Z. Wahrscheinlichkeitstheorie verw. Gebiete
57, 87-101 (1981)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete
© Springer-Verlag 1981

On the Capacity of the Arbitrarily Varying Channel for Maximum Probability of Error

I. Csiszár and J. Körner

Mathematical Institute of the Hungarian Academy of Sciences,
Reáltanoda u. 13-15, 1053 Budapest, Hungary

1. Introduction

A discrete memoryless channel (DMC) is determined by a stochastic matrix W with rows

$$W(\cdot\,|\,x)=\{W(y\,|\,x)\colon y\in\mathcal{Y}\},\qquad x\in\mathcal{X}.$$

Here $\mathcal{X}$ and $\mathcal{Y}$ are finite sets, called the input and output alphabet, respectively. For this DMC, denoted by $\{W\colon \mathcal{X}\to\mathcal{Y}\}$, the probability that a length-n input sequence $\mathbf{x}=x_1x_2\dots x_n\in\mathcal{X}^n$ yields on output sequence $\mathbf{y}=y_1y_2\dots y_n\in\mathcal{Y}^n$ is defined to be

$$W^n(\mathbf{y}\,|\,\mathbf{x})\triangleq\prod_{i=1}^{n}W(y_i\,|\,x_i).$$

An *arbitrarily varying channel* (AVC) with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and set of states $\mathcal{S}$ is, formally, a DMC $\{W\colon \mathcal{X}\times\mathcal{S}\to\mathcal{Y}\}$. It is understood that the components $x\in\mathcal{X}$ of the inputs (x,s) of this DMC are selected by the "sender" while the components $s\in\mathcal{S}$ are selected in an unpredictable manner by a malevolent "jammer". For a discussion and history cf. Ahlswede [2, 3], Wolfowitz [17] and Csiszár-Körner [7]. This paper deals with the case when $\mathcal{S}$ is finite. The results carry over to infinite $\mathcal{S}$ by a standard approximation argument, cf. e.g. the proof of Theorem 2.6.11 in [7].

Any length-n block code for channels with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ can be used for transmitting messages over an AVC $\{W\colon \mathcal{X}\times\mathcal{S}\to\mathcal{Y}\}$. Consider for each state sequence $\mathbf{s}\in\mathcal{S}^n$ the maximum and the average over the message set $\mathcal{M}$ of the probability of not decoding correctly the message $m\in\mathcal{M}$. Maximizing the rate $\frac{1}{n}\log|\mathcal{M}|$ under the constraint that this maximum resp. average probability of error be small, uniformly in $\mathbf{s}\in\mathcal{S}^n$, one arrives at the concept of capacity of the AVC for maximum resp. average probability of error (formal definitions will be given in Sect. 2). These capacities, denoted by C_m resp. C_a, are usually different (unlike for a DMC).

0044-3719/81/0057/0087/$03.00

C_a has been determined by Ahlswede [2] while C_m is unknown, in general. For AVC's with binary output alphabet, C_m was found by Ahlswede-Wolfowitz [4]. As shown by Ahlswede [1], determining C_m for an arbitrary AVC would include as a special case the solution of the famous graph-theoretic problem of determing the zero-error capacity of an arbitrary DMC. In the latter problem, raised by Shannon [15], remarkable progress has recently been made by Lovász [13] but the general solution still appears to be a long way ahead.

Recently Ahlswede [3] succeeded in determining C_m for a fairly large class of AVC's. An essential point of his proof was that into the random selection of the codeword set he included an expurgation that, combined with an ingenious choice of the decoder, enabled him to bound the maximum probability of error. The aim of this paper is to give a simpler proof which leads to a more general result. This will be done by the combinatorial approach introduced by Csiszár-Körner-Marton [8] (cf. also Csiszár-Körner [6]) which was further developed in Csiszár-Körner [7] and applied also in Körner-Sgarro [12], Csiszár [5], etc.

In [8] and [7], looking at constant composition codes with a fixed "balanced" codeword set and with various decoders, bounds on error probabilities were obtained by simple counting arguments applied separately to each of those joint types of sequences that contributed to the error event. Here we shall proceed similarly, but since the problem is more difficult, more careful bounding is needed in the selection of the codeword set (using large deviation bounds as did also Ahlswede [3]). The approach naturally suggests a candidate for a good decoder (significantly different from that of Ahlswede [3]). A crucial step will be to prove that the definition of this decoder is consistent.

Now we describe Ahlswede's theorem (in a formulation different from but equivalent to his) and our generalization. We introduce a graph with vertex set $\mathscr{X}$ to be called the *graph of* W or $G(W)$. In this graph x_1 and x_2 are connected by an edge – in symbols $x_1 \overset{W}{\sim} x_2$ – iff there exist distributions Q_1 and Q_2 on $\mathscr{S}$ such that

$$\sum_{s\in\mathscr{S}} W(y\,|\,x_1,s)\,Q_1(s)=\sum_{s\in\mathscr{S}} W(y\,|\,x_2,s)\,Q_2(s) \quad \text{for every } y\in\mathscr{Y}. \tag{1.1}$$

Notice that $G(W)$ is a graph with loops but without multiple edges. By a well-known result of Kiefer-Wolfowitz [11], $C_m>0$ iff $G(W)$ is not a complete graph. The condition of Ahlswede's theorem is that $G(W)$ should consist of isolated vertices. Under this condition he proves that

$$C_m=\max_P C(P) \tag{1.2}$$

where

$$C(P)\overset{\Delta}{=}\min I(X\wedge Y), \tag{1.3}$$

the minimum being taken for RV's X, Y taking values in $\mathscr{X}$ resp. $\mathscr{Y}$ such that X has distribution P and for some RV S with values in $\mathscr{S}$

$$\begin{aligned}\Pr\{Y=y\,|\,X=x,S=s\}=W(y\,|\,x,s) \quad &\text{whenever}\\ \Pr\{X=x,S=s\}>0.&\end{aligned} \tag{1.4}$$

We shall prove formula (1.2) under a considerably weaker condition. Denote

$$D(P) \overset{\Delta}{=} \min_{\Pr\{X \overset{W}{\sim} X'\}=1} I(X \wedge X') \tag{1.5}$$

where X and X' stand for RV's both having distribution P. Now we can state our

Main Result. *For every distribution P on $\mathscr{X}$, $\min(C(P), D(P))$ is an achievable rate for maximum probability of error. In particular, if there is a P_0 maximizing $C(P)$ such that $D(P_0) \geqq C(P_0)$ then* (1.2) *holds.*

If the graph of W consists of isolated vertices then clearly $D(P)=H(P)$ for every P. Thus the above result contains Ahlswede's theorem as a special case.

