\|\, W \times P \times Q). \end{aligned} \tag{16}$$

$\square$

Remarks: Note that in order for $G(V,W,P,Q) = +\infty$ it is necessary and sufficient that $D(V \,\|\, QW \mid P) = +\infty$. Sufficiency is immediate; for necessity, observe $G(V,W,P,Q) = +\infty$ implies there is an x, y, and $\hat{Q}$ such that

$$P(x)V(y \mid x) > 0$$

and $\hat{Q}W(y \mid x) = 0$. Since $Q \ll \hat{Q}$, it follows that $QW(y \mid x) = 0$ and hence $D(V \,\|\, QW \mid P) = +\infty$.

It is immediate from (4) and (14) that any RM code $(\boldsymbol{T}f, \varphi \boldsymbol{T}^{-1})$ and $\boldsymbol{s} \in \mathcal{T}_Q$ satisfy

$$e(\boldsymbol{s}, \boldsymbol{T}f, \varphi \boldsymbol{T}^{-1}) \le \inf_{\hat{Q}:\, Q \ll \hat{Q}} [\hat{Q}^n(\mathcal{T}_Q)]^{-1} e((\hat{Q}W)^n, f, \varphi).$$

This bound improves upon (13), as can be seen by considering the special case $\hat{Q} = Q$ and observing $[Q^n(\mathcal{T}_Q)]^{-1} \le (n+1)^{|\mathcal{S}|}$. Using Theorem 5, we can prove a counterpart of Theorem 1 for RM codes.

Theorem 6: For every $R > 0$, $\delta > 0$, and $P \in \mathcal{D}_n(\mathcal{X})$, there exists an (n, M) code (f, φ) of constant type P with $M \ge \exp\{n(R-\delta)\}$ such that for every AVC $\mathcal{W}$ and $Q \in \mathcal{D}_n(\mathcal{S})$

$$\max_{\boldsymbol{s} \in \mathcal{T}_Q} e(\boldsymbol{s}, \boldsymbol{T}f, \varphi \boldsymbol{T}^{-1}) \le \exp\{-n[E_r(R,W,P,Q) - \delta]\} \quad (17)$$

whenever $n \ge n_3(|\mathcal{Y}|, |\mathcal{X}|, |\mathcal{S}|, \delta)$, where $E_r(R,W,P,Q)$ is as defined in (6). □

Remark: The codes of Theorem 6 are universal in the sense that the same codes achieve the exponent $E_r(R,W,P,Q)$ for every AVC $\mathcal{W}$ and every state type Q. These codes consist of codewords chosen according to a packing lemma [4, p. 162, Lemma 5.1] and an MMI decoder. These same codes were shown by Csiszár, Körner, and Marton [4, p. 172, Theorem 5.8] to achieve the exponent $\bar{E}_r(R,V,P)$ for every DMC $V\colon \mathcal{X} \to \mathcal{Y}$.

We now present the main result of this section. The proof is similar to that of Theorem 3 and so is omitted.

Theorem 7: Theorem 3 holds with $E^*(R)$ replacing $E(R)$. □

We conclude that the simpler class of RM codes can achieve the same capacity and random coding exponent $E_r(R)$ as the class of random codes with no structural restrictions. In particular, for $\hat{R} \le R \le C$ (see remark following Theorem 3), RM codes achieve the random code reliability function $E(R)$.

A potential weakness of Theorems 6 and 7 concerns the size of the ensemble of $\boldsymbol{T}$, namely $n!$, which grows superexponentially with n. However, one can apply Ericson's approach [7, Theorem 1] to choose a random mapping $\boldsymbol{T}'$ with a smaller ensemble, albeit at the cost of losing the universality of Theorem 6.

D. An Application: The Expurgated Bound

As yet, no analog of the expurgated bound has appeared in the literature for the AVC, with or without constraints. Moreover, the methods used in deriving the expurgated error exponent do not appear to extend readily to the AVC. As a final application of Theorem 5, we now present an expurgated bound for RM codes (and hence also for random codes). To this end, we say that φ is a *maximum-likelihood* (ML) decoder for the channel $U\colon \mathcal{X} \to \mathcal{Y}$ and encoder f, if

$$\varphi(\boldsymbol{y}) = m \Longrightarrow U^n(\boldsymbol{y} \mid f(m)) = \max_{m' \in \mathcal{M}} U^n(\boldsymbol{y} \mid f(m')).$$

Theorem 8 (Expurgated Bound): For every $R > 0$, $\delta > 0$, $P \in \mathcal{D}_n(\mathcal{X})$, and $\hat{Q} \in \mathcal{D}(\mathcal{S})$, there exists an (n, M) code (f, φ) of constant type P such that φ is the ML decoder for the channel $\hat{Q}W\colon \mathcal{X} \to \mathcal{Y}$ and f, $M \ge \exp\{n(R-\delta)\}$, and such that

$$\max_{\boldsymbol{s} \in \mathcal{T}_Q} e(\boldsymbol{s}, \boldsymbol{T}f, \varphi \boldsymbol{T}^{-1}) \le \exp\{-n[E_x(R,W,\hat{Q},P,Q) - \delta]\} \quad (18)$$

for every AVC $\mathcal{W}$, every $Q \in \mathcal{D}_n(\mathcal{S})$, $Q \ll \hat{Q}$, and all $n \ge n_4(|\mathcal{Y}|, |\mathcal{X}|, |\mathcal{S}|, \delta)$, where

$$E_x(R,W,\hat{Q},P,Q) \triangleq \min_{\substack{V\colon \mathcal{X} \to \mathcal{X},\, I(P,V) \le R, \\ P \times V = (P \times V)^\dagger}} [J(\hat{Q}W, V, P) - D(Q \,\|\, \hat{Q}) + I(P,V) - R] \quad (19)$$

$$J(V', V, P) \triangleq -\sum_{(x,\hat{x}) \in \mathcal{X}^2} P(x)V(\hat{x} \mid x) \cdot \log\left\{\sum_{\boldsymbol{y} \in \mathcal{Y}} \sqrt{V'(y \mid x)V'(y \mid \hat{x})}\right\}$$

and $(P \times V)^\dagger(x, \hat{x}) \triangleq (P \times V)(\hat{x}, x)$. □

As before, the constant composition result implies a bound on the reliability function for the AVC subject to constraints. The proof of the next theorem is similar to Theorem 3 and so is omitted.

Theorem 9: $E^*(R) \ge E_x(R)$ for all $R > 0$, where

$$E_x(R) \triangleq \max_{P \in \mathcal{P}_\Gamma, \hat{Q} \in \mathcal{D}(\mathcal{S})} \min_{Q \in \mathcal{Q}_\Lambda} E_x(R,W,\hat{Q},P,Q). \quad \square$$

Remark: Since $E(R) \ge E^*(R)$, it follows that $E_x(R)$ is also a lower bound to the random code reliability function. We conjecture that if $C > 0$ then $E_x(R) > E_r(R)$ for R sufficiently small, but we have not succeeded in proving this in general. This is true, however, for all examples we have calculated, including the one presented in Section IV.

III. Proofs

Throughout this section and in the Appendix, we use the elementary type identities and estimates given in [4, pp. 29–32] without further reference. Before proceeding to the proof of Theorem 1, we need the following lemma. We omit the proof, which is a straightforward variation of [4, p. 162, Lemma 5.1].

