Z. Wahrscheinlichkeitstheorie verw. Gebiete
44, 159–175 (1978)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Elimination of Correlation in Random Codes for Arbitrarily Varying Channels

Rudolf Ahlswede

Fakultät für Mathematik der Universität Bielefeld, Universitätsstraße 1, D-4800 Bielefeld 1

Summary. The author determines for arbitrarily varying channels

a) the average error capacity and
b) the maximal error capacity in case of randomized encoding.

A formula for the *average* error capacity in case of randomized encoding was announced several years ago by Dobrushin ([3]). Under a mild regularity condition this formula turns out to be valid and follows as consequence from either a) or b).

1. The Channel Model and the Coding Problems

Since several articles have been written on this subject we begin right away with the mathematical notions needed. The reader not sufficiently familiar with the concepts used will find some heuristic explanations in the last Section.

Let $\mathfrak{X}$ and $\mathfrak{Y}$ be finite sets, which serve as input and output alphabets of the channels described below. Let S be an arbitrary set, and let $\mathfrak{C}=\{w(\cdot|\cdot|s): s\in S\}$ be a set of stochastic $|\mathfrak{X}|\times|\mathfrak{Y}|$-matrices. For every $s^n=(s_1,\ldots,s_n)\in S^n=\prod_1^n S$ we define transmission probabilities $P(\cdot|\cdot|s^n)$ by

$$P(y^n|x^n|s^n)=\prod_{t=1}^{n} w(y_t|x_t|s_t)$$

$$\text{for all } x^n=(x_1,\ldots,x_n)\in\mathfrak{X}^n=\prod_1^n\mathfrak{X},$$

$$y^n=y_1,\ldots,y_n)\in\mathfrak{Y}^n=\prod_1^n\mathfrak{Y}, \quad \text{and all } n=1,2,\ldots,. \tag{1.1}$$

Set $\mathfrak{C}^n=\{P(\cdot|\cdot|s^n): s^n\in S^n\}$. We call the sequence $(\mathfrak{C}^n)_{n=1}^{\infty}$ an arbitrarily varying channel (AVC) and denote it by $\mathfrak{A}$.

0044-3719/78/0044/0159/$03.40

For every fixed $x\in\mathfrak{X}$, $\mathfrak{T}(x)$ denotes the closed convex hull of the set $\{w(\cdot|x|s): s\in S\}$ of probability distributions (PD) on $\mathfrak{Y}$.

The set of matrices

$$\overline{\mathfrak{C}}=\{(w(y|x))_{x\in\mathfrak{X},\,y\in\mathfrak{Y}}: w(\cdot|x)\in\mathfrak{T}(x), x\in\mathfrak{X}\}$$

is called the row-convex closure of the set $\mathfrak{C}$. (1.2)

The closed convex hull of $\mathfrak{C}$ is denoted by $\bar{\mathfrak{C}}$.

$\mathfrak{A}$ is called a standard arbitrarily varying channel (SAVC), if $\mathfrak{C}=\bar{\mathfrak{C}}$. Suppose now that sender and receiver want to communicate over an AVC without knowing which $P(\cdot|\cdot|s^n)$ will govern the transmission of any word sent. This leads to a coding problem which heavily depends on the code concept and also on the notion of error probability (maximal or average) used. Here we consider three code concepts and we list them in an order corresponding to their increasing generality.

1. A K_1-code (N,n) is a system $\{(u_i,D_i): 1\leqq i\leqq N\}$, where $u_i\in\mathfrak{X}^n$, $D_i\subset\mathfrak{Y}^n$ for $i=1,2,\ldots,N$ and $D_i\cap D_j=\emptyset$ for $i\neq j$.

2. If we now permit randomisation in the encoding we come to the following concept:

A K_2-code $(n,N,\mathbf{r})$ is a system $\{(r_i,D_i): 1\leqq i\leqq N\}$, where the D_i's have the same properties as before and the r_i's are probability distributions on $\mathfrak{X}^n$.

3. Finally, a K_3-code (n,N,μ,Γ) is a collection of K_1-codes (n,N) $\{(u_i^\gamma,D_i^\gamma): 1\leqq i\leqq N\}_{\gamma\in\Gamma}$, where Γ is finite, together with a PD μ on Γ.

Those codes were introduced in [35] under the name "random codes". It is also explained there in which sense K_2-codes are special K_3-codes. For a given channel $\mathfrak{A}$ one can compute the maximal and the average error probabilities for each of the codes described. We denote those error probabilities by λ_i, $\bar{\lambda}_i (i=1,2,3)$. They are given by the following expressions

$$\lambda_1=\sup_{s^n\in S^n}\max_i P(D_i^c|u_i|s^n),$$

$$\lambda_2=\sup_{s^n\in S^n}\max_i\sum_{x^n\in\mathfrak{X}^n}P(D_i^c|x^n|s^n)\,r_i(x^n),$$

$$\lambda_3=\sup_{s^n\in S^n}\max_i\sum_{\gamma\in\Gamma}P((D_i^\gamma)^c|u_i^\gamma|s^n)\,\mu(\gamma),$$

$$\bar{\lambda}_1=\sup_{s^n\in S^n}\frac{1}{N}\sum_{i=1}^{N}P(D_i^c|u_i|s^n),$$

$$\bar{\lambda}_2=\sup_{s^n\in S^n}\frac{1}{N}\sum_{i=1}^{N}\sum_{x^n\in\mathfrak{X}^n}P(D_i^c|x^n|s^n)\,r_i(x_n),$$

$$\bar{\lambda}_3=\sup_{s^n\in S^n}\frac{1}{N}\sum_{i=1}^{N}\sum_{\gamma\in\Gamma}P((D_i^\gamma)^c|u_i^\gamma|s^n)\,\mu(\gamma).$$