In the literature several variations of the capacity problem for an AVC have been considered. One of them, namely when "the states are known at the receiver" (cf. Kiefer-Wolfowitz [11], Stambler [16]) is actually a special case of the above problem. In fact, codes for an AVC $\{W: \mathscr{X} \times \mathscr{S} \to \mathscr{Y}\}$ with decoder depending on the state sequence $\mathbf{s}$ are the same as codes in the original sense for a new AVC $\{W': \mathscr{X} \times \mathscr{S} \to \mathscr{Y} \times \mathscr{S}\}$ where

$$W'(y, s' \mid x, s) \overset{\Delta}{=} \begin{cases} W(y \mid x, s) & \text{if } s'=s \\ 0 & \text{else.} \end{cases} \tag{1.6}$$

For average probability of error, the capacity of an AVC with states known at the receiver has been determined by Stambler [16]. For maximum probability of error, we shall determine this capacity for a wide class of AVC's, specializing or main result to AVC's of form (1.6). This will be done in Sect. 5.

2. Preliminaries

A *length-n block code* for channels with input alphabet $\mathscr{X}$ and output alphabet $\mathscr{Y}$ is a pair of mappings $f: \mathscr{M} \to \mathscr{X}^n$, $\varphi: \mathscr{Y}^n \to \mathscr{M}'$ where $\mathscr{M}' \supset \mathscr{M}$. The elements of $\mathscr{M}$ are called *messages*, their images under f are the *codewords*, f itself is the *encoder*, while φ is the *decoder*. If message $m \in \mathscr{M}$ is sent and a sequence $\mathbf{y} \in \mathscr{Y}^n$ is received, an error occurs whenever $\varphi(\mathbf{y}) \neq m$.

Accordingly, using the code (f, φ) on an AVC $\{W: \mathscr{X} \times \mathscr{S} \to \mathscr{Y}\}$, the probability of erroneous transmission of an $m \in \mathscr{M}$ for state sequence $\mathbf{s} \in \mathscr{S}^n$ is

$$e(m, \mathbf{s}) \overset{\Delta}{=} W^n(\{\mathbf{y}: \varphi(\mathbf{y}) \neq m\} \mid f(m), \mathbf{s}). \tag{2.1}$$

The *maximum probability of error* is defined as

$$e = e(f, \varphi) \overset{\Delta}{=} \max_{\mathbf{s} \in \mathscr{S}^n} \max_{m \in \mathscr{M}} e(m, \mathbf{s}).$$

Definition. A number $R \geqq 0$ is an *achievable rate* (for maximum probability of error) for a given AVC if to every $\varepsilon > 0$, $\delta > 0$ and every sufficiently large n

there exist length-n block codes (f, φ) with

$$\frac{1}{n}\log|\mathcal{M}|\geqq R-\delta, \qquad e(f,\varphi)\leqq\varepsilon.$$

The largest achievable rate is called the *capacity* (for maximum probability of error) of the AVC; it will be denoted by C_m.

Remark. The capacity for average probability of error (C_a) is defined similarly; its definition is omitted for we shall not need it.

Throughout this paper, we shall use the same basic notation as in [6] and [7], summarized below.

Distributions, Types

For RV's X, Y with values in finite sets $\mathcal{X}, \mathcal{Y}$, we denote by P_X resp. P_{XY} the *distribution* of X resp. the *joint distribution* of X and Y. The *conditional distribution* of Y given X is denoted by $P_{Y|X}$:

$$P_X(a)\overset{\Delta}{=}\Pr\{X=a\}, \qquad P_{XY}(a,b)\overset{\Delta}{=}\Pr\{X=a, Y=b\},$$
$$P_{Y|X}(b\,|\,a)\overset{\Delta}{=}\Pr\{Y=b\,|\,X=a\}=\frac{P_{XY}(a,b)}{P_X(a)};$$

$P_{Y|X}(b\,|\,a)$ is undefined if $P_X(a)=0$.

The *type* of a sequence $\mathbf{x}\in\mathcal{X}^n$ is the distribution $P_\mathbf{x}$ on $\mathcal{X}$ defined by letting $P_\mathbf{x}(a)$ be the relative frequency of the symbol $a\in\mathcal{X}$ in $\mathbf{x}$. The *joint type* of two sequences $\mathbf{x}\in\mathcal{X}^n$ and $\mathbf{y}\in\mathcal{Y}^n$ is the distribution $P_{\mathbf{x},\mathbf{y}}$ on $\mathcal{X}\times\mathcal{Y}$ defined similarly. The *conditional type* $P_{\mathbf{y}|\mathbf{x}}$ of $\mathbf{y}$ given $\mathbf{x}$ is defined by

$$P_{\mathbf{y}|\mathbf{x}}(b\,|\,a)=\frac{P_{\mathbf{x},\mathbf{y}}(a,b)}{P_\mathbf{x}(a)};$$

$P_{\mathbf{y}|\mathbf{x}}(b\,|\,a)$ is undefined if $P_\mathbf{x}(a)=0$.

A *stochastic matrix* V of which the rows are distributions on $\mathcal{Y}$, indexed by elements of $\mathcal{X}$, will be denoted by $V\colon\mathcal{X}\to\mathcal{Y}$. For a stochastic matrix $V\colon\mathcal{X}\to\mathcal{Y}$ the equality $P_{Y|X}=V$ or $P_{\mathbf{y}|\mathbf{x}}=V$ will mean that $P_{Y|X}(b\,|\,a)=V$ resp. $P_{\mathbf{y}|\mathbf{x}}(b\,|\,a)=V$ whenever the left-hand term is defined.

The set of sequences of type P in $\mathcal{X}^n$ will be denoted by $\mathcal{T}_P^n$ or simply $\mathcal{T}_P$. Of course, $\mathcal{T}_P^n\neq 0$ holds only for "a few" distributions on $\mathcal{X}$; the family of these distributions will be denoted by $\mathcal{P}_n(\mathcal{X})$. For a given $\mathbf{x}\in\mathcal{X}^n$ and stochastic matrix $V\colon\mathcal{X}\to\mathcal{Y}$, the set of sequences $\mathbf{y}\in\mathcal{Y}^n$ of conditional type $P_{\mathbf{y}|\mathbf{x}}=V$ will be called the *V-shell of* $\mathbf{x}$, denoted by $\mathcal{T}_V(\mathbf{x})$. The family of conditional types of sequences $\mathbf{y}\in\mathcal{Y}^n$ given an $\mathbf{x}\in\mathcal{X}^n$ depends on $\mathbf{x}$ only through its type $P_\mathbf{x}=P$; this family will be denoted by $\mathcal{V}_n(\mathcal{Y}\,|\,P)$. With some abuse of terminology, we consider $\mathcal{V}_n(\mathcal{Y}\,|\,P)$ as a family of stochastic matrices $V\colon\mathcal{X}\to\mathcal{Y}$ even if $P(a)=0$ for some $a\in\mathcal{X}$; we understand that the rows of these matrices indexed by elements $a\in\mathcal{X}$ with $P(a)=0$ are defined in some arbitrary but fixed way. Further, we write

$$\mathcal{V}_n(\mathcal{Y}\,|\,\mathcal{X})\overset{\Delta}{=}\bigcup_{P\in\mathcal{P}_n(\mathcal{X})}\mathcal{V}_n(\mathcal{Y}\,|\,P).$$

Information Quantities

If $P_X=P$, $P_{Y|X}=V$, the entropy $H(X)$, conditional entropy $H(Y|X)$ resp. mutual information $I(X\wedge Y)$ will also be denoted by $H(P)$, $H(V|P)$ resp. $I(P,V)$. We shall also use "non-probabilistic" information quantities (cf. Goppa [10]) defined for length-n sequences (rather than RV's). They will mean, by definition, the corresponding information quantities for RV's with joint distribution equal to the joint type of the sequences in question. E.g., $I(\mathbf{x}\wedge\mathbf{y}|\mathbf{s})$ means the conditional mutual information $I(X\wedge Y|S)$ for RV's X, Y, S having joint distribution $P_{XYS}\overset{\Delta}{=}P_{\mathbf{x},\mathbf{y},\mathbf{s}}$.