Lemma 1: For any $R > 0$, $\delta > 0$, $P \in \mathcal{D}_n(\mathcal{X})$, and $M \le \exp\{n(R-\delta)\}$, let the random variables $Z_1, \cdots, Z_M$ be independent and uniformly distributed on $\mathcal{T}_P^n$. Then for all $\boldsymbol{x} \in \mathcal{X}^n$, $\boldsymbol{s} \in \mathcal{S}^n$, and all stochastic matrices $U\colon \mathcal{X} \times \mathcal{S} \to \mathcal{Y}$ and $V\colon \mathcal{X} \to \mathcal{Y}$

$$\boldsymbol{E}\left|\mathcal{T}_U(\boldsymbol{x}, \boldsymbol{s}) \cap \bigcup_{j=1}^{M} \mathcal{T}_V(Z_j)\right| \le |\mathcal{T}_U(\boldsymbol{x}, \boldsymbol{s})| \exp\{-n|I(P,V) - R|^+\}$$

for all $n \ge n_5(|\mathcal{X}|, \delta)$. □

Proof of Theorem 1: Consider first a fixed code (f,φ) of constant type P where φ is an MMI decoder. When message i is sent and the state sequence is $\boldsymbol{s} \in \mathcal{T}_Q^n$, observe that $\boldsymbol{y} \in \mathcal{Y}^n$ is decoded incorrectly only if

$$(\boldsymbol{y}, f(i), \boldsymbol{s}) \in \mathcal{T}_{U_{YXS}}, \boldsymbol{y} \in \mathcal{T}_V(f(j))$$

and

$$I(P,V) \geq I(P,U_{YX})$$

for some $j \neq i$, some joint type U_{YXS} on $\mathcal{Y}\times\mathcal{X}\times\mathcal{S}$ satisfying $U_X = P$ and $U_S = Q$, and some conditional type $V\colon \mathcal{X} \to \mathcal{Y}$. Let

$$\mathcal{A}(\boldsymbol{x},\boldsymbol{s}) \triangleq \{\boldsymbol{y} \in \mathcal{Y}^n\colon (\boldsymbol{y},\boldsymbol{x},\boldsymbol{s}) \in \mathcal{T}_{U_{YXS}}\}.$$

Now let $R > 0$, $\delta > 0$, $P \in \mathcal{D}_n(\mathcal{X})$, $M = \lceil \exp\{n(R-\delta)\}\rceil$, and let (F,Φ) be an (n,M) random code such that $F(i) \triangleq Z_i$, $i \in \mathcal{M}$, are independent and uniform on $\mathcal{T}_P$ and Φ is an MMI decoder for F. For all $Q \in \mathcal{D}_n(\mathcal{S})$ and $\boldsymbol{s} \in \mathcal{T}_Q$, the error probability of (F,Φ) is bounded by

$$\begin{aligned}
&\boldsymbol{E}e_i(\boldsymbol{s},F,\Phi)\\
&\quad\triangleq \boldsymbol{E}W^n(\{\boldsymbol{y}\colon \Phi(\boldsymbol{y}) \neq i\} \mid Z_i,\boldsymbol{s})\\
&\quad\leq \sum_{\substack{U_{YXS}\colon U_X=P, U_S=Q\\ V\colon I(P,V)\geq I(P,U_{Y|X})}} \boldsymbol{E}W^n(\mathcal{A}(Z_i,\boldsymbol{s})\\
&\qquad\cap \bigcup_{j\neq i} \mathcal{T}_V(Z_j) \mid Z_i,\boldsymbol{s})\\
&\quad\leq (n+1)^{|\mathcal{X}||\mathcal{Y}|(|\mathcal{S}|+1)}\\
&\qquad\times \max_{\substack{U_{YXS}\colon U_X=P, U_S=Q\\ V\colon I(P,V)\geq I(P,U_{Y|X})}} \boldsymbol{E}W^n(\mathcal{A}(Z_i,\boldsymbol{s})\\
&\qquad\cap \bigcup_{j\neq i} \mathcal{T}_V(Z_j) \mid Z_i,\boldsymbol{s}) \qquad (20)
\end{aligned}$$

where the second inequality follows by observing that there are at most $(n+1)^{|\mathcal{X}||\mathcal{Y}||\mathcal{S}|}$ types U_{YXS} and $(n+1)^{|\mathcal{X}||\mathcal{Y}|}$ types V.

Note that $\mathcal{A}(\boldsymbol{x},\boldsymbol{s}) = \mathcal{T}_{U_{Y|XS}}(\boldsymbol{x},\boldsymbol{s})$ when $\boldsymbol{x} \in \mathcal{T}_{U_{X|S}}(\boldsymbol{s})$ and otherwise $\mathcal{A}(\boldsymbol{x},\boldsymbol{s}) = \emptyset$; hence

$$\begin{aligned}
&\boldsymbol{E}W^n\left(\mathcal{A}(Z_i,\boldsymbol{s}) \cap \bigcup_{j\neq i} \mathcal{T}_V(Z_j) \mid Z_i,\boldsymbol{s}\right)\\
&\quad= \frac{1}{|\mathcal{T}_P|} \sum_{\boldsymbol{x}\in\mathcal{T}_{U_{X|S}}(\boldsymbol{s})} \boldsymbol{E}W^n(\mathcal{T}_{U_{Y|XS}}(\boldsymbol{x},\boldsymbol{s})\\
&\qquad\cap \bigcup_{j\neq i} \mathcal{T}_V(Z_j) \mid \boldsymbol{x},\boldsymbol{s})\\
&\quad= \frac{1}{|\mathcal{T}_P|} \sum_{\boldsymbol{x}\in\mathcal{T}_{U_{X|S}}(\boldsymbol{s})} \boldsymbol{E}\left|\mathcal{T}_{U_{Y|XS}}(\boldsymbol{x},\boldsymbol{s}) \cap \bigcup_{j\neq i} \mathcal{T}_V(Z_j)\right|\\
&\qquad\times \exp\{-n[D(U_{Y\mid XS}\,\|\,W \mid U_{XS})\\
&\qquad+ H(U_{Y\mid XS} \mid U_{XS})]\}
\end{aligned}$$

which follows by observing that $W^n(\boldsymbol{y} \mid \boldsymbol{x},\boldsymbol{s})$ equals the last exponent above for all $(\boldsymbol{y},\boldsymbol{x},\boldsymbol{s}) \in \mathcal{T}_{U_{YXS}}$. Applying Lemma 1 with $M' \triangleq M-1 \leq \exp\{n(R-\delta)\}$, we obtain for $n \geq n_5(|\mathcal{X}|,\delta)$

$$\begin{aligned}
&\boldsymbol{E}W^n(\mathcal{A}(Z_i,\boldsymbol{s}) \cap \bigcup_{j\neq i} \mathcal{T}_V(Z_j) \mid Z_i,\boldsymbol{s})\\
&\quad\leq \frac{1}{|\mathcal{T}_P|} \sum_{\boldsymbol{x}\in\mathcal{T}_{U_{X|S}}(\boldsymbol{s})} \left|\mathcal{T}_{U_{Y|XS}}(\boldsymbol{x},\boldsymbol{s})\right|\\
&\qquad\times \exp\{-n[D(U_{Y\mid XS}\,\|\,W \mid U_{XS})\\
&\qquad+H(U_{Y|XS}|U_{XS}) + |I(P,V)-R|^+]\}\\
&\quad\leq \frac{|\mathcal{T}_{U_{X|S}}(\boldsymbol{s})|}{|\mathcal{T}_P|} \exp\{-n[D(U_{Y\mid XS}\,\|\,W \mid U_{XS})\\
&\qquad+|I(P,V)-R|^+]\}\\
&\quad\leq (n+1)^{|\mathcal{X}|} \exp\{-n[D(U_{Y\mid XS}\,\|\,W \mid U_{XS}) + H(P)\\
&\qquad-H(U_{X|S} \mid Q) + |I(P,V)-R|^+]\}\\
&\quad= (n+1)^{|\mathcal{X}|} \exp\{-n[D(U_{YXS}\,\|\,W\times P\times Q)\\
&\qquad+|I(P,V)-R|^+]\} \qquad (21)
\end{aligned}$$

where the last step follows from

$$H(P) - H(U_{X|S} \mid Q) = D(U_{XS}\,\|\,P\times Q)$$

and the chain rule for divergence. Substituting (21) into (20) and observing that $I(P,V) \geq I(P,U_{Y|X})$ implies

$$|I(P,V)-R|^+ \geq |I(P,U_{Y|X})-R|^+$$

we obtain

$$\begin{aligned}
\boldsymbol{E}e_i(\boldsymbol{s},F,\Phi) &\leq (n+1)^{|\mathcal{X}|(|\mathcal{S}||\mathcal{Y}|+|\mathcal{Y}|+1)}\\
&\quad\times \exp\{-nE_r(R,W,P,Q)\}.
\end{aligned}$$

Since the bound holds for all $i \in \mathcal{M}$ and $\boldsymbol{s} \in \mathcal{T}_Q^n$, Theorem 1 is proved for

$$\begin{aligned}
n_1(|\mathcal{Y}|,|\mathcal{X}|,|\mathcal{S}|,\delta) \triangleq \min\{&n \geq n_5(|\mathcal{X}|,\delta)\colon \exp\{-n\delta\}\\
&\times (n+1)^{|\mathcal{X}|(|\mathcal{S}||\mathcal{Y}|+|\mathcal{Y}|+1)} \leq 1\}.
\end{aligned}$$

□

Proof of Theorem 2: For any $Q \in \mathcal{D}_n(\mathcal{S})$, define the channel

$$\hat{W}_{n,Q}(\boldsymbol{y} \mid \boldsymbol{x}) \triangleq \frac{1}{|\mathcal{T}_Q|} \sum_{\boldsymbol{s}\in\mathcal{T}_Q} W^n(\boldsymbol{y} \mid \boldsymbol{x},\boldsymbol{s}) \qquad (22)$$

so that the middle expression in (7) can be written as $\max_m e_m(\hat{W}_{n,Q},F,\Phi)$. The following lemma is proved in the Appendix.