A number C_1 us called the capacity of the channel $\mathfrak{A}$, if for any $\varepsilon>0$, any λ, $0<\lambda<1$, and for all sufficiently large n there exists a K_1-code $(n,\exp\{(C_1-\varepsilon)n\})$

with maximal error probability λ_1 not exceeding λ and there does not exist a code $(n, \exp\{(C_1+\varepsilon)n\})$ meeting the bound λ on the error probability.

Analogously one defines for K_2-codes and for K_3-codes capacities C_2 and C_3. If we use average errors instead of maximal errors we denote the corresponding capacities by $\bar{C}_i; i=1,2,3$.

One of the principal tasks in Shannon's Information Theory is to show that a (specified) capacity exists for a channel in question and to find a formula for it, which is such that it can be used for a numerical evaluation.

In their pioneering paper [1] Blackwell, Breiman, and Thomasian obtained for AVC's such a formula for $\bar{C}_3$. However, a serious drawback to the use of K_3-codes is that they require *correlated* randomisation between encoding and decoding, that is the outcome of the random experiment (Γ, μ) has to be made known to both communicators. If there are no further channels available (which is realistic to assume), the only way to achieve this is that the sender, before transmitting any message i, chooses a K_1-code at random, communicates the result of his random experiment to the receiver, and then transmits the message according to the code sected. This procedure is repeated at each message. If the size of Γ could be kept so small that sending those additional messages would not cause essential loss in capacity, one could agree to such a procedure. However, no such bound on $|\Gamma|$ had been given until now.

The reasons mentioned led the authors of [2] to investigate K_1-codes with maximal error and the author of [3] to study K_2-codes with average error. C_1 was determined for binary output AVC's in [7]. For general output alphabets no solution exists until now. The problem seems to be very hard. It includes the famous zero-error capacity problem as a special case ([8, 9]).

For list codes of relatively small list size and in the presence of complete feedback K_1-type maximal error capacities are known ([11, 12]).

In this paper we determine all the other capacities: $\bar{C}_1, C_2, \bar{C}_2, C_3$.

2. Auxiliary Results

Denote by $\mathfrak{P}$ the set of all PD's on $\mathfrak{X}$. For any $p \in \mathfrak{P}$ and any $w \in \bar{\mathfrak{C}}$ define the mutual information

$$R(p, w) = H(q) - \sum_{x \in \mathfrak{X}} p(x) H(w(\cdot|x)),$$

where $q = p \cdot w$ and H denotes the entropy. (2.1)

The following two quantities shall play an important role in the sequel.

$$C_R = \min_{w \in \mathfrak{C}} \max_{p \in \mathfrak{P}} R(p, w), \tag{2.2}$$

$$C_D = \min_{w \in \bar{\mathfrak{C}}} \max_{p \in \mathfrak{P}} R(p, w). \tag{2.3}$$

These definitions are meaningful, because $\mathfrak{P}$, $\mathfrak{C}$ and $\bar{\mathfrak{C}}$ are compact in the norm topology and $R(p, w)$ is continuous in both variables. We shall make essentially use of the following result of [1]:

Random Code Theorem

For the AVC $\mathfrak{A}$

$$\bar{C}_3 = C_R. \tag{2.4}$$

Moreover, any rate smaller than C_R is achievable with exponentially small error probability.

Actually, the authors of [1] proved $\bar{C}_3 \geqq C_R$ and then a weak converse, but the strong converse follows immediately because the coding problems for $(\mathfrak{C}^n)_{n=1,2,\ldots}$ and $(\bar{\mathfrak{C}}^n)_{n=1,2,\ldots}$ are equivalent and the latter contains a DMC with capacity C_R (*cf.* [6]).

Notice that we allowed $\mathfrak{C}$ to contain infinitely many matrices. Our proofs in later Sections are given first for the case $|\mathfrak{C}| < \infty$. Then we remove this restriction with the help of the following Approximation lemma and its Corollary.

Approximation Lemma ([14])

Let $A \geqq a\,b^2$, where $a = |\mathfrak{X}|$, $b = |\mathfrak{Y}|$. There exists a set $\mathfrak{C}_A \subset \mathfrak{C}$, $|\mathfrak{C}_A| \leqq (A+1)^{a \cdot b}$, such that for every $w \in \mathfrak{C}$ there exists a $w' \in \mathfrak{C}_A$ such that for all $x \in \mathfrak{X}$ and all $y \in \mathfrak{Y}$

$$|w(y|x) - w'(y|x)| \leqq b\,A^{-1} \tag{2.5}$$

and

$$w(y|x) \leqq \exp\{2b^2 A^{-1}\}\, w'(y|x). \tag{2.6}$$

For our purpose we have to choose A as function of n, a choice $A(n) = n^2$ will be appropriate.