For two distributions P and Q on $\mathcal{X}$ resp. stochastic matrices $V:\mathcal{X}\to\mathcal{Y}$, $W:\mathcal{X}\to\mathcal{Y}$ we denote by $D(P\|Q)$ resp. $D(V\|W|P)$ the Kullback-Leibler informational divergence

$$D(P\|Q)\overset{\Delta}{=}\sum_{x\in\mathcal{X}}P(x)\log\frac{P(x)}{Q(x)}$$

resp. conditional informational divergence

$$D(V\|W|P)\overset{\Delta}{=}\sum_{x\in\mathcal{X}}P(x)D(V(\cdot|x)\|W(\cdot|x)).$$

A useful inequality of Pinsker [14] is

$$\sum_{x\in\mathcal{X}}|P(x)-Q(x)|\leqq c\sqrt{D(P\|Q)},\tag{2.3}$$

where c is an absolute constant.

We shall use the same elementary bounds as in [6]. For their proof cf. [9] or [7]:

$$|\mathcal{P}_n(\mathcal{X})|\leqq(n+1)^{|\mathcal{X}|}\tag{2.4}$$

$$|\mathcal{V}_n(\mathcal{Y}|\mathcal{X})|\leqq(n+1)^{|\mathcal{X}||\mathcal{Y}|}\tag{2.5}$$

$$(n+1)^{-|\mathcal{X}|}\exp\{nH(P)\}\leqq|\mathcal{T}_P|\leqq\exp\{nH(P)\}\quad\text{for}\quad P\in\mathcal{P}_n(\mathcal{X})\tag{2.6}$$

$$(n+1)^{-|\mathcal{X}||\mathcal{Y}|}\exp\{nH(V|P)\}\leqq|\mathcal{T}_V(\mathbf{x})|\leqq\exp\{nH(V|P)\}\tag{2.7}$$

for every $\mathbf{x}\in\mathcal{T}_P$, $V\in\mathcal{V}_n(\mathcal{Y}|P)$

$$Q^n(\mathbf{x})=\exp\{-n[D(P\|Q)+H(P)]\}\qquad\text{for }\mathbf{x}\in\mathcal{T}_P^n\tag{2.8}$$

$$W^n(\mathbf{y}|\mathbf{x})=\exp\{-n[D(V\|W|P)+H(V|P)]\}\qquad\text{for }\mathbf{x}\in\mathcal{T}_P^n,\quad\mathbf{y}\in\mathcal{T}_V(\mathbf{x})\tag{2.9}$$

Throughout the paper, $|\mathcal{A}|$ denotes the cardinality of the finite set $\mathcal{A}$. All exps and logs are to the base 2.

Finally, we shall use the notation

$|t|^+\overset{\Delta}{=}\max(0,t)$
$\lfloor t\rfloor\overset{\Delta}{=}$largest integer not exceeding t.

3. Statement and Discussion of the Results

Our capacity result will be a consequence of the following coding theorem for fixed composition codes.

Theorem 1. *Given an AVC* $\{W: \mathcal{X} \times \mathcal{S} \to \mathcal{Y}\}$ *and any* $\varepsilon > 0$, $\delta > 0$, *for every* $P \in \mathcal{P}_n(\mathcal{X})$ *and* $0 \leqq R \leqq \min(C(P), D(P))$ - *cf.* (1.3), (1.5) - *there exists a length-n block code* (f, φ) *with codewords of type P such that*

$$\frac{1}{n} \log |\mathcal{M}| \geqq R - \delta, \qquad e(f, \varphi) \leqq \varepsilon,$$

provided that $n \geqq n_0(W, \varepsilon, \delta)$.

The proof will be given in the next section. A corresponding converse-type result is the simple

Theorem 2. *Given an AVC* $\{W: \mathcal{X} \times \mathcal{S} \to \mathcal{Y}\}$ *and any* $\varepsilon > 0$, $\delta > 0$, *for every* $P \in \mathcal{P}_n(\mathcal{X})$ *and* $R \geqq C(P)$ *every length-n block code with codewords of type P such that*

$$\frac{1}{n} \log |\mathcal{M}| \geqq R + \delta$$

has

$$e(f, \varphi) \geqq \varepsilon$$

provided that $n \geqq n_0(|\mathcal{X}|, |\mathcal{Y}|, \varepsilon, \delta)$.

Proof. Consider the triple X, S, Y achieving the minimum in the Definition (1.3) of $C(P)$. Clearly, every code (f, φ) for the AVC $\{W: \mathcal{X} \times \mathcal{S} \to \mathcal{Y}\}$ is also a code for the DMC $\{P_{Y|X}: \mathcal{X} \to \mathcal{Y}\}$ with at most the same maximum probability of error. Hence the statement follows by the strong converse to the DMC coding theorem for fixed composition codes, cf. e.g. [7], Corollary 2.1.4.

Combining Theorems 1 and 2 we get

Theorem 3. *Suppose that the AVC* $\{W: \mathcal{X} \times \mathcal{S} \to \mathcal{Y}\}$ *has the property that for a distribution* P_0 *achieving* $\max C(P)$ *we have* $D(P_0) \geqq C(P_0)$. *Then this AVC has capacity*

$$C_m = \max_P C(P) = C(P_0).$$

Proof. One easily checks that $C(P)$ is a continuous function of P. Thus the direct part follows by applying Theorem 1 to distributions $P_n \in \mathcal{P}_n(\mathcal{X})$ with $P_n \to P_0$. The converse follows from Theorem 2 and (2.4).

Let us now discuss what improvements of these results might be hoped for. In this respect, a constant composition code analogue of the zero-error capacity problem seems relevant.