Lemma 2: Let U_{YXS} be any joint type in $\mathcal{D}_n(\mathcal{Y}\times\mathcal{X}\times\mathcal{S})$ satisfying $U_S = Q$ and $U_X = P$. Then for all $\boldsymbol{x} \in \mathcal{T}_P$

$$\begin{aligned}
\hat{W}_{n,Q}(\mathcal{T}_{U_{Y|X}}(\boldsymbol{x}) \mid \boldsymbol{x}) &\geq (n+1)^{-(1+|\mathcal{Y}|)|\mathcal{X}||\mathcal{S}|}\\
&\quad\times \exp\{-nD(U_{YXS}||W\times P\times Q)\}.
\end{aligned}$$

The first inequality in (7) is immediate. To prove the second, fix $\delta > 0$ and let (F,Φ) be any (n,M) random code of constant type P with $M \geq \exp\{n(R+\delta)\}$. By the usual argument made in passing from a random code to a deterministic code, there exists a deterministic $(n,M/2)$ code (f,φ) of constant type P such that

$$\max_m e_m(\hat{W}_{n,Q},f,\varphi) \leq 2\max_m e_m(\hat{W}_{n,Q},F,\Phi). \qquad (23)$$

Let U_{YXS} be any joint type in $\mathcal{D}_n(\mathcal{Y} \times \mathcal{X} \times \mathcal{S})$ satisfying $U_X = P$, $U_S = Q$, and $I(P, U_{Y|X}) \leq R + \delta/2$. For all n sufficiently large, we must have

$$|\mathcal{T}_{U_{Y|X}}(f(m)) \cap \varphi^{-1}(m)| < \frac{1}{2}|\mathcal{T}_{U_{Y|X}}(f(m))|$$

for some m, since otherwise

$$\begin{aligned}\exp\{nH(U_Y)\} \geq |\mathcal{T}_{U_Y}| &\geq \left|\bigcup_m \mathcal{T}_{U_{Y|X}}(f(m)) \cap \varphi^{-1}(m)\right| \\ &\geq \frac{(n+1)^{-|\mathcal{Y}||\mathcal{X}|}}{4} \exp\{n(R+\delta)\} \\ &\quad \times \exp\{nH(U_{Y|X} \mid P)\}\end{aligned}$$

which contradicts the assumption $I(P, U_{Y|X}) \leq R + \delta/2$ for all n such that

$$\exp\{n\delta/2\} > 4(n+1)^{|\mathcal{Y}||\mathcal{X}|}.$$

Thus

$$|\mathcal{T}_{U_{Y|X}}(f(m)) \cap \{\boldsymbol{y}\colon \varphi(\boldsymbol{y}) \neq m\}| > \frac{1}{2}|\mathcal{T}_{U_{Y|X}}(f(m))|$$

for some m. Since $\hat{W}_{n,Q}(\cdot \mid \boldsymbol{x})$ is constant on $\mathcal{T}_{U_{Y|X}}(\boldsymbol{x})$, it follows from Lemma 2 that for this m

$$\begin{aligned}\hat{W}_{n,Q}(\{\boldsymbol{y}\colon \varphi(\boldsymbol{y}) \neq m\} \mid f(m)) &\geq \frac{(n+1)^{-(1+|\mathcal{Y}|)|\mathcal{X}||\mathcal{S}|}}{2} \\ &\quad \times \exp\{-nD(U_{YXS}||W \times P \times Q)\}.\end{aligned}$$

Combining this with (23), we conclude that

$$\begin{aligned}&\max_{m \in \mathcal{M}} e_m(\hat{W}_{n,Q}, F, \Phi) \\ &\qquad \geq \exp\{-n[D(U_{YXS}||W \times P \times Q) + \delta/2]\} \quad (24)\end{aligned}$$

holds for all nonempty types U_{YXS} such that $U_X = P$, $U_S = Q$, and $I(P, U_{Y|X}) \leq R + \delta/2$, and for sufficiently large n.

The proof is now completed with an approximation argument: Let U'_{YXS} achieve the minimum in (8) so that $I(P, U'_{Y|X}) \leq R$. It is easily shown that, for each n, there is an approximation $U^{(n)}_{YXS}$ to this probability such that $U^{(n)}_X = P$, $U^{(n)}_S = Q$, $\mathcal{T}_{U^{(n)}_{YXS}} \neq \emptyset$, and

$$\|U^{(n)}_{YXS} - U'_{YXS}\| < \theta_n \triangleq 2|\mathcal{X}||\mathcal{Y}||\mathcal{S}|/n$$

where $\|\cdot\|$ denotes the variational distance. Using this and [4, p. 33, Lemma 2.7], we can show for $\theta_n \leq 1/2$

$$\begin{aligned}I(P, U^{(n)}_{Y|X}) &\leq I(P, U'_{Y|X}) + g(\theta_n) \\ D(U^{(n)}_{YXS}||W \times P \times Q) &\leq E_{sp}(R, W, P, Q) + g(\theta_n)\end{aligned}$$

where

$$g(\theta) \triangleq -\theta \log(\gamma\theta/|\mathcal{X}||\mathcal{Y}||\mathcal{S}|)$$

and γ is the smallest nonzero value of W. Taking n large enough to ensure $g(\theta_n) \leq \delta/2$, we obtain $I(P, U^{(n)}_{Y|X}) \leq R + \delta/2$ and hence (24) holds for $U_{YXS} = U^{(n)}_{YXS}$. Substituting the second inequality above into (24), we obtain (7) thereby completing the proof. □

Proof of Theorem 3: Throughout this proof, let P_n and Q_n denote generic elements in $\mathcal{D}_n(\mathcal{X})$ and $\mathcal{D}_n(\mathcal{S})$, respectively. For fixed blocklength n, a random code of constant type P_n satisfies the input constraint Γ if and only if $P_n \in \mathcal{P}_\Gamma$; a channel sequence $\boldsymbol{s} \in \mathcal{T}_{Q_n}$ satisfies the channel constraint Λ if and only if $Q_n \in \mathcal{Q}_\Lambda$. From Theorem 1, it follows immediately that

$$E_L(R) \triangleq \varliminf_{n \to +\infty} \max_{P_n \in \mathcal{P}_\Gamma} \min_{Q_n \in \mathcal{Q}_\Lambda} E_r(R, W, P_n, Q_n)$$

is an achievable error exponent at rate R; hence $E_L(R) \leq E(R)$.