Write $\mathfrak{C}_A \subset \mathfrak{C} = \{w(\cdot|\cdot|s)\colon s \in S\}$ as $\mathfrak{C}_A = \{w(\cdot|\cdot|s)\colon s \in S'\}$ with $|S'| \leqq (A+1)^{ab}$ and denote the approximating matrix of $w(\cdot|\cdot|s)$ by $w(\cdot|\cdot|s')$. By (2.6) we have for $s^n = (s_1, \ldots, s_n) \in S^n$ and $s'^n = (s'_1, \ldots, s'_n) \in S'^n$

$$P(y^n|x^n|s^n) \leqq P(y^n|x^n|s'^n) \cdot \exp\{2b^2 n A^{-1}\}$$
$$\text{for all } x^n \in \mathfrak{X}^n, \quad y^n \in \mathfrak{Y}^n \tag{2.7}$$

and with the choice $A(n) = n^2$ for large n

$$P(y^n|x^n|s^n) \leqq P(y^n|x^n|s'^n) \cdot 2 \tag{2.8}$$

and therefore for the complement of a decoding set and all x^n

$$P(D^c|x^n|s^n) \leqq 2\,P(D^c|x^n|s'^n). \tag{2.9}$$

Thus we obtain the

Corollary. *For the* AVC $\mathfrak{A} = (\mathfrak{C}^n)_{n=1,2,\ldots}$ *there exists a subset* $\mathfrak{C}^{*n}$ *of* $\mathfrak{C}^n$, *such that*

$$|\mathfrak{C}^{*n}| \leqq (n^2+1)^{abn} \tag{2.10}$$

and every K_i*-code* $(i = 1, 2, 3)$ *for* $\mathfrak{C}^{*n}$ *is a* K_i*-code for* $\mathfrak{C}^n$ *with at most twice the (maximal or average) error probability for n large enough.*

The following result of [2] is useful in investigating the positiveness of capacities.

Seperation Lemma. *For the* AVC $\mathfrak{A}$ C_1 *is positive exactly when the following condition holds:*

(S) *there exist* $x, x' \in \mathfrak{X}$ *such that* $\mathfrak{T}(x) \cap \mathfrak{T}(x') = \emptyset$.

We need also a last and quite elementary result

Innerproduct Lemma. *Let* $\alpha = (\alpha_1, \ldots, \alpha_R)$ *and* $\beta = (\beta_1, \ldots, \beta_R)$ *be two vectors with* $0 \leqq \alpha_i, \beta_i \leqq 1$ *for* $i = 1, 2, \ldots, R$, *which satisfy*

$$\frac{1}{R} \sum_{j=1}^{R} \beta_j \geqq 1 - \varepsilon, \quad \frac{1}{R} \sum_{j=1}^{R} \alpha_j \geqq 1 - \varepsilon, \quad 0 < \varepsilon < 1,$$

then

$$\frac{1}{R} \sum_{j=1}^{R} \alpha_j \beta_j \geqq 1 - 2\varepsilon.$$

Proof. The worst case occurs for instance if the α_j's are monotonically increasing and the β_j's are monotonically decreasing, and if maximally many α_j's and β_j's take the value 1. An easy calculation leads to the inequality.

3. The Results

Before reading the Theorems below it might be useful to be aware of the following two examples.

Example 1. Let $|\mathfrak{X}| = 2$, $|\mathfrak{Y}| = 3$ and let $\mathfrak{C} = \{w_1, w_2\}$, where

$$w_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad w_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

It was shown in [1] that in this case $\bar{C}_1 = 0$ and $C_R > 0$. The conclusion can be drawn that $\bar{C}_1$ is much smaller than $\bar{C}_3$ and that $\bar{C}_1$ and C_R are not related.

Our Theorem 1 classifies Example 1 as pathological in so far as $\bar{C}_1 = 0$. It also shows that in case $\bar{C}_1 = 0$ Example 1 reflects the typical behaviour.

Example 2. $|\mathfrak{X}| = |\mathfrak{Y}| = 3$, $\mathfrak{C} = \{w_1, w_2\}$, where

$$w_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad w_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

In this case condition (S) does not hold and hence $C_1 = 0$. However by randomisation over input letters 1 and 2 with probabilities $(\frac{1}{2}, \frac{1}{2})$ one can produce

$$w_1^* = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad w_2 = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix}$$

and for those matrices (S) holds. Therefore $C_2 > 0$.

Theorem 1. *For the* AVC $\mathfrak{A}$

$$\bar{C}_1 = C_R, \quad \text{if } \bar{C}_1 > 0.$$

Theorem 2. *For the* AVC $\mathfrak{A}$

a) $C_2 = C_R$, if $C_2 > 0$,

b) $C_3 = \bar{C}_3 = C_R$.

Theorem 3. *For the* AVC $\mathfrak{A}$ *always*

a) $C_2 = \bar{C}_1 = \bar{C}_2$.

In particular

b) $C_2 > 0 \Leftrightarrow \bar{C}_1 > 0 \Leftrightarrow \bar{C}_2 > 0$.