Recall that a subset of the vertex set of a graph G is called an *independent set* if no pair of its elements is connected by an edge. The maximum size of such sets is called the *independence number* of G, denoted by $\alpha(G)$. The n'th power of a graph G having vertex set $\mathscr{X}$ is the graph G^n with vertex set $\mathscr{X}^n$ in which $\mathbf{x}=x_1\dots x_n$ and $\mathbf{x}'=x'_1\dots x'_n$ are connected by an edge iff x_i and x'_i are connected in G, $i=1,2,\dots,n$ (with the understanding that each x is connected with itself). The *zero-error capacity* of a graph G is the (always existing) limit

$$C_0(G)\overset{\Delta}{=}\lim_{n\to\infty}\frac{1}{n}\log\alpha(G^n).$$

Clearly, the codewords of a length-n block code for an AVC $\{W\colon\mathscr{X}\times\mathscr{S}\to\mathscr{Y}\}$ with maximum probability of error less than $\frac{1}{2}$ form an independent set of $[G(W)]^n$ (which is the same graph as $G(W^n)$). Thus $C_0(G(W))$ is a trivial upper bound of C_m. Further, if all codewords are of type P, then

$$|\mathscr{M}|\leqq\alpha_n(P) \tag{3.1}$$

must hold, where $\alpha_n(P)$ stands for the independence number of the subgraph of $[G(W)]^n$ spanned by the subset $\mathscr{T}_P^n$ of $\mathscr{X}^n$. Introduce the notation

$$A(P)\overset{\Delta}{=}\sup\left(\limsup_{n\to\infty}\frac{1}{n}\log\alpha_n(P_n)\right) \tag{3.2}$$

where the supremum is taken for all sequences $P_n\to P$ with $P_n\in\mathscr{P}_n(\mathscr{X})$. Then inequality (3.1) enables us to sharpen Theorem 2 replacing the condition $R\geqq C(P)$ by $R\geqq\min(C(P),A(P))$. In particular, we certainly have

$$C_m\leqq\max_P\min(C(P),A(P)). \tag{3.3}$$

This implies that C_m can be strictly smaller than both $\max_P C(P)$ and $C_0(G(W))$.

Although (3.3) is not a computable bound (for no computable characterization of $A(P)$ is known), it would be interesting to decide whether it is sharp. The reason for our method giving only the weaker direct result

$$C_m\geqq\max_P\min(C(P),D(P))$$

consists in our applying random selection. In this way one gets an independent subset of $\mathscr{T}_P^n$ but of size $\exp\{nD(P)\}$ rather than $\exp\{nA(P)\}$.

4. Proof of Theorem 1

We start by a simple combinatorial lemma stating, intuitively, that for any prescribed number $R\geqq 0$ there exist $\exp(nR)$ not necessarily distinct sequences in $\mathscr{T}_P^n$ such that in no V-shall does their "local density" substantially exceed their "global density" in $\mathscr{T}_P^n$.

Lemma 1. *Given arbitrary finite sets $\mathscr{U},\mathscr{X}$, to every $R>0$, $n\geqq\max(|\mathscr{U}|,|\mathscr{X}|)$ and $P\in\mathscr{P}_n(\mathscr{X})$ there exist $M\overset{\Delta}{=}\lfloor\exp(nR)\rfloor$ not necessarily distinct sequences $\mathbf{x}_i\in\mathscr{T}_P^n$, $i=1,\dots,M$ such that for every $\mathbf{u}\in\mathscr{U}^n$ and $V\colon\mathscr{U}\to\mathscr{X}$ we have*

$$|\{i\colon\mathbf{x}_i\in\mathscr{T}_V(\mathbf{u})\}|\leqq 3(n+1)^{|\mathscr{X}|}\exp\{n|R-I(P_{\mathbf{u}},V)|^+\}. \tag{4.1}$$

Proof. We shall show that if $M \overset{\Delta}{=} \lfloor \exp(nR) \rfloor$ elements of $\mathcal{T}_P$ are chosen at random then with positive probability the inequalities (4.1) hold simultaneosly.

Formally, let $Z_1, \dots, Z_M$ be independent RV's taking values in $\mathcal{T}_P$ and uniformly distributed on $\mathcal{T}_P$. If V is not in $\mathscr{V}_n(\mathscr{X} \mid \mathscr{U})$, inequality (4.1) trivially holds. Fixing now an $\mathbf{u} \in \mathscr{U}^n$ and $V \in \mathscr{V}_n(\mathscr{X} \mid \mathscr{U})$, consider the independent and identically distributed RV's

$$\chi_i \overset{\Delta}{=} \begin{cases} 1 & \text{if } Z_i \in \mathcal{T}_V(\mathbf{u}) \\ 0 & \text{else.} \end{cases}$$

Since the number of possible pairs $(\mathbf{u}, V)$ is less than $|\mathscr{U}|^n (n+1)^{|\mathscr{U}||\mathscr{X}|}$, it suffices to show that

$$\Pr\left\{ \sum_{i=1}^{M} \chi_i > 3(n+1)^{|\mathscr{X}|} \exp\left[n |R - I(P_{\mathbf{u}}, V)|^+\right] \right\} \\ < |\mathscr{U}|^{-n} (n+1)^{-|\mathscr{U}||\mathscr{X}|}. \tag{4.2}$$

This can be done by a standard Bernstein argument. Introducing the notation

$$a(n) \overset{\Delta}{=} 3(n+1)^{|\mathscr{X}|} \exp\left[n |R - I(P_{\mathbf{u}}, V)|^+\right], \tag{4.3}$$

the left-hand side of (4.2) equals

$$\Pr\left\{ \exp \sum_{i=1}^{M} \chi_i > \exp a(n) \right\},$$

which, by Markov's inequality, is upper bounded by

$$\varepsilon(n) \overset{\Delta}{=} (E \exp \chi_1)^M \exp(-a(n)). \tag{4.4}$$

Here (recalling that exp's are to the base 2)

$$\begin{aligned} E \exp \chi_1 &= \Pr\{\chi_1 = 0\} + 2 \Pr\{\chi_1 = 1\} \\ &= 1 + \Pr\{\chi_1 = 1\} = 1 + \frac{|\mathcal{T}_V(\mathbf{u})|}{|\mathcal{T}_P|} \\ &\leqq 1 + (n+1)^{|\mathscr{X}|} \exp\{-n I(P_{\mathbf{u}}, V)\} \end{aligned}$$

where the last inequality follows from (2.6), (2.7). Hence, using the inequality $(1+t) \leqq e^t = \exp(t \log e)$ and the definition of M, we get

$$\begin{aligned} (E \exp \chi_1)^M &\leqq \exp\{M (n+1)^{|\mathscr{X}|} \exp[-n I(P_{\mathbf{u}}, V)] \log e\} \\ &\leqq \exp\{(n+1)^{|\mathscr{X}|} \cdot \log e \cdot \exp[n(R - I(P_{\mathbf{u}}, V))]\}. \end{aligned}$$

Substituting this and (4.3) into (4.4) gives

$$\varepsilon(n) \leqq \exp\{-(3 - \log e)(n+1)^{|\mathscr{X}|}\}.$$

Since $\varepsilon(n)$ is an upper bound of the left-hand side of (4.2), the last bound establishes (4.2) for n sufficiently large. A simple calculation shows that already for

$n \geqq \max[|\mathcal{U}|, |\mathcal{X}|]$ we have

$$\exp\{-(3-\log e)(n+1)^{|\mathcal{X}|}\} < |\mathcal{U}|^{-n}(n+1)^{-|\mathcal{U}||\mathcal{X}|}.$$

What we shall actually need in our code construction is the following consequence of Lemma 1. It will play the same role for an AVC as did for a DMC our Corollary of Lemma 2 in [6].