Conversely, corresponding to any (n, M) random code (F, Φ) there is an $(n, M/2)$ deterministic code (f, φ) which satisfies (23), where $\hat{W}_{n,Q}$ is as defined in (22). Moreover, since (F, Φ) satisfies the input constraint Γ, (f, φ) can be chosen to satisfy it as well. Since $|\mathcal{D}_n(\mathcal{X})| \leq (n+1)^{|\mathcal{X}|}$, (f, φ) contains a subcode (f', φ') of constant type P_n, for some $P_n \in \mathcal{P}_\Gamma$, which has at least $M(n+1)^{-|\mathcal{X}|}/2$ codewords, so that

$$\max_m e_m(\hat{W}_{n,Q}, F, \Phi) \geq \frac{1}{2} \max_m e_m(\hat{W}_{n,Q}, f', \varphi').$$

Using Theorem 2 to bound the error probability of (f', φ'), we obtain $E(R) \leq E_U(R)$ where

$$E_U(R) \triangleq \varlimsup_{n \to +\infty} \max_{P_n \in \mathcal{P}_\Gamma} \min_{Q_n \in \mathcal{Q}_\Lambda} E_{sp}(R_n, W, P_n, Q_n)$$

and $R_n \geq R - (|\mathcal{X}|/n)\log(n+1) - 1/n$. The following lemma, which is proved in the Appendix, completes the proof of Theorem 3.

Lemma 3: For all $R > 0$,

a) $E_r(R) \leq E_L(R)$, and

b) $E_U(R) \leq E_{sp}(R)$ except possibly at $R = R_\infty(W)$.

Proof of Theorem 4: For any $Q \in \mathcal{D}(\mathcal{S})$, let $\mathcal{V}_Q$ be the set of all $V\colon \mathcal{X} \to \mathcal{Y}$ achieving the minimum in (11) for $\hat{W} \triangleq QW$. We say that Q satisfies the *equivalence condition* if there exists a $V^* \in \mathcal{V}_Q$ such that for every $s \in \mathcal{S}$

$$\sum_{x,y} \frac{P(x)V^*(y \mid x)W(y \mid x, s)Q(s)}{QW(y \mid x)} = Q(s). \quad (25)$$

Theorem 4 will follow if we prove the two assertions:

a) $E_r(R, W, P, Q) \geq \bar{E}_r(R, QW, P)$ with equality if and only if Q satisfies the equivalence condition.

b)

$$\bar{E}_r(R, QW, P) = \min_{Q' \in \mathcal{D}(\mathcal{S})\colon Q' \ll Q} \bar{E}_r(R, Q'W, P)$$

if and only if Q satisfies the equivalence condition.

To prove a), it is convenient to replace $U_{Y|X}$ by V and $U_{S|YX}$ by U in (6) to obtain

$$\begin{aligned}&E_r(R, W, P, Q) \\ &\quad = \min_{\substack{V\colon \mathcal{X} \to \mathcal{Y} \\ U\colon \mathcal{X} \times \mathcal{Y} \to \mathcal{S}, (P \times V)U = Q}} D(V \times P \times U \,\|\, W \times P \times Q) \\ &\qquad + |I(P, V) - R|^+ \\ &\quad = \min_{V\colon \mathcal{X} \to \mathcal{Y}} [D(V \,\|\, QW \mid P) + |I(P, V) - R|^+\end{aligned}$$

$$+ \min_{\substack{U:\mathcal{X}\times\mathcal{Y}\to\mathcal{S},\\ (P\times V)U=Q}} D(U \,\|\, (W\times Q)/QW \mid P\times V)]$$
$$\geq \bar{E}_r(R, QW, P) \tag{26}$$

where the second equality follows from the chain rule for divergence. Observe that equality is achieved if and only if for some $V^* \in \mathcal{V}_Q$ there exists a $U\colon \mathcal{X}\times\mathcal{Y}\to\mathcal{S}$ such that $(P\times V^*)U = Q$ and

$$D(U \,\|\, (W\times Q)/QW \mid P\times V^*) = 0.$$

Since the latter condition occurs if and only if

$$U(\cdot \mid x, y) = W(y \mid x, \cdot)Q(\cdot)/QW(y \mid x)$$

whenever $(P\times V^*)(y,x) > 0$, the requirement $(P\times V^*)U = Q$ is equivalent to (25). This completes the proof of a).

Now consider b). By interchanging the order of minimization over Q' and V (see (11)), it is easily shown that

$$\bar{E}_r(R, QW, P) = \min_{Q'\in\mathcal{D}(\mathcal{S})\colon\, Q'\ll Q} \bar{E}_r(R, Q'W, P)$$

holds if and only if

$$D(V^* \,\|\, QW \mid P) = \min_{Q'\in\mathcal{D}(\mathcal{S})\colon\, Q'\ll Q} D(V^* \,\|\, Q'W \mid P)$$

for some $V^* \in \mathcal{V}_Q$. By the Kuhn–Tucker Theorem [8, p. 314], Q minimizes the convex functional

$$\Psi(Q') \triangleq D(V^* \,\|\, Q'W \mid P)$$

subject to the constraints $Q'(s) \geq 0$, $\sum_s Q'(s) = 1$, and $Q' \ll Q$, if and only if there exists a real number λ such that

$$-\frac{\partial\Psi(Q')}{\partial Q'(s)}\bigg|_{Q'=Q} = \sum_{x,y} \frac{P(x)V^*(y \mid x)W(y \mid x, s)}{QW(y \mid x)} = \lambda,$$
$$\text{for all } s\colon Q(s) > 0.$$

Averaging over Q, we find $\lambda = 1$. Since this is equivalent to (25), the proof is complete. □

Proof of Theorem 5: First, observe that for any $\boldsymbol{s} \in \mathcal{T}_Q$

$$\boldsymbol{E}W^n(\boldsymbol{T}\boldsymbol{y} \mid \boldsymbol{T}\boldsymbol{x}, \boldsymbol{s}) = \boldsymbol{E}W^n(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{T}^{-1}\boldsymbol{s})$$
$$= \hat{W}_{n,Q}(\boldsymbol{y} \mid \boldsymbol{x}) \tag{27}$$

where $\hat{W}_{n,Q}$ is as defined in (22). The first step holds because applying the same permutation (namely $\boldsymbol{T}^{-1}$) to the input, output, and state sequences leaves W^n unchanged. The second follows by observing that $\boldsymbol{T}^{-1}\boldsymbol{s}$ is uniformly distributed on $\mathcal{T}_Q$.

For any $Q \in \mathcal{D}_n(\mathcal{S})$ and $\hat{Q} \in \mathcal{D}(\mathcal{S})$, we can write

$$(\hat{Q}W)^n(\mathcal{A} \mid \boldsymbol{x}) = \sum_{\boldsymbol{s}'\in\mathcal{S}^n} \hat{Q}^n(\boldsymbol{s}')\, W^n(\mathcal{A} \mid \boldsymbol{x}, \boldsymbol{s}')$$
$$\geq \frac{\hat{Q}^n(\mathcal{T}_Q)}{|\mathcal{T}_Q|} \sum_{\boldsymbol{s}'\in\mathcal{T}_Q} W^n(\mathcal{A} \mid \boldsymbol{x}, \boldsymbol{s}')$$
$$= \hat{Q}^n(\mathcal{T}_Q)\, \hat{W}_{n,Q}(\mathcal{A} \mid \boldsymbol{x}). \tag{28}$$

This completes the proof of (14).

Equation (15) follows immediately from (14) by substituting the elementary type estimates

$$(\hat{Q}W)^n(\mathcal{T}_V(\boldsymbol{x}) \mid \boldsymbol{x}) \leq \exp\{-nD(V \,\|\, \hat{Q}W \mid P)\}$$

and

$$\hat{Q}^n(\mathcal{T}_Q^n) \geq (n+1)^{-|\mathcal{S}|} \exp\{-nD(Q \,\|\, \hat{Q})\}$$

and minimizing the resulting bound over all $\hat{Q}$ satisfying $Q \ll \hat{Q}$.