Notice that condition (S) gives a useful sufficient condition for all capacities to be positive, because they are all not smaller than C_1. *With some effort it should be possible to find a necessary and sufficient condition for* $\bar{C}_1$ *or* C_2 *to be positive. The next lemma completely settles this problem for SAVC's. Since those channels provide the more robust model (see Section 8) we felt little urge to study the question any further.*

Strong Separation Lemma. *For the SAVC* $\bar{\bar{\mathfrak{A}}}$

$$C_2 > 0 (\Leftrightarrow \bar{C}_1 > 0 \Leftrightarrow \bar{C}_2 > 0) \Leftrightarrow \textit{condition (S) holds.}$$

Theorem 4 (Solution to Dobrushin's problem [3]). *For the SAVC* $\bar{\bar{\mathfrak{A}}}$

$$C_2 = \bar{C}_1 = \bar{C}_2 = \begin{cases} C_D, & \textit{if (S) holds} \\ 0, & \textit{if (S) does not hold.} \end{cases}$$

Remarks. 1. In [3] the formula $\bar{C}_2 = C_D$ was stated for the AVC $\mathfrak{A}$. An example of [6] shows that $C_2 > C_D$ can occur. This is excluded here by the assumption $\mathfrak{A} = \bar{\bar{\mathfrak{A}}}$. A further regularity condition for $\bar{C} = C_D$ to hold is condition (S), because it can happen that C_D is positive and (S) does not hold. $C_D = 0$ for $\bar{\bar{\mathfrak{A}}}$ is equivalent to $\bigcap_{x \in \mathfrak{X}} \mathfrak{T}(x) \neq \emptyset$.

2. In the noticable paper [13] the formula $\bar{C}_1 = C_R$ is proved for a certain class of AVC's $\mathfrak{A}$, which is characterized by quite complicated conditions. Considering the mathematical effort it is somewhat disappointing that a simple channel such as

Example 3 ([13]), $\mathfrak{X} = \mathfrak{Y} = \{1, 2\}$,

$$\mathfrak{C} = \left\{ \begin{pmatrix} 1 & 0 \\ 0.4 & 0.6 \end{pmatrix}, \begin{pmatrix} 0.6 & 0.4 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}.$$

Already by [7]

$$\bar{C}_1 = \max_p R(p, \bar{w}), \quad \text{where } \bar{w} = \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix}.$$

cannot be treated by the approach choosen.

4. Proof of Theoem 1

We assume first that $|S| < \infty$. Our proof is such that this assumption can then easily be removed with the help of the corollary to the Approximation lemma.

We describe now first the key ideas of the proof. From a given K_3-code (n, N, μ, Γ) with exponentially small error probability $\bar{\lambda}_3 = e^{-\varepsilon n}$ one selects a relatively small number $R = n^2$ of K_1-codes by independent repetitions of the random experiment (Γ, μ). Averaging over those K_1-codes leads to a new K_3-code (n, N, μ', Γ') with $|\Gamma'| = R = n^2$ relatively small. Allowing now a constant (rather than exponentially small) error probability one can guarantee that the randomly produced (n, N, μ', Γ') fails to meet this new error bound for a fixed $P(\cdot|\cdot|s^n)$ with a super exponentially small probability only. Since $|\mathfrak{C}|^n$ grows only exponentially there is a realisation (n, N, μ^*, Γ^*), $|\Gamma^*| = R$, with a small error probability for all $P(\cdot|\cdot|s^n)$.

Now we make use of the assumption $\bar{C}_1 > 0$, which guarantees the existence of a K_1-code $(f(n), n^2)$ with $f(n) = o(n)$ and small average error probability. By concatenation of this code with the (n, N, μ^*, Γ^*)-code just produced we obtain a K_1-code with the desired properties.

We give now the formal proof. By the Random Code Theorem there exists an (n, N, μ, Γ)-code $\{(u_i^\gamma, D_i^\gamma)\colon 1 \leqq i \leqq N;\ \gamma \in \Gamma\}$ with $\bar{\lambda}_3 \leqq e^{-\varepsilon n}$ and $N \geqq \exp\{(\bar{C}_3 - \delta) n\}$, $\delta > 0$, for ε sufficiently small and all large n. Consider a sequence $Z_1, \ldots, Z_R$ of independent RV's with values in Γ and distribution $Pr(Z_i = \gamma) = \mu(\gamma)$. Define for a fixed s^n and $j = 1, \ldots, R$ $T_j(s^n) = T_j$ by

$$T_j = 1 - \frac{1}{N} \sum_{i=1}^{N} P(D_i^{Z_j}, u_i^{Z_j} | s^n). \tag{4.1}$$

The RV's $T_1, \ldots, T_R$ are again independent and identically distributed with expected value $ET_j \leqq \bar{\lambda}_3$. For any $\lambda, 0 < \lambda < 1$, any $\alpha > 0$ and $T = \sum_{j=1}^{R} T_j$ Bernstein's trick gives

$$\Pr(T \geqq \lambda R) \leqq E \exp\{\alpha(T - R\lambda)\} = e^{-\alpha R \lambda} \prod_{j=1}^{R} E \exp\{\alpha T_j\}. \tag{4.2}$$