Lemma 2. *Given finite sets $\mathcal{X}, \mathcal{S}$, to every $\delta > 0$, $R > 0$, $n \geqq n_0(|\mathcal{X}|, |\mathcal{S}|, \delta)$ and $P \in \mathcal{P}_n(\mathcal{X})$ with $H(P) \geqq R$ there exists a subset $\mathcal{C}$ of $\mathcal{T}_P \subset \mathcal{X}^n$ such that*

(i) $\frac{1}{n}\log|\mathcal{C}| \geqq R - \delta$

(ii) *for every $\mathbf{x} \in \mathcal{X}^n$, $\mathbf{s} \in \mathcal{S}^n$, $V: \mathcal{X} \times \mathcal{S} \to \mathcal{X}$ we have*

$$|\mathcal{T}_V(\mathbf{x}, \mathbf{s}) \cap \mathcal{C}| \leqq 3(n+1)^{|\mathcal{X}|} \exp\{n|R - I(P_{\mathbf{x},\mathbf{s}}, V)|^+\}$$

(iii) *for every pair of elements $\mathbf{x} \neq \tilde{\mathbf{x}}$ of $\mathcal{C}$*

$$I(\mathbf{x} \wedge \tilde{\mathbf{x}}) < R.$$

Proof. Applying Lemma 1 with $\mathcal{U} \overset{\Delta}{=} (\mathcal{X} \times \mathcal{S}) \cup \mathcal{X}$ we get $M = \lfloor \exp(nR) \rfloor$ (not necessarily distinct) sequences $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M$ in $\mathcal{T}_P$ such that for every $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{s} \in \mathcal{S}^n$

$$|\{i: \mathbf{x}_i \in \mathcal{T}_V(\mathbf{x}, \mathbf{s})\}| \leqq 3(n+1)^{|\mathcal{X}|} \exp\{n|R - I(P_{\mathbf{x},\mathbf{s}}, V)|^+\} \quad \text{for every } V: \mathcal{X} \times \mathcal{S} \to \mathcal{X}, \tag{4.5}$$

$$|\{i: \mathbf{x}_i \in \mathcal{T}_{\bar{V}}(\mathbf{x})\}| \leqq 3(n+1)^{|\mathcal{X}|} \exp\{n|R - I(P_{\mathbf{x}}, \bar{V})|^+\} \quad \text{for every } \bar{V}: \mathcal{X} \to \mathcal{X}. \tag{4.6}$$

It suffices to show that at least $M \exp\left(-\frac{n\delta}{2}\right)$ sequences $\mathbf{x}_{i_j}$ can be selected out of these $\mathbf{x}_i$'s such that

$$I(\mathbf{x}_{i_j} \wedge \mathbf{x}_{i_k}) < R \qquad \text{for } i_j \neq i_k; \tag{4.7}$$

in fact, this also implies (by the assumption $R \leqq H(P)$) that the sequences $\mathbf{x}_{i_j}$ are all distinct. To do this, notice that (4.6) implies for every $l \leqq M$

$$|\{i: \mathbf{x}_i \in \mathcal{T}_{\bar{V}}(\mathbf{x}_l)\}| \leqq 3(n+1)^{|\mathcal{X}|} \qquad \text{if } I(P, \bar{V}) \geqq R.$$

It follows by (2.5) that for every $l \leqq M$

$$|\{i: I(\mathbf{x}_i \wedge \mathbf{x}_l) \geqq R\}| \leqq 3(n+1)^{|\mathcal{X}|}(n+1)^{|\mathcal{X}|^2}. \tag{4.8}$$

Now sequences $\mathbf{x}_{i_j}$ meeting (4.7) can be chosen successively: Suppose that $\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_m}$ satisfying (4.7) have already been selected. Then, on account of (4.8), a further $\mathbf{x}_{i_{m+1}}$ can also be selected without violating (4.7) unless

$$m \cdot 3(n+1)^{|\mathcal{X}|}(n+1)^{|\mathcal{X}|^2} > M.$$

To prove Theorem 1, we shall use the set $\mathscr{C}$ of Lemma 2 as codeword set, letting the encoder f be the identity mapping on $\mathscr{C}$. Our decoder φ will map each $\mathbf{y}\in\mathscr{Y}^n$ into an $\mathbf{x}\in\mathscr{C}$ such that for some $\mathbf{s}\in\mathscr{S}^n$

$$D(P_{\mathbf{y}|\mathbf{x},\mathbf{s}}\|W|P_{\mathbf{x},\mathbf{s}})<\eta \tag{4.9}$$

and

$$I(\hat{\mathbf{x}}\wedge\mathbf{y}\,|\,\mathbf{x},\mathbf{s})<\eta \qquad \text{for every } \hat{\mathbf{x}}\in\mathscr{C} \text{ which for some } \hat{\mathbf{s}}\in\mathscr{S}^n \text{ satisfies (4.9)} \tag{4.10}$$

provided that such an $\mathbf{x}\in\mathscr{C}$ exists; here $\eta>0$ will be specified later.

Notice that (4.9) is a kind of "joint typicality" condition; for a DMC every decoder satisfying this condition would do. For an AVC the decoder should be chosen more carefully. In our proof it is exactly condition (4.10) which will be needed to make the maximum probability of error small. The content of the next lemma is that (4.9) and (4.10) unambigously define $\varphi(\mathbf{y})$ whenever an $\mathbf{x}\in\mathscr{C}$ satisfying these condition does exist.

Lemma 3. *Given an AVC $\{W:\mathscr{X}\times\mathscr{S}\to\mathscr{Y}\}$, to every $\delta>0$ there exists an $\eta>0$ such that for any subset $\mathscr{C}$ of $\mathscr{T}_P\subset\mathscr{X}^n$ satisfying (iii) of Lemma 2 with $R\leqq D(P)-\delta$, to every $\mathbf{y}\in\mathscr{Y}^n$ at most one $\mathbf{x}\in\mathscr{C}$ can be found with the properties* (4.9), (4.10).

Proof. We have to prove that if two pairs $(\mathbf{x},\mathbf{s})$ and $(\tilde{\mathbf{x}},\tilde{\mathbf{s}})$ in $\mathscr{C}\times\mathscr{S}^n$ both satisfy (4.9), (4.10), then necessarily $\mathbf{x}=\tilde{\mathbf{x}}$.