It only remains to prove the second equality in (16). To this end, let

$$D_1 \triangleq \min_{U_{YXS}\colon\, U_{YX}=P\times V, U_S=Q} D(U_{YXS} \,\|\, W\times P\times Q)$$
$$D_2 \triangleq \max_{\hat{Q}\colon\, Q\ll\hat{Q}} [D(V \,\|\, \hat{Q}W \mid P) - D(Q \,\|\, \hat{Q})]. \tag{29}$$

For any U_{YXS} satisfying $U_{YX} = P\times V$ and $U_S = Q$, the chain rule for divergence yields

$$D(V \,\|\, \hat{Q}W \mid P) - D(Q \,\|\, \hat{Q})$$
$$= D(U_{YX} \,\|\, \hat{Q}W\times P) - D(U_S \,\|\, \hat{Q})$$
$$\leq D(U_{YXS} \,\|\, W\times P\times \hat{Q}) - D(U_S \,\|\, \hat{Q})$$
$$= D(U_{YXS} \,\|\, W\times P\times Q)$$

from which follows $D_1 \geq D_2$. This completes the proof if $D(V \,\|\, QW \mid P) = +\infty$, since $D_1 \geq D_2 \geq D(V \,\|\, QW \mid P)$. Now we prove the converse inequality for $D(V \,\|\, QW \mid P) < +\infty$. Let $\hat{Q}'$ achieve the maximum of

$$\Upsilon(\hat{Q}) \triangleq D(V \,\|\, \hat{Q}W \mid P) - D(Q \,\|\, \hat{Q})$$

subject to the constraints $\hat{Q}(s) \geq 0$, $Q \ll \hat{Q}$, $\sum_s \hat{Q}(s) = 1$. By the Kuhn–Tucker theorem [8, p. 314], there is a λ such that

$$-\frac{\partial\Upsilon(\hat{Q})}{\partial\hat{Q}(s)}\bigg|_{\hat{Q}=\hat{Q}'} = \sum_{x,y} \frac{P(x)V(y \mid x)W(y \mid x, s)}{\hat{Q}'W(y \mid x)} - \frac{Q(s)}{\hat{Q}'(s)} = \lambda,$$
$$\text{for all } s\colon \hat{Q}'(s) > 0. \tag{30}$$

Recalling the remarks following Theorem 5, it is easily shown that $\hat{Q}'W(y \mid x) = 0$ implies $V(y \mid x) = 0$. Thus averaging (30) over $\hat{Q}'$, we obtain $\lambda = 0$. Now define

$$U'_{YXS}(y, x, s) \triangleq \frac{P(x)V(y \mid x)W(y \mid x, s)\hat{Q}'(s)}{\hat{Q}'W(y \mid x)}$$

for $\hat{Q}'W(y \mid x) > 0$ and $U'_{YXS} \triangleq 0$ elsewhere. It follows from (30) that $U'_{YX} = P\times V$ and $U'_S = Q$. Hence

$$D_1 \leq D(U'_{YXS} \,\|\, W\times P\times Q)$$
$$= D(V \,\|\, \hat{Q}'W \mid P) - D(Q \,\|\, \hat{Q}') \triangleq D_2$$

thereby completing the proof. □

Proof of Theorem 6: By (27)

$$\boldsymbol{E}W^n(\boldsymbol{T}\boldsymbol{y} \mid \boldsymbol{T}\boldsymbol{x}, \boldsymbol{s}) = \hat{W}_{n,Q}(\boldsymbol{y} \mid \boldsymbol{x})$$

for all $\boldsymbol{s} \in \mathcal{T}_Q$. Thus it suffices to prove the theorem with $\max_m e_m(\hat{W}_{n,Q}, f, \varphi)$ replacing the left side of (17).

We may assume $R < H(P)$, since $E_r(R, W, P, Q)$ is zero otherwise. Fix $\delta > 0$ and let $\{f(m)\colon m \in \mathcal{M}\}$ be a collection of $M \geq \exp\{n(R-\delta)\}$ codewords chosen as in the Packing Lemma [4, p. 162, Lemma 5.1], which is possible

for all $n \geq n_0(|\mathcal{Y}|, |\mathcal{X}|, \delta)$. Let φ denote the MMI decoder corresponding to f. We claim that

$$\max_m e_m(\hat{W}_{n,Q}, f, \varphi) \leq (n+1)^{2|\mathcal{X}\,\|\,\mathcal{Y}|+|\mathcal{S}|} \exp\{-nE_r^*(R, W, P, Q)\}$$

where

$$E_r^*(R, W, P, Q) \triangleq \min_{V:\ \mathcal{X}\to\mathcal{Y}} G(V, W, P, Q) + |I(P, V) - R|^+.$$

The proof of this claim, which is omitted, is identical to the proof of [4, p. 165, Theorem 5.2] with W^n replaced everywhere by $\hat{W}_{n,Q}$ and with $\exp\{-nD(V \,\|\, W|P)\}$ replaced everywhere by $(n+1)^{|\mathcal{S}|}\exp\{-nG(V, W, P, Q)\}$. The latter replacement is justified by Theorem 5. From (16), it follows that

$$E_r^*(R, W, P, Q) = E_r(R, W, P, Q).$$

Theorem 6 now follows for $n_3(|\mathcal{Y}|, |\mathcal{X}|, |\mathcal{S}|, \delta)$ chosen to be the smallest $n \geq n_0(|\mathcal{Y}|, |\mathcal{X}|, \delta)$ such that

$$(n+1)^{2|\mathcal{X}\,\|\,\mathcal{Y}|+|\mathcal{S}|} \leq \exp\{n\delta\}. \qquad \square$$

Proof of Theorem 8: We may assume $R < H(P)$, since $E_x(R, W, \hat{Q}, P, Q)$ is zero otherwise. From [4, p. 162, Lemma 5.1], for all $\delta > 0$ and $n \geq n_0(|\mathcal{Y}|, |\mathcal{X}|, \delta)$, there exist at least $\exp\{n(R-\delta)\}$ sequences $\boldsymbol{x}_m \in \mathcal{X}^n$ of type P such that for every m and every $V\colon \mathcal{X} \to \mathcal{X}$ the number of $\boldsymbol{x}_j$'s in $\mathcal{T}_V(\boldsymbol{x}_m)$ is not more than $\exp\{n[R - I(P,V)]\}$. Let $(\boldsymbol{T}f, \varphi\boldsymbol{T}^{-1})$ be an RM code such that $f(m) \triangleq \boldsymbol{x}_m$, $m \in \mathcal{M}$, and φ is the ML decoder for the channel $\hat{Q}W$. Applying Theorem 5 with $\mathcal{A} = \{\boldsymbol{y}\}$, we obtain for all $\boldsymbol{s} \in \mathcal{T}_Q$, $Q \ll \hat{Q}$,

$$\begin{aligned}
&\boldsymbol{E}e_m(\boldsymbol{s}, \boldsymbol{T}f, \varphi\boldsymbol{T}^{-1}) \\
&\leq \sum_{m'\neq m} \sum_{\boldsymbol{y}:\ \hat{Q}W^n(\boldsymbol{y}|f(m'))\geq\hat{Q}W^n(\boldsymbol{y}|f(m))} \boldsymbol{E}W^n(\boldsymbol{T}\boldsymbol{y} \mid \boldsymbol{T}f(m), \boldsymbol{s}) \\
&\leq \sum_{m'\neq m} \sum_{\boldsymbol{y}:\ \hat{Q}W^n(\boldsymbol{y}|f(m'))\geq\hat{Q}W^n(\boldsymbol{y}|f(m))} \frac{\hat{Q}W^n(\boldsymbol{y} \mid f(m))}{\hat{Q}(\mathcal{T}_Q)} \\
&\leq [\hat{Q}(\mathcal{T}_Q)]^{-1} \sum_{m'\neq m} \sum_{\boldsymbol{y}\in\mathcal{Y}^n} \sqrt{\hat{Q}W^n(\boldsymbol{y} \mid f(m'))\hat{Q}W^n(\boldsymbol{y} \mid f(m))}.
\end{aligned} \tag{31}$$

For any $f(m') \in \mathcal{T}_V(f(m))$, the inner sum can be recognized as $\exp\{-nJ(\hat{Q}W, V, P)\}$.