Now, since $0 \leqq T_j^k \leqq T_j \leqq 1$

$$E \exp\{\alpha T_j\} = 1 + \alpha ET_j + \frac{\alpha^2 ET_j^2}{2!} + \cdots$$

$$\leqq 1 + \sum_{k=1}^{\infty} \frac{\alpha^k}{k!} ET_j \leqq 1 + \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} \bar{\lambda}_3$$

and therefore also

$$\Pr(T \geqq \lambda R) \leqq e^{-\alpha R \lambda}(1+e^{\alpha}\bar{\lambda}_3)^R, \tag{4.3}$$

Choosing $R=n^2$, $\alpha=2$, and using $\bar{\lambda}_3=e^{-\varepsilon n}$ we get

$$\Pr(T \geqq \lambda R) \leqq e^{-2n^2 R}\,(1+e^2\,e^{-\varepsilon n})^{n^2}. \tag{4.4}$$

For large enough n

$$(1+e^2\,e^{-\varepsilon n}) \leqq e^{\lambda}$$

and therefore

$$\Pr\left(\frac{1}{n^2}\sum_{j=1}^{n^2} T_j \geqq \lambda\right) \leqq e^{-\lambda n^2}. \tag{4.5}$$

That is, the "random random code"

$$\{(u_i^{Z_r}, D_i^{Z_r}): 1 \leqq i \leqq N;\ r=1,\ldots,R\}, \qquad \Gamma^*=\{1,\ldots,R\},\ \mu^*(r)=\frac{1}{R}$$

for $r \in \Gamma^*$ has an average error probability

$$\lambda^*(Z_1,\ldots,Z_R)=\frac{1}{R}\sum_{j=1}^{R} T_j$$

which exceeds λ with a super exponentially small probability $\leqq e^{-\lambda n^2}$. The same is true for all $s^n \in S^n$ and since $|S|^n$ grows exponentially only, there exists a K_3-code

$$\{(u_i^{\gamma_r}, D_i^{\gamma_r}): 1 \leqq i \leqq N;\ r=1,\ldots,n^2\},\ \mu^* \tag{4.6}$$

with average error probability $\leqq \lambda$ for the AVC $\mathfrak{A}$.

In case $|\mathfrak{C}|=\infty$ we apply the Corollary to the Approximation lemma. Since $|\mathfrak{C}^{*n}| \leqq (n^2+1)^{abn}$, also in this case

$$(n^2+1)^{abn}\,e^{-\lambda n^2}<1 \qquad \text{for } n \text{ large}, \tag{4.7}$$

and therefore again there exists a code as described in (4.6) with average error probability $\leqq 2\lambda$.

Since $\bar{C}_1>0$ there exists a $(f(n), n^2)$-code $\{(v_j, A_j): 1 \leqq j \leqq n^2\}$ with average error probability $\leqq \lambda$ and $f(n)=o(n)$.

We compute now the average error probability of the $(f(n)+n,\ N \cdot R)$-code

$$\{(v_j \times u_i^j,\ A_j \times D_i^j): 1 \leqq i \leqq N;\ 1 \leqq j \leqq R\}$$

obtained by concatenation. Set $m=f(n)+n$. For any $s^m=(s_1,\ldots,s_m)$ define $t_1=(s_1,\ldots,s_{f(n)})$ and $t_2=(s_{f(n)+1},\ldots,s_m)$. Consider

$$\frac{1}{R}\sum_{j=1}^{R}\frac{1}{N}\sum_{i=1}^{N}P(A_j\times D_i^j|v_j u_i^j|s^m)$$
$$=\frac{1}{R}\sum_{j=1}^{R}\frac{1}{N}\sum_{i=1}^{N}P(A_j|v_j|t_1)\,P(D_i^j|u_i^j|t_2)$$
$$=\frac{1}{R}\sum_{j=1}^{R}P(A_j|v_j|t_1)\cdot\frac{1}{N}\sum_{i=1}^{N}P(D_i^j|u_i^j|t_2).$$

With the abbreviations

$$\alpha_j=p(A_j|v_j|t_1),\qquad \beta_j=\frac{1}{N}\sum_{i=1}^{N}P(D_i^j|u_i^j|t_2)$$

this becomes $\frac{1}{R}\sum_{j=1}^{R}\alpha_j\beta_j$. By the Innerproduct Lemma

$$\frac{1}{R}\sum_{j=1}^{R}\alpha_j\beta_j\geqq 1-2\lambda \tag{4.8}$$

and the concatenated code has therefore an error probability less than 2λ.

Since for any $\eta>0$ $(f(n)+n)^{-1}\log NR\geqq(1-\eta)n^{-1}\log N$ for n large enough, the desired result follows.

5. Proof of Theorem 2

The argument which led to the K_3-code (n,N,μ^*,Γ^*) in the previous proof will now be refined such that we even achieve small *maximal* error probability. In particular we thus show that always

$$C_3=\bar{C}_3. \tag{5.1}$$

Then again we use a concatenation argument, which is in this case even simpler.

The codes are selected at random as before via the RV's $Z_1,\ldots,Z_R$, but now we add a random permutation on the indices for each code choosen, that is, we allow all possible assignments of the set of messages $\{1,\ldots,N\}$ to the codewords. Thus we get RV's $V_j=(Z_j,P_j)$; $j=1,\ldots,R$; which are independent and indentically distributed.