In fact, suppose that

$$D(P_{\mathbf{y}|\mathbf{x},\mathbf{s}}\,\|\,W\,|\,P_{\mathbf{x},\mathbf{s}})<\eta, \qquad D(P_{\mathbf{y}|\tilde{\mathbf{x}},\tilde{\mathbf{s}}}\,\|\,W\,|\,P_{\tilde{\mathbf{x}},\tilde{\mathbf{s}}})<\eta, \tag{4.11}$$

$$I(\tilde{\mathbf{x}}\wedge\mathbf{y}\,|\,\mathbf{x},\mathbf{s})<\eta, \qquad I(\mathbf{x}\wedge\mathbf{y}\,|\,\tilde{\mathbf{x}},\tilde{\mathbf{s}})<\eta. \tag{4.12}$$

Let $X, S, \tilde{X}, \tilde{S}, Y$ denote RV's having joint distribution equal to the joint type of $(\mathbf{x},\mathbf{s},\tilde{\mathbf{x}},\tilde{\mathbf{s}},\mathbf{y})$. Then the first inequality of (4.11) resp. (4.12) means, by definition, that

$$\sum_{x,s,y} P_{XSY}(x,s,y)\log\frac{P_{Y|X,S}(y|x,s)}{W(y\,|\,x,s)}<\eta$$

resp.

$$I(\tilde{X}\wedge Y|X,S)=\sum_{x,s,\tilde{x},y} P_{XS\tilde{X}Y}(x,s,\tilde{x},y)\log\frac{P_{Y|XS\tilde{X}}(y\,|\,x,s,\tilde{x})}{P_{Y|XS}(y\,|\,x,s)}<\eta.$$

Adding these two inequalities we get

$$\sum_{x,s,\tilde{x},y} P_{XS\tilde{X}Y}(x,s,\tilde{x},y)\log\frac{P_{XS\tilde{X}Y}(x,s,\tilde{x},y)}{W(y\,|\,x,s)\,P_{XS\tilde{X}}(x,s,\tilde{x})}<2\eta.$$

Here the left-hand side is the informational divergence of two distributions on $\mathscr{X}\times\mathscr{S}\times\mathscr{X}\times\mathscr{Y}$. Projecting them to $\mathscr{X}\times\mathscr{X}\times\mathscr{Y}$ the divergence does not increase, and thus

$$D(P_{X\tilde{X}Y}\,\|\,P_{X\tilde{X}}\times V)<2\eta, \tag{4.13}$$

where $V\colon \mathcal{X}\times\mathcal{X}\to\mathcal{Y}$ is the stochastic matrix defined by

$$V(y|x,\tilde{x}) \overset{\Delta}{=} \sum_{s\in\mathcal{S}} W(y\,|\,x,s)\,P_{S|X\tilde{X}}(s\,|\,x,\tilde{x}).$$

(Recall the notation (2.2); if $P_{X\tilde{X}}(x,\tilde{x})=0$ then $V(y\,|\,x,\tilde{x})$ and $\tilde{V}(y\,|\,x,\tilde{x})$ below can be arbitrary.) Similarly, we get from the remaining inequalities of (4.11) and (4.12) that

$$D(P_{X\tilde{X}Y}\,\|\,P_{X\tilde{X}}\times\tilde{V})<2\eta \tag{4.14}$$

where $\tilde{V}\colon \mathcal{X}\times\mathcal{X}\to\mathcal{Y}$ is defined by

$$\tilde{V}(y\,|\,x,\tilde{x})=\sum_{\tilde{s}\in\mathcal{S}} W(y\,|\,\tilde{x},\tilde{s})\,P_{\tilde{S}|X\tilde{X}}(\tilde{s}\,|\,x,\tilde{x}).$$

Using (2.3), the inequalities (4.13), (4.14) give rise to

$$\sum_{x,\tilde{x}} P_{X\tilde{X}}(x,\tilde{x})\sum_{y}|V(y|x,\tilde{x})-\tilde{V}(y|x,\tilde{x})|\leqq 2c\sqrt{2\eta}. \tag{4.15}$$

Lemma 3 will be proved if we show that (4.15) can not hold for $\mathbf{x}\neq\tilde{\mathbf{x}}$ provided that $\eta=\eta(W,\delta)$ is sufficiently small. To this end, notice that as $\mathcal{C}\subset\mathcal{T}_P$ meets condition (iii) of Lemma 2 with $R\leqq D(P)-\delta$, we have

$$I(X\wedge\tilde{X})=I(\mathbf{x}\wedge\tilde{\mathbf{x}})<R\leqq D(P)-\delta=\min_{\substack{\Pr\{X\overset{W}{}X'\}=1\\ P_X=P_{X'}=P}} I(X\wedge\tilde{X}')-\delta.$$

This implies that

$$1-\Pr\{X\overset{W}{\sim}\tilde{X}\}>\varepsilon_1 \tag{4.16}$$

for some $\varepsilon_1=\varepsilon_1(G(W),\delta)>0$. Further, if x and $\tilde{x}$ are not connected by an edge of $G(W)$, then by the definition (1.1) of $G(W)$

$$\min_{Q_1,Q_2}\sum_{y\in\mathcal{Y}}\Big|\sum_{s\in\mathcal{S}} W(y\,|\,x,s)\,Q_1(s)-\sum_{s\in\mathcal{S}} W(y\,|\,\tilde{x},\tilde{s})\,Q_2(\tilde{s})\Big|>0,$$

where Q_1 and Q_2 range over the distributions on $\mathcal{S}$. Denoting by $\varepsilon_2=\varepsilon_2(W)$ the minimum of these minima for pairs $(x,\tilde{x})$ not connected by and edge in $G(W)$, it follows that if $P_{X\tilde{X}}(x,\tilde{x})>0$ and $x\overset{W}{\sim}\tilde{x}$ does not hold then

$$\sum_{y\in\mathcal{Y}}|V(y\,|\,x,\tilde{x})-V(y\,|\,x,\tilde{x})|\geqq\varepsilon_2.$$

This and (4.16) yield the desired contradiction with (4.15) if η is chosen sufficiently small.

Now we turn to the

Proof of Theorem 1. Consider an AVC $\{W\colon\mathcal{X}\times\mathcal{S}\to\mathcal{Y}\}$, fix some $\varepsilon>0,\delta>0$, $n\geqq n_0$, $P\in\mathcal{P}_n(\mathcal{X})$ as in Theorem 1, and suppose that

$$R\leqq\min(C(P),D(P))-\delta. \tag{4.17}$$

This is not a real restriction compared with the condition $R \leqq \min(C(P), D(P))$ of Theorem 1. In fact, if the assertion of Theorem 1 is true under the hypothesis (4.17) then, applying if for $R-\frac{\delta}{2}$ and $\frac{\delta}{2}$ in the role of R and δ, the same assertion follows for every $R \leqq \min(C(P), D(P))$. Define a length-n block code (f, φ) as follows: Let the message set be the set $\mathscr{C}$ of Lemma 2 and the encoder f be the identity mapping on $\mathscr{C}$. Let the decoder φ map each $\mathbf{y} \in \mathscr{Y}^n$ into an $\mathbf{x} \in \mathscr{C}$ which for some $\mathbf{s} \in \mathscr{S}^n$ satisfies (4.9), (4.10). If no such $\mathbf{s} \in \mathscr{S}^n$ and $\mathbf{x} \in \mathscr{X}^n$ exist, $\varphi(\mathbf{y})$ may be arbitrary. Here, the threshold η of (4.9), (4.10) is chosen to satisfy Lemma 3, $\eta < \frac{\delta}{2}$, and the additional condition that for every triple of RV's $\hat{X}\hat{S}Y$ and a RV $\hat{Y}$ with $P_{\hat{Y}|\hat{S}\hat{X}} \overset{\Delta}{=} W$ we have

$$I(\hat{X} \wedge Y) \geqq (\hat{X} \wedge \hat{Y}) - \frac{\delta}{2} \quad \text{if} \quad D(P_{Y|\hat{X}\hat{S}} \| W | P_{\hat{X}\hat{S}}) \leqq \eta. \tag{4.18}$$