Observing that

$$\hat{Q}^n(\mathcal{T}_Q^n) \geq (n+1)^{-|\mathcal{S}|}\exp\{-nD(Q \,\|\, \hat{Q})\}$$

and decomposing the sum over m' in (31) into all possible conditional types $V\colon \mathcal{X} \to \mathcal{X}$, we obtain

$$\begin{aligned}
\boldsymbol{E}e_m(\boldsymbol{s}, \boldsymbol{T}f, \varphi\boldsymbol{T}^{-1}) &\leq [\hat{Q}(\mathcal{T}_Q)]^{-1} \sum_{V:\ PV=P} \sum_{m'\neq m:\ f(m')\in\mathcal{T}_V(f(m))} \exp\{-nJ(\hat{Q}W, V, P)\} \\
&\leq \sum_{V:\ PV=P} (n+1)^{|\mathcal{S}|}\exp\{-n[J(\hat{Q}W, V, P) \\
&\quad - D(Q \,\|\, \hat{Q}) + I(P, V) - R]\} \\
&\leq (n+1)^{|\mathcal{S}|+|\mathcal{X}|^2}\exp\{-nE_x(R, W, \hat{Q}, P, Q)\}
\end{aligned}$$

where the last inequality follows by observing that

$$J(\hat{Q}W, V, P) - D(Q \,\|\, \hat{Q}) + I(P, V)$$

is convex in V and symmetric in $(P \times V)$ and $(P \times V)^\dagger$; hence, the minimum over $(P \times V)$ is always achieved by at least one distribution satisfying $(P \times V) = (P \times V)^\dagger$. This completes the proof of Theorem 8 for $n_4(|\mathcal{Y}|, |\mathcal{X}|, |\mathcal{S}|, \delta)$ equal to the smallest $n \geq n_0(|\mathcal{Y}|, |\mathcal{X}|, \delta)$ such that $(n+1)^{|\mathcal{S}|+|\mathcal{X}|^2} \leq \exp\{n\delta\}$. $\square$

TABLE I
PROBABILITIES USED IN CALCULATION OF $E_r(R, W, P, Q)$

(y, x, s)	$W \times P \times Q$	U_{YXS}
$(0,0,0)$	$(1-p)(1-q)$	β
$(1,0,1)$	$(1-p)q$	$1-p-\beta$
$(1,1,0)$	$p(1-q)$	$1-q-\beta$
$(0,1,1)$	pq	$p+q+\beta-1$
otherwise	0	0

IV. AN EXAMPLE

In this section, the exponents of Section II are calculated for a simple example. Consider the discrete AVC defined by $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0, 1\}$

$$W(y \mid x, s) \triangleq \begin{cases} 1, & y = x + s \text{ modulo } 2 \\ 0, & \text{otherwise} \end{cases}$$

and where the constraint functions in (2) and (3) are $g(x) = x$ and $l(s) = s$, respectively. This can be interpreted as a binary channel where codewords of length n are restricted to have Hamming weight at most $n\Gamma$, and where the channel can cause any pattern of errors with Hamming weight at most $n\Lambda$. It is shown in [6] that the random code capacity of this channel is $C = h(\Gamma \star \Lambda) - h(\Lambda)$ for $\Lambda \leq 1/2$ and $\Gamma \leq 1/2$, where $\Gamma \star \Lambda \triangleq \Gamma(1-\Lambda) + \Lambda(1-\Gamma)$ and $h(\Lambda) \triangleq -\Lambda\log\Lambda - (1-\Lambda)\log(1-\Lambda)$. Here we focus on the particular case $\Gamma = 1/2$.

We begin by calculating $E_r(R, W, P, Q)$. Let $P \triangleq (1-p, p)$ and $Q \triangleq (1-q, q)$, and observe that $W \times P \times Q$ is nonzero for only four values of (y, x, s), as shown in Table I. Thus the distribution U_{YXS} achieving the minimum in (6) is nonzero only for these same four values. Further observe that the constraints $U_X = P$, $U_S = Q$, $\sum_{y,x,s} U_{YXS}(y, x, s) = 1$, imply that U_{YXS} must take the form given in Table I for some $\beta_0 \leq \beta \leq \beta_1$ where $\beta_0 \triangleq \max\{0, 1-p-q\}$, $\beta_1 \triangleq \min\{1-p, 1-q\}$. Consequently, (6) reduces to

$$\begin{aligned}
E_r(R, W, P, Q) = \min_{\beta_0\leq\beta\leq\beta_1} &D(P_\beta \,\|\, Q_q) \\
&+ |D(P_\beta \,\|\, Q_{2-p-q-2\beta}) - R|^+
\end{aligned} \tag{32}$$

where

$$\begin{aligned}
P_\beta &\triangleq (\beta, 1-p-\beta, 1-q-\beta, p+q+\beta-1) \\
Q_q &\triangleq ((1-p)(1-q), (1-p)q, p(1-q), pq).
\end{aligned}$$

This exponent is easily evaluated numerically for any choice of p and q.

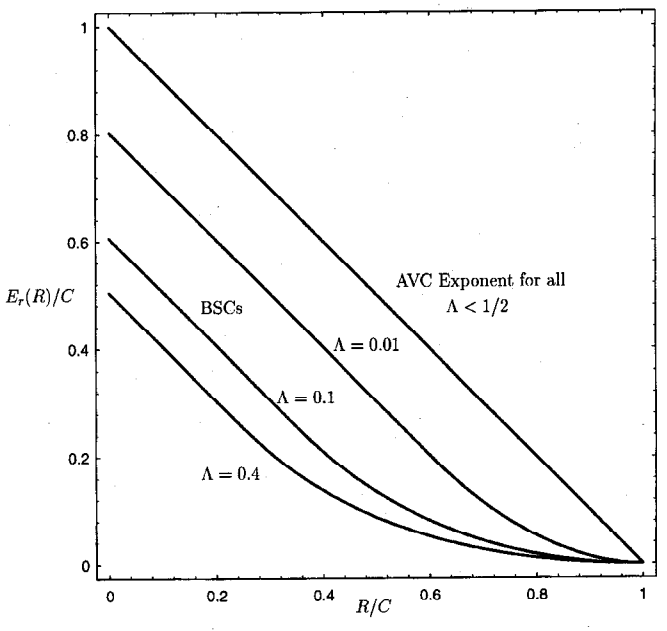

Fig. 1. Normalized exponents for BSC's and AVC.

Next we calculate $E_r(R)$. For any AVC, observe that choosing $U_{YXS} = W \times P \times Q$ in (6) gives the upper bound

$$E_r(R, W, P, Q) \le |I(P, QW) - R|^+.$$

From (5) and (9), it follows that $E_r(R) \le |C - R|^+$. Conversely, observe that if $P = P' \triangleq (1/2, 1/2)$, both divergences in (32) are convex in β and symmetric in β and $1 - q - \beta$. Thus the minimum is achieved at $\beta = (1 - q)/2$ which yields

$$E_r(R, W, P', Q) = |1 - h(q) - R|^+.$$

Since

$$\mathcal{Q}_\Lambda = \{(1 - q, q)\colon 0 \le q \le \Lambda\}$$

and

$$\mathcal{P}_\Gamma = \{(1 - p, p)\colon 0 \le p \le 1/2\}$$

we see that

$$E_r(R) \ge \min_{Q \in \mathcal{Q}_\Lambda} E_r(R, W, P', Q) = |C - R|^+$$

where $C = 1 - h(\Lambda)$ is the random code capacity. Hence $E_r(R) = |C - R|^+$.

This exponent, normalized to capacity, is plotted in Fig. 1. Also shown for several values of Λ is the random coding exponent of the binary-symmetric channel (BSC) with crossover probability Λ, which has the same capacity as the AVC with input constraint $\Gamma = 1/2$ and channel constraint Λ. Note that the AVC exponent is universally larger for $R < C$ than the exponent of the BSC with the same capacity. It is perhaps somewhat surprising that such a pessimistic model of channel interference actually yields a smaller error probability (for the best codes) than the BSC. An explanation of this phenomenon is given in the remarks following Theorem 4.

The exponents $E_{sp}(R, W, P, Q)$ and $E_{sp}(R)$ can be calculated in a similar manner. Here (8) becomes

$$E_{sp}(R, W, P, Q) = \min_{\beta_0 \le \beta \le \beta_1:\ D(P_\beta \| Q_{2-p-q-2\beta}) \le R} D(P_\beta \| Q_q). \quad (33)$$

Observe that

$$E_{sp}(R) \ge \min_{Q \in \mathcal{Q}_\Lambda} E_{sp}(R, W, P', Q)$$

where again $P' \triangleq (1/2, 1/2)$. For this choice of P, the minimum value of $D(P_\beta \| Q_{2-p-q-2\beta})$ for $\beta_0 \le \beta \le \beta_1$ is $1 - h(q)$ which is achieved by $\beta = (1 - q)/2$. Since $1 - h(q) \ge 1 - h(\Lambda)$ for all $0 \le q \le \Lambda$, it follows that $D(P_\beta \| Q_{2-p-q-2\beta}) > R$ for all $\beta_0 \le \beta \le \beta_1$ and $R < C$. Consequently, there exists no distribution U_{YXS} such that $D(U_{YXS} \| W \times P' \times Q)$ is finite and $I(P, U_{Y|X}) \le R$. We conclude that $E_{sp}(R) = +\infty$ for $0 < R < C$ and so the bound of Theorem 2 is useless.