Consider the RV's

$$S_{ji}=1-P(D^{Z_j}_{P_j(i)}|u^{Z_j}_{P_j(i)}|s^n),$$
$$j=1,\ldots,R;\qquad i=1,\ldots,N.$$

They are identically distributed and have expected values $ES_{ji}\leqq\bar{\lambda}_3$. Moreover, for any fixed i the RV's $S_{1i},\ldots,S_{Ri}$ are independent. Let $\lambda>\bar{\lambda}_3$ be constant.

Now, for any fixed i,

$$\Pr\left(\sum_{j=1}^{R} S_{ji} \geqq R\lambda\right) \leqq E\exp\left\{\alpha\left(\sum_{j=1}^{R} S_{ji} - R\lambda\right)\right\} = e^{-\alpha R\lambda} \sum_{j=1}^{R} E\exp\{\alpha S_{ji}\}$$

$$\leqq e^{-\alpha R\lambda}(1+e^{\alpha}\cdot ES_{11})^{R}$$

and hence (as in the previous proof) $\Pr\left(\frac{1}{R}\sum_{j=1}^{R} S_{ji} \geqq \lambda\right) \leqq e^{-\lambda n^2}$ for all large n.

Therefore also

$$\Pr\left(\max_{i=1,\ldots,N} \frac{1}{R}\sum_{j=1}^{R} S_{ji} \geqq \lambda\right) \leqq N e^{-\lambda n^2}. \tag{5.2}$$

Since $N = e^{(C_3-\delta)n}$ grows exponentially only, we again obtain a superexponential bound and therefore we can pass from one s^n to all $s^n \in S^n$ if $|S| < \infty$. Again there exists a code as described in (4.6) which now has even *maximal* error probability $\leqq \lambda$ for the AVC $\mathfrak{A}$. Since also

$$(n^2+1)^{abn} N e^{-\lambda n^2} < 1 \quad \text{for } n \text{ large} \tag{5.3}$$

the extension to the case $|S| = \infty$ is immediate by the Corollary. Since $C_2 > 0$ there exists a K_2-code $\{(p_j, A_j): 1 \leqq j \leqq n^2\}$ with error probability $\lambda_2 \leqq \lambda$ and block length $f(n) = o(n)$.

Define now a K_2-code as follows: Denote by $\delta_{u_i^j}$ the PD concentrated at u_i^j and define for $i = 1, \ldots, N$

$$q_i = \frac{1}{R}\sum_{j=1}^{R} p_j \times \delta_{u_i^j} \quad \text{as PD on } \mathfrak{X}^{f(n)} \times \mathfrak{X}^n. \tag{5.4}$$

Define decoding sets B_i by

$$B_i = \bigcup_{j=1}^{R} A_j \times D_i^j. \tag{5.5}$$

Then for fixed i, $s^m = t_1 \times t_2$ (as before), and $f = f(n)$

$$\frac{1}{R}\sum_{j=1}^{R}\sum_{x^f} P(B_i | x^f u_i^j | s^m)\, p_j(x^f)$$

$$\geqq \frac{1}{R}\sum_{j=1}^{R}\sum_{x^f} P(A_j \times D_i^j | x^f u_i^j | s^m)\, p_j(x^f)$$

$$= \frac{1}{R}\sum_{j=1}^{R}\sum_{x^f} P(A_j | x^f | t_1)\, p_j(x^f) \cdot P(D_i^j | u_i^j | t_2)$$

$$\geqq \frac{1}{R}\sum_{j=1}^{R} (1-\lambda) P(D_i^j | u_i^j | t_2) \geqq (1-\lambda)^2 \geqq 1 - 2\lambda.$$

The Theorem follows now, because also $\lim_{n\to\infty} n(n+f(n))^{-1}=1$.

Remarks. 1. Actually it would have been sufficient for the proof to require that $\bar{C}_2>0$. As in the proof of Theorem 1 one could again use the Innerproduct lemma. However, in the light of Theorem 3 this observation is of no consequence.

2. In the special case where the code we start with is already of type K_1 the proof shows that $C_2\geqq\bar{C}_1$ if $C_2>0$ and by 1. even $C_2\geqq\bar{C}_1$ if $\bar{C}_2>0$. Since $\bar{C}_2\geqq\bar{C}_1$, therefore always $C_2\geqq\bar{C}_1$.

6. Proof of Theorem 3

By Remark 2 in Section 5 we know already that $C_2\geqq\bar{C}_1$. This can also be derived as follows. It follows from Theorem 2 and the Random Code Theorem that

$$C_2\geqq\bar{C}_1 \quad \text{if} \quad C_2>0. \tag{6.1}$$

If $\bar{C}_1>0$, consider a code $\{(u_i,D_i)\colon 1\leqq i\leqq 2N\}$ with $\bar{\lambda}_1<\frac{1}{4}$. Define a K_2-code by

$$p_1=\frac{1}{N}\sum_{i=1}^{N}\partial_{u_i}, \qquad p_2=\sum_{i=n+1}^{2N}\partial_{u_i},$$

$$A_1=\bigcup_{i=1}^{N}D_i, \qquad A_2=\bigcup_{i=N+1}^{2N}D_i.$$

Then for $i=1,2$ and all s^n

$$\sum_{x^n}P(A_i|x^n|s^n)\,p_i(x^n)\geqq 1-2\bar{\lambda}_1>\tfrac{1}{2}. \tag{6.2}$$

This implies condition (S) for the "random letters" p_1 and p_2 and therefore $C_2>0$. We show now that

$$\bar{C}_2>0\Rightarrow\bar{C}_1>0. \tag{6.3}$$

Since obviously $\bar{C}_2\geqq\bar{C}_1$ this will complete the proof of b). Statement a) follows then from b) and Theorems 1, 2.