(Such a choice is possible by inequality (2.3) and the uniform continuity of $I(X \wedge Y)$ as a function of P_{XY}.) We claim that this code (f, φ) has maximum probability of error less than ε, i.e.,

$$e(\mathbf{x}, \mathbf{s}) \leqq \varepsilon \quad \text{for every} \quad \mathbf{x} \in \mathscr{C}, \mathbf{s} \in \mathscr{S}^n \tag{4.19}$$

(cf. (2.1)), provided that $n_0 = n_0(W, \varepsilon, \delta)$ is sufficiently large.

Fix an $\mathbf{x} \in \mathscr{C}$ and $\mathbf{s} \in \mathscr{S}^n$, and consider the subsets of $\mathscr{Y}^n$

$$\mathscr{A} \overset{\Delta}{=} \{\mathbf{y} : D(P_{\mathbf{y}|\mathbf{x},\mathbf{s}} \| W | P_{\mathbf{x},\mathbf{s}}) \geqq \eta\}$$

$$\mathscr{B} \overset{\Delta}{=} \{\mathbf{y} : I(\hat{\mathbf{x}} \wedge \mathbf{y} | \mathbf{x}, \mathbf{s}) \geqq \eta \quad \text{for some} \quad \hat{\mathbf{x}} \in \mathscr{C}(\mathbf{y})\}$$

where

$$\mathscr{C}(\mathbf{y}) \overset{\Delta}{=} \{\hat{\mathbf{x}} : \hat{\mathbf{x}} \in \mathscr{C}, D(P_{\mathbf{y}|\hat{\mathbf{x}},\hat{\mathbf{s}}} \| W | P_{\hat{\mathbf{x}},\hat{\mathbf{s}}}) < \eta \quad \text{for some} \quad \hat{\mathbf{s}} \in \mathscr{S}^n\}.$$

By the definition of our decoder φ an error $\varphi(\mathbf{y}) \neq \mathbf{x}$ can occur only if $\mathbf{y} \in \mathscr{A} \cup \mathscr{B}$, thus

$$e(\mathbf{x}, \mathbf{s}) \leqq W^n(\mathscr{A} | \mathbf{x}, \mathbf{s}) + W^n(\mathscr{B} | \mathbf{x}, \mathbf{s}). \tag{4.20}$$

By (2.7), (2.9) and (2.5) we have

$$W^n(\mathscr{A} | \mathbf{x}, \mathbf{s}) \leqq (n+1)^{|\mathscr{X}||S||\mathscr{Y}|} \exp(-n\delta). \tag{4.21}$$

In order to bound $W^n(\mathscr{B} | \mathbf{x}, \mathbf{s})$, we consider separately the intersections of $\mathscr{B}$ with the various V-shells of $(\mathbf{x}, \mathbf{s})$:

$$W^n(\mathscr{B} | \mathbf{x}, \mathbf{s}) = \sum_{V \in \mathscr{V}_n(\mathscr{Y} | P_{\mathbf{x},\mathbf{s}})} W^n(\mathscr{B} \cap \mathscr{T}_V(\mathbf{x}, \mathbf{s}) | \mathbf{x}, \mathbf{s}). \tag{4.22}$$

For notational convenience, joint types of length-n sequences will be represented as joint distributions of RV's. The cardinality of $\mathscr{B} \cap \mathscr{T}_V(\mathbf{x}, \mathbf{s})$ is bounded from above by the number of pairs $(\hat{\mathbf{x}}, \mathbf{y}) \in \mathscr{C} \times \mathscr{Y}^n$ such that $P_{\mathbf{x},\mathbf{s},\hat{\mathbf{x}},\mathbf{y}} = P_{XS\hat{X}Y}$ for

RV's X, S, Y with joint distribution given by

$$P_{XSY}(x,s,y) \overset{\Delta}{=} P_{\mathbf{x},\mathbf{s}}(x,s)\,V(y\,|\,x,s) \tag{4.23}$$

and some RV $\hat{X}$ such that $P_{\hat{X}} = P$ and

$$I(\hat{X} \wedge Y\,|\,XS) \geqq \eta, \tag{4.24}$$

$$D(P_{Y|\hat{X}\hat{S}} \,\|\, W\,|\,P_{\hat{X}\hat{S}}) \leqq \eta \quad \text{for some RV } \hat{S}. \tag{4.25}$$

By Lemma 2 (ii), the number of sequences $\hat{\mathbf{x}} \in \mathscr{C}$ with $P_{\mathbf{x},\mathbf{s},\hat{\mathbf{x}}} = P_{XS\hat{X}}$ is at most

$$3(n+1)^{|\mathscr{X}|} \exp\{n\,|R - I(XS \wedge \hat{X})|^+\},$$

while for every such $\hat{\mathbf{x}}$ the number of sequences $\mathbf{y} \in \mathscr{Y}^n$ with $P_{\mathbf{x},\mathbf{s},\hat{\mathbf{x}},\mathbf{y}} = P_{XS\hat{X}Y}$ is, by (2.7), at most

$$\exp\{nH(Y\,|\,XS\hat{X})\} = \exp\{n[H(Y\,|\,XS) - I(\hat{X} \wedge Y\,|\,XS)]\}.$$

Combining the last two bounds and using (2.4) we get

$$\begin{aligned} |\mathscr{B} \cap \mathscr{T}_V(\mathbf{x},\mathbf{s}) \leqq (n+1)^{|\mathscr{X}|^2\,|\mathscr{S}|\,|\mathscr{Y}|}\,3(n+1)^{|\mathscr{X}|} \exp\{n[H(Y\,|\,XS) \\ -\min\,(I(\hat{X} \wedge Y\,|\,XS) - |R - I(XS \wedge \hat{X})|^+)]\}, \end{aligned} \tag{4.26}$$

where P_{XSY} is given by (4.23) and the minimum is taken for all RV's $\hat{X}$ satisfying $P_{\hat{X}} = P$ and (4.24), (4.25).