Finally, consider the expurgated exponents $E_x(R, W, \hat{Q}, P, Q)$ and $E_x(R)$ in Theorems 8 and 9. Setting $\hat{Q} = (1 - \hat{q}, \hat{q})$, we see that $\hat{Q}W$ is a BSC with crossover probability $\hat{q}$. For $P = P'$, the condition $(P' \times V) = (P' \times V)^\dagger$ implies that V is a BSC with crossover probability v, say. Thus

$$E_x(R, W, \hat{Q}, P', Q) = \min_{v:\ h(v) \ge 1 - R} -v \log \left\{2\sqrt{\hat{q}(1 - \hat{q})}\right\} + 1 - h(v) - R - D((1 - q, q) \| (1 - \hat{q}, \hat{q})).$$

The minimum above is achieved at

$$v = \max\{h^{-1}(1 - R), 2\sqrt{\hat{q}(1 - \hat{q})}/(1 + 2\sqrt{\hat{q}(1 - \hat{q})}\}.$$

To calculate $E_x(R)$, we claim

$$E_x(R) = \max_{\hat{Q}} \min_{Q \in \mathcal{Q}_\Lambda} E_x(R, W, \hat{Q}, P', Q). \quad (34)$$

To see this, observe that the right side is clearly a lower bound to $E_x(R)$. Conversely, we may obtain an upper bound to $E_x(R)$ by restricting V in the minimization in (19) to be a BSC. The resulting upper bound is maximized by P' and reduces to the right side of (34). For each $\hat{q}$, the minimum over Q in (34) is always achieved by $q = \Lambda$ or 0. Maximizing over $\hat{q}$, we obtain $E_x(R) = +\infty$ for $R < R_\infty(W) \triangleq 1 - h(2\Lambda)$ for $\Lambda < 1/4$, and otherwise we must calculate the maximum numerically. This value of $R_\infty(W)$ is intuitively reasonable: for $R < 1 - h(2\Lambda)$, the Gilbert bound [4, p. 180] guarantees the existence of a code (f, φ) with minimum distance greater than $2n\Lambda$. Thus we can achieve a zero error probability even with deterministic codes. The exponent $E_x(R)$ is plotted in Fig. 2 for several values of Λ, together with the corresponding exponent of the BSC. Again, the AVC exponent is everywhere larger than the BSC exponent, and is even unbounded for a range of positive rates.

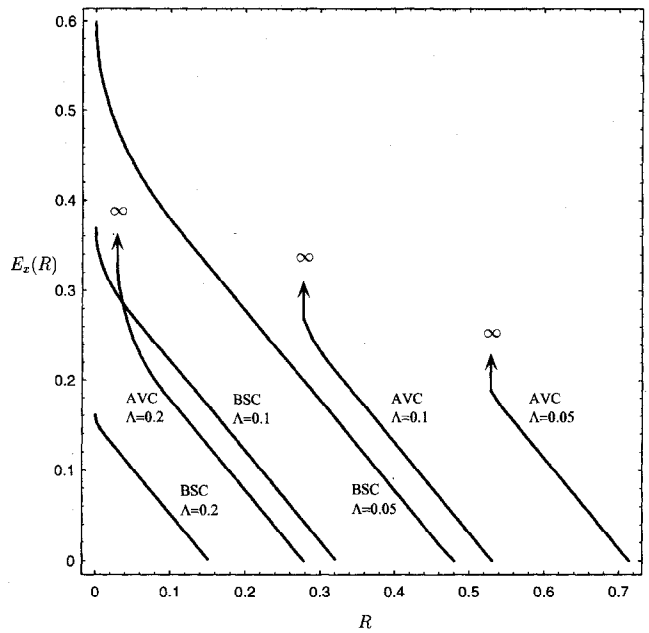

Fig. 2. Expurgated exponents for BSC's and AVC.

V. Conclusion

For many AVC's of practical interest, random codes can achieve a much larger capacity and reliability function than deterministic codes. Consequently, random codes are important not only as tools for proving coding theorems, but also as models of practical communication systems. In this paper, we considered the minimum error probability achievable over the AVC through the use of random block codes. Specifically, we presented analogs for the AVC of the random coding, sphere packing, and expurgated error exponents.

We first derived exponents for random codes and state sequences of fixed composition. These results were used to strengthen Stiglitz's bound for the AVC without constraints, and to refine Csiszár and Narayan's coding theorem for the AVC subject to the input constraint Γ and the state constraint Λ. When the state constraint is inactive in a certain sense for a given $R > 0$, the random coding exponent of the AVC reduces to the worst DMC exponent among the DMC's in the channel class $\bar{\mathcal{W}} \triangleq \{QW\colon Q \in \mathcal{Q}_\Lambda\}$. This behavior is reminiscent of the channel capacity C, which always equals the minimum capacity of the DMC's in $\bar{\mathcal{W}}$. However, when the state constraint is active, the random coding exponent of the AVC is strictly larger than the corresponding DMC exponent for all DMC's in $\bar{\mathcal{W}}$.

Coding theorems for AVC's are typically proved using random codes in which codewords are independent and identically distributed. Because of the obvious practical disadvantages of such codes, we considered a simpler subclass of random codes in which coding and randomization are separate. Specifically, an RM code consists of a deterministic code combined with a random permutation mapping. We proved a theorem relating the error probability of RM codes to the error probability of a compound DMC, and also a universal coding theorem which states that there exist RM codes that achieve the constant composition random coding exponent for every AVC and every state sequence type. Thus the class of RM codes can achieve the same capacity and random coding exponent as the class of random codes without structural restrictions. Finally, we derived the first available expurgated exponent for the AVC.

Appendix

Proof of Lemma 2: For any $\boldsymbol{s} \in \mathcal{T}_{U_{S|X}}(\boldsymbol{x})$, observe that $\mathcal{T}_{U_{Y|XS}}(\boldsymbol{x},\boldsymbol{s}) \subset \mathcal{T}_{U_{Y|X}}(\boldsymbol{x})$. Hence

$$\begin{aligned}
&\hat{W}_{n,Q}(\mathcal{T}_{U_{Y|X}}(\boldsymbol{x}) \mid \boldsymbol{x}) \\
&\quad \triangleq \frac{1}{|\mathcal{T}_Q|} \sum_{\boldsymbol{s}\in\mathcal{T}_Q} W^n(\mathcal{T}_{U_{Y|X}}(\boldsymbol{x}) \mid \boldsymbol{x},\boldsymbol{s}) \\
&\quad \geq \frac{1}{|\mathcal{T}_Q|} \sum_{\boldsymbol{s}\in\mathcal{T}_{U_{S|X}}(\boldsymbol{x})} W^n(\mathcal{T}_{U_{Y|XS}}(\boldsymbol{x},\boldsymbol{s}) \mid \boldsymbol{x},\boldsymbol{s}) \\
&\quad \geq \frac{(n+1)^{-|\mathcal{Y}\|\mathcal{X}\|\mathcal{S}|}}{|\mathcal{T}_Q|} \\
&\qquad \times \sum_{\boldsymbol{s}\in\mathcal{T}_{U_{S|X}}(\boldsymbol{x})} \exp\{-nD(U_{Y|XS}\,\|\,W \mid U_{XS})\} \\
&\quad \geq (n+1)^{-(1+|\mathcal{Y}|)|\mathcal{X}\|\mathcal{S}|} \exp\{-n[H(Q) - H(U_{S|X} \mid P) \\
&\qquad\qquad + D(U_{Y|XS}\,\|\,W \mid U_{XS})]\} \\
&\quad = (n+1)^{-(1+|\mathcal{Y}|)|\mathcal{X}\|\mathcal{S}|} \exp\{-nD(U_{YXS}\,\|\,W\times P\times Q)\}
\end{aligned}$$

where the last step follows from the chain rule for divergence. This completes the proof. □

Proof of Lemma 3: Here we prove only b), as the proof of a) is very similar. For b), it suffices to show $E_U(R) \leq E_{sp}(R)$ for $R > R_\infty(W)$, since the bound is trivial for $R < R_\infty(W)$. We will need the following lemma, whose proof is deferred until later in the Appendix.