Suppose we have a K_2-code $(n,N,\mathbf{r})$ for $\mathfrak{A}$ with error probability $\bar{\lambda}_2$:

$$\{(r_i,D_i)\colon 1\leqq i\leqq N\}.$$

Let $U_1,\ldots,U_N$ be independent RV's with

$$\Pr(U_i=x^n)=r_i(x^n). \tag{6.4}$$

For fixed s^n denote

$$P(D_i^c|U_i|s^n) \quad \text{by} \quad V_i.$$

Then for $c, \alpha > 0$; $\lambda \geqq \bar{\lambda}_2$ a constant:

$$\Pr\left(\sum_{i=1}^{N} V_i > c\lambda N\right) \leqq e^{-c\lambda N\alpha} \sum_{j=1}^{N} E \exp\{\alpha V_i\}$$

$$\leqq e^{-c\lambda N\alpha} \prod_{j=1}^{N} (1 + e^{\alpha} EV_i), \tag{6.5}$$

because $EV_i \leqq EV_i^k$ for $k = 1, 2, \ldots$.

The inequality of the arithmetic and geometric means yields

$$\prod_{j=1}^{N} (1 + e^{\alpha} EV_i) \leqq \left(1 + e^{\alpha} \frac{1}{N} \sum_{i=1}^{N} EV_i\right)^N \leqq (1 + e^{\alpha} \bar{\lambda}_2)^N \leqq (1 + e^{\alpha} \lambda)^N. \tag{6.6}$$

Choose $\alpha = 1$ and c such that

$$e^{c\frac{\lambda}{2}} \geqq (1 + e \cdot \lambda),$$

then

$$\Pr\left(\frac{1}{N} \sum_{i=1}^{N} V_i > c\lambda\right) \leqq e^{-\frac{c}{2}\lambda N}. \tag{6.7}$$

Since N can grow exponentially in n, this probability is double exponentially small and we can complete the proof in the usual way (see Section 4 or 5).

Actually, we have thus proved even directly $\bar{C}_1 = \bar{C}_2$.

7. Proof of the Strong Seperation Lemma

Since $C_2 \geqq C_1$, by the Seperation lemma it suffices to show that $C_2 > 0$ implies condition (S). We make use of an elementary identity for Minkowski sums of sets in a linear space $\mathfrak{L}$.

Let $A_1, \ldots, A_I \subset \mathfrak{L}$, then

$$\operatorname{conv}\left(\sum_{i=1}^{I} A_i\right) = \sum_{i=1}^{I} \operatorname{conv}(A_i), \tag{7.1}$$

where "conv" denotes the convex hull operation.

If $C_2 > 0$, then by the Seperation lemma there exists an n and two PD's p^n and q^n on $\mathfrak{X}^n$ such that

$$\operatorname{conv}\Big(\sum_{x^n \in \mathfrak{X}^n} p^n(x^n) \mathfrak{T}(x^n)\Big) \cap \operatorname{conv}\Big(\sum_{x^n \in \mathfrak{X}^n} q^n(x^n) \mathfrak{T}(x^n) = \emptyset, \tag{7.2}$$

where $\mathfrak{T}(x^n) = \mathfrak{T}(x_1) \times \cdots \times \mathfrak{T}(x^n)$, $x^n = (x_1, \ldots, x_n)$.

If (S) were violated, then

$$\mathfrak{T}(x) \cap \mathfrak{T}(x') \neq \emptyset \quad \text{for all } x, x' \in \mathfrak{X}$$

and therefore also

$$\mathfrak{T}(x^n)\cap\mathfrak{T}(x'^n)\neq\emptyset \quad \text{for all } x^n, x'^n\in\mathfrak{X}^n. \tag{7.3}$$

We show now that (7.3) contradicts (7.2) and thus complete the proof. Abstractly speaking we are in the following situation:

We are given sets $A_1,\ldots,A_I$ in a linear space $\mathfrak{L}$, which satisfy

$$A_i\cap A_j\neq\emptyset \quad \text{for all } i,j,$$

and for suitable convex combinations

$$\operatorname{conv}\left(\sum_{i=1}^{I}\alpha_i A_i\right)\cap\operatorname{conv}\left(\sum_{j=1}^{I}\beta_j A_j\right)=\emptyset.$$

By (7.1) this equation is equivalent to

$$\sum_{i=1}^{I}\alpha_i\operatorname{conv}(A_i)\cap\sum_{j=1}^{I}\beta_j\operatorname{conv}(A_j)=\emptyset.$$