Now we show that this last minimum can not be less than η. To verify this, on occount of (4.24) it suffices to consider RV's $\hat{X}$ with $I(XS \wedge \hat{X}) < R$. Then

$$\begin{aligned} &I(\hat{X} \wedge Y\,|\,XS) - |R - I(XS \wedge \hat{X})|^+ \\ &= I(\hat{X} \wedge Y\,|\,XS) - R + I(XS \wedge \hat{X}) = I(\hat{X} \wedge YXS) - R \\ &\geqq I(\hat{X} \wedge Y) - R \geqq C(P) - \frac{\delta}{2} - R, \end{aligned}$$

where the last step follows from (4.25) by assumption (4.18) and the definition of $C(P)$, cf. (1.3), (1.4). Since $R \leqq C(P) - \delta$ and $\eta \leqq \frac{\delta}{2}$ by assumption, our claim that the minimum in (4.26) is not less than η is herewith established. Using this result and (2.9), the bound (4.26) yields

$$W^n(\mathscr{B} \cap \mathscr{T}_V(\mathbf{x},\mathbf{s})\,|\,\mathbf{x},\mathbf{s}) \leqq 3(n+1)^{|\mathscr{X}|^2\,|\mathscr{S}|\,|\mathscr{Y}| + |\mathscr{X}|} \exp(-n\eta). \tag{4.27}$$

Comparing (4.20), (4.21), (4.22) and (4.27) completes the proof of (4.19).

Remark. We have actually proved that under the conditions of Theorem 1 the maximum probability of error of the optimal length-n block code tends to zero exponentially as $n \to \infty$. Counting more carefully, we could have got tighter exponential bounds quite easily. Still, it remains to be seen whether these improved bounds are tight in any interesting case.

5. States Known at the Receiver

As pointed out in the Introduction, our results have immediate implications for an AVC whose states are known at the receiver. Formally, a length-n block code for an AVC $\{W: \mathscr{X} \times \mathscr{S} \to \mathscr{Y}\}$ with states known at the receiver is a pair of mappings $f: \mathscr{M} \to \mathscr{X}^n$, $\varphi: \mathscr{Y}^n \times \mathscr{S}^n \to \mathscr{M}'$ where $\mathscr{M}' \supset \mathscr{M}$. The remaining definitions are the same as in Section 2 except that the analogue of $e(m, \mathbf{s})$ of formula (2.1) is now defined by

$$e'(m, \mathbf{s}) \overset{\Delta}{=} W^n(\{y: \varphi(\mathbf{y}, \mathbf{s}) \neq m\} \mid f(m), \mathbf{s}).$$

Obviously, the above (f, φ) and $e'(m, \mathbf{s})$ are the same as a length-n block code in the original sense and the corresponding $e(m, \mathbf{s})$ for the new AVC $\{W': \mathscr{X} \times \mathscr{S} \to \mathscr{Y} \times \mathscr{S}\}$ defined by (1.6). Hence, applying our results to this new AVC, we get their analogues for the AVC $\{W: \mathscr{X} \times \mathscr{S} \to \mathscr{Y}\}$ with states known at the receiver. The relevant quantities of formulas (1.3) and (1.5) now have the analogues

$$C'(P) \overset{\Delta}{=} \min_{\substack{P_X = P \\ P_{Y|XS} = W}} I(X \wedge YS),$$

$$D'(P) \overset{\Delta}{=} \min_{\substack{\Pr\{X \underset{\sim}{W}' X'\} = 1 \\ P_X = P_{X'} = P}} I(X \wedge X').$$

The definition of $D'(P)$ becomes simple upon observing that in the graph of W' two elements x_1 and x_2 of $\mathscr{X}$ are connected by an edge iff there exists an $s \in \mathscr{S}$ such that

$$W(y \mid x_1, s) = W(y \mid x_2, s) \quad \text{for every} \quad y \in \mathscr{Y}.$$

In particular, the analogue of Theorem 3 is

Theorem 4. *Suppose that the AVC $\{W: \mathscr{X} \times \mathscr{S} \to \mathscr{Y}\}$ has the property that for a distribution P_0 achieving* $\max C'(P)$ *we have $D'(P_0) \geqq C'(P_0)$. Then the capacity for maximum probability of error C'_m of this AVC with states known at the receiver is*

$$C'_m = \max_P C'(P) = C'(P_0).$$

Acknowledgement. The authors are indebted to R. Ahlswede for kindly sending them the manuscript of his paper [3], which stimulated their interest in the problem.

References

1. Ahlswede, R.: A note on the existence of the weak capacity for channels with arbitrarily varying channel probability functions and its relation to Shannon's zero error capacity. Annals Math. Statist. **41**, 1027–1033 (1970)
2. Ahlswede, R.: Elimination of correlation in random codes for arbitrarily varying channels. Z. Wahrscheinlichkeitstheorie verw. Gebiete **44**, 159–175 (1978)

3. Ahlswede, R.: A method of coding and an application to arbitrarily varying channels. Journal of Combinatorics, Information and System Sciences **5**, 10–35 (1980)
4. Ahlswede, R., Wolfowitz, J.: The capacity of a channel with arbitrarily varying cpf. and binary output alphabet. Z. Wahrscheinlichkeitstheorie verw. Gebiete **15**, 186–194 (1970)
5. Csiszár, I.: Joint source-channel error exponent. Problems of Control and Information Theory, **9**, 315–328 (1980)
6. Csiszár, I., Körner, J.: Graph decomposition: a new key to coding theorems. IEEE Trans. Information Theory **27**, 5–12 (1981)
7. Csiszár, I., Körner, J.: Information Theory: Coding Theorems for Discrete Memoryless Systems. New York: Academic Press, 1981
8. Csiszár, I., Körner, J., Marton, K.: A new look at the error exponent of discrete memoryless channels. Preprint. Presented at the International Symposium on Information Theory, Cornell Univ., Ithaca, N.Y., 1977
9. Dueck, G., Körner, J.: Reliabily function of a discrete memoryless channel at rates above capacity. IEEE Trans. Information Theory **25**, 82–85 (1979)
10. Goppa, V.D.: Nonprobabilistic mutual information without memory (in Russian). Problems of Control and Information Theory **4**, 97–102 (1975)
11. Kiefer, J., Wolfowitz, J.: Channels with arbitrarily varying channel probability functions. Information and Control **5**, 44–54 (1962)
12. Körner, J., Sgarro, A.: Universally attainable error exponents for broadcast channels with degraded message sets. IEEE Trans. Information Theory **26**, 670–679 (1980)
13. Lovász, L.: On the Shannon capacity of a graph. IEEE Trans. Information Theory **25**, 1–7 (1979)
14. Pinsker, M.S.: Information and Information Stability of Random Variables and Processes (in Russian). Problemy Peredâci Informacii Vol **7**, AN SSSR, Moscow, 1960. English Translation: San Francisco: Holden-Day, 1964
15. Shannon, C.E.: The zero error capacity of a noisy channel. IRE Transactions Information Theory **2**, 8–19 (1956)
16. Stambler, S.Z.: Shannon theorem for a full class of channels with state known at the output (in Russian). Problemy Peredači Informacii **14**, no. 4, 3–12 (1975)
17. Wolfowitz, J.: Coding Theorems of Information Theory, 3rd edition. Berlin-Heidelberg-New York: Springer 1978

Received December 10, 1979; in revised form January 12, 1981