Lemma 4: For any nonnegative function l defined on $\mathcal{S}$ and any probability distribution $U_{XS} \in \mathcal{D}(\mathcal{X}\times\mathcal{S})$, there exists an approximation $U'_{XS} \in \mathcal{D}(\mathcal{X}\times\mathcal{S})$ such that $U'_X = U_X$, $U'_S \in \mathcal{D}_n(\mathcal{S})$,

$$\sum_s l(s)U'_S(s) \leq \sum_s l(s)U_S(s)$$

and $\|U'_{XS} - U_{XS}\| < 2|\mathcal{S}|/n$.

Fix $\delta > 0$, $P \in \mathcal{P}_\Gamma$, and $Q \in \mathcal{Q}_\Lambda$. By the uniform continuity of the mutual information, there exists a $\tau_1(|\mathcal{X}|,|\mathcal{S}|,\delta)$ such that

$$\|U_{XS} - U'_{XS}\| < \tau_1(|\mathcal{X}|,|\mathcal{S}|,\delta)$$

implies

$$|I(U_S,U_{X|S}) - I(U'_S,U'_{X|S})| < \delta.$$

Set

$$\tau \triangleq \min\{\tau_1(|\mathcal{X}|,|\mathcal{S}|,\delta),\tau_1(|\mathcal{X}|,|\mathcal{Y}|,\delta/2)\}.$$

Let U_{YXS} achieve the minimum in (8) for $R' \triangleq R-\delta$, P, and Q, and observe that for all x,s satisfying

$$U_{XS}(x,s) \geq \eta \triangleq \tau/(2|\mathcal{Y}\|\mathcal{X}||\mathcal{S}| + \delta)$$

$$D(U_{Y|XS}(\cdot \mid x,s)\,\|\,W(\cdot \mid x,s)) \leq E_{sp}(R-\delta,W,P,Q)/\eta. \tag{35}$$

By Lemma 4, for all sufficiently large n there exists an approximation U'_{XS} of U_{XS} such that $U'_X = P$, $U'_S \in \mathcal{D}_n(\mathcal{S}) \cap \mathcal{Q}_\Lambda$, and $\| U_{XS} - U'_{XS} \| < \delta\eta$. Let U'_{YXS} be the probability distribution defined by U'_{XS} and

$$U'_{Y|XS}(\cdot \mid x,s) \triangleq \begin{cases} U_{Y|XS}(\cdot \mid x,s) & U_{XS}(x,y) \geq \eta \\ W(\cdot \mid x,s) & U_{XS}(x,y) < \eta \end{cases}$$

and observe that

$$\| U_{YXS} - U'_{YXS} \| < 2|\mathcal{Y}||\mathcal{X}||\mathcal{S}|\eta + \delta\eta = \tau.$$

It then readily follows that $\| U_{YX} - U'_{YX} \| < \tau$ and so

$$I(U'_X, U'_{Y|X}) < I(U_X, U_{Y|X}) + \delta/2 \leq R - \delta/2.$$

Setting $Q'_n \triangleq U'_S$ and observing that $U'_X = P$ and $I(P, U'_{Y|X}) \leq R_n$ for sufficiently large n, we obtain from (8)

$$\begin{aligned} &E_{sp}(R_n, W, P, Q'_n) \\ &\quad \leq D(U'_{YXS} \| W \times P \times Q'_n) \\ &\quad = D(U'_{Y|XS} \| W \mid U_{XS}) + D(U'_{XS} \| P \times Q'_n) \\ &\qquad + \sum_{x,s} D(U'_{Y|XS}(\cdot \mid x,s) \| W(\cdot \mid x,s)) \\ &\qquad \cdot (U'_{XS}(x,s) - U_{XS}(x,s)) \\ &\quad \leq D(U_{Y|XS} \| W \mid U_{XS}) + D(U'_{XS} \| P \times Q'_n) \\ &\qquad + \| U_{XS} - U'_{XS} \| E_{sp}(R - \delta, W, P, Q)/\eta \\ &\quad \leq (1+\delta) E_{sp}(R - \delta, W, P, Q) + \delta. \end{aligned} \tag{36}$$

Here, the second inequality follows from (35) and the definition of $U'_{Y|XS}$; the third inequality follows from $\| U_{XS} - U'_{XS} \| < \delta\eta < \tau$ and the definition of τ. Thus to every $Q \in \mathcal{Q}_\Lambda$, there is a $Q'_n \in \mathcal{D}_n(\mathcal{S}) \cap \mathcal{Q}_\Lambda$ satisfying (36). We have therefore proved that for all $R > 0$ and all sufficiently large n

$$\max_{P_n \in \mathcal{P}_\Gamma} \min_{Q_n \in \mathcal{Q}_\Lambda} E_{sp}(R_n, W, P_n, Q_n) \leq (1+\delta) E_{sp}(R - \delta) + \delta.$$

It follows that

$$E_U(R) \leq (1+\delta) E_{sp}(R - \delta) + \delta, \quad \text{for all } \delta > 0.$$

Since $E_{sp}(R)$ is finite and convex on $R > R_\infty(W)$, it is also continuous. Hence $E_U(R) \leq E_{sp}(R)$ for $R > R_\infty(W)$, which is the desired result.

Proof of Lemma 4: Let s_o be such that $l(s_o) = \min_s l(s)$. For any n and $s \neq s_o$, let $U'_S(s)$ be obtained by rounding *down* $U_S(s)$ to the nearest multiple of $1/n$; let $U'_S(s_o)$ be chosen so that U'_S sums to unity. Observe that $U'_S \in \mathcal{D}_n(\mathcal{S})$, $\| U'_S - U_S \| < 2|\mathcal{S}|/n$, and

$$\sum_s l(s) U'_S(s) \leq \sum_s l(s) U_S(s).$$

Now define for all $x \in \mathcal{X}$ and $s \in \mathcal{S}$

$$U'_{XS}(x,s) \triangleq \begin{cases} U'_S(s) U_{X|S}(x \mid s), & s \neq s_o \\ U_X(x) - \sum_{s' \neq s_o} U'_S(s') U_{X|S}(x \mid s'), & s = s_o. \end{cases}$$

By direct calculation, it is easily verified that U'_{XS} is nonnegative and has marginal distributions U_X and U'_S. Moreover

$$\| U'_{XS} - U_{XS} \| = \| U'_S - U_S \| \leq 2|\mathcal{S}|/n.$$

This completes the proof of Lemma 4.

ACKNOWLEDGMENT

The authors wish to thank the referees for several helpful suggestions, including a simplified proof of Theorem 2.

REFERENCES

[1] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Z. Wahrscheinihchkeitstheorie verw., Geb.*, vol. 44, pp. 159–175, 1978.

[2] ——, "Arbitrarily varying channels with states sequence known to the sender," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 621–629, Sept. 1986.

[3] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacities of certain channel classes under random coding," *Ann. Math. Statist.*, vol. 31, pp. 558–567, 1960.

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

[5] I. Csiszár and P. Narayan, "Arbitrarily varying channels with constrained inputs and states," *IEEE Trans. Inform. Theory*, vol. 34, pp. 27–34, Jan. 1988.

[6] ——, "The capacity of the arbitrarily varying channel revisited: Positivity, constraints," *IEEE Trans. Inform. Theory*, vol. 34, pp. 181–193, Mar. 1988.

[7] T. Ericson, "Exponential error bounds for random codes on the arbitrarily varying channel," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 42–48, Jan. 1985.

[8] D. G. Luenberger, *Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1984.

[9] I. G. Stiglitz, "Coding for a class of unknown channels," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 189–195, Apr. 1966.