Choose points a_{ij}: $a_{ij}\in A_i\cap A_j$ for $i,j=1,\ldots,I$. Then

$$B=\sum_{i=1}^{I}\alpha_i\operatorname{conv}(A_i)\supset\left\{\sum_{i=1}^{I}\alpha_i a_{ij}\colon j=1,\ldots,I\right\}$$

and

$$\left\{\sum_{i=1}^{I}\alpha_i a_{ij}\colon j=1,\ldots,I\right\}\cap\operatorname{conv}(A_j)\neq\emptyset \quad \text{for all } j=1,\ldots,I.$$

Since B is convex this implies that also

$$B\cap\sum_{j=1}^{I}\beta_j\operatorname{conv}(A_j)\neq\emptyset. \quad \text{Q.E.D.}$$

8. Some Additional Remarks, Observations, and Problems

a) The Channel Model

The channels considered in this paper can be viewed as a model for a transmission system which has several states and those states change arbitrarily from one time point to the next. In a so called "finite state channel" the changes of states are assumed to follow probabilistic laws. Whenever changes of states are not governed by a probability distribution or if this distribution is not known, then one can describe the situation by an AVC-model. It is even conceivable that a malevolant being like Maxwell's demon for instance chooces the states so as to make communication as difficult as possible – this is still incorporated into the model. Moreover, this demon could decide at any time to

randomize over the states – thus enlarging the set of matrices $\mathfrak{C}$ to the set $\overline{\mathfrak{C}}$ – this would still be within the boundaries of the model.

Now we go one step further. Suppose the demon chooses at time t his randomisation depending on the letter x given into the system, that is x must be known to him, before he randomizes, then this means that the class $\mathfrak{C}$ is replaced by $\overline{\overline{\mathfrak{C}}}$. For this situation the SAVC provides the appropriate model, which is mathematically just a special AVC.

But now we go even further. Suppose a K_1-code $\{(u_i, D_i)\colon i=1,\ldots,N\}$ has been chosen. Assume that the demon even *knows the* u_i to be sent before it is sent and that he can choose at any time instance his states or randomisations over states *depending* on this knowledge of the word. The channel model for this situation has been called by Kiefer and Wolfowitz ([2]) a "channel with arbitrarily varying channel probability functions".

Let us use the abbreviation A*VC. The difference between the two channels – AVC and A*VC – can mathematically be well explained if we consider the average error probability of a K_1-code for both. (For maximal errors there is no difference in the K_1-coding problem for the two channels as can be seen from Lemma 3 of [7]). We recall that for the AVC

$$\bar{\lambda}_1 = \sup_{s^n \in S^n} \frac{1}{N} \sum_{i=1}^{N} P(D_i^c | u_i | s^n).$$

For the A*VC s^n may depend on u_i and therefore

$$\bar{\lambda}_1^* = \frac{1}{N} \sum_{i=1}^{N} \sup_{s^n \in S^n} P(D_i^c | u_i | s^n).$$

(In general $\bar{\lambda}_1^*$ may be much larger than $\bar{\lambda}_1$.)

But then we can pass to an $\left(n, \left[\frac{N}{2}\right]\right)$-subcode with maximal error probability $\lambda^* \leqq 2\bar{\lambda}_1^*$. That means that the capacities C_1^* and $\bar{C}_1^*$ are the same. There is no advantage in allowing average error. A similar argument applies to randomized encoding and to correlated codes. All those performances do not help anything for A*VC's. Definitely those channels give the much more robust model. For an even more robust model we refer to the "subchannels" introduced in [23]. There again no single-letter characterisation of the capacity is available.

The papers [2, 7, 11], and [12] deal with A*VC's. The papers [1, 3, 4–6, 13] consider AVC's.

b) An Analogy

A similar distinction as for arbitrarily varying channels can be made for compound channels. Let us denote them by CC and C^*C. The paper [15] deals with C^*C and the papers [14, 16], and [18] consider CC. The "star" was noticed in [16]: for average error CC has no strong converse. This phenomenon was studied systematically in [18] and let to the following result:

for the λ-capacities (in the sense of [17])

$$\bar{C}_1(\lambda) = C_2(\lambda) = \bar{C}_2(\lambda), \quad 0 < \lambda < 1. \tag{8.1}$$

Whereas for compound channels still $\bar{C}_1 = \inf_{\lambda} \bar{C}_1(\lambda) = C_1^*$, this identity fails for the arbitrarily varying channels, because always $C_1 = C_1^* \leqq C_D$ and in case $C_D \neq C_R = \bar{C}_1$ clearly $C_1^* < \bar{C}_1$.

However, (8.1) finds its analog in the relations $\bar{C}_1 = C_2 = \bar{C}_2$ (Theorem 3).

c) There are Two Theories of Coding

By the previous remarks we know that in general for the AVC (8.2) $C_1 \neq \bar{C}_1$.

This may be viewed as a pathology of *the channel.* The fact that the zero-error capacity problem for a DMC ([8]) is equivalent to the coding problem of a very special (only $0-1$-matrices in $\mathfrak{C}$) A*VC (see [9]) may confirm this view, because this 0-error problem is considered by many workers in the field not to be information theoretic in nature, because there seem to be no connections to standard information measures. However, in studying multi-way channels, in particular Shannon's two-way channel in [20] and the so called multiple-access channel in [21], we became strongly convinced that *also for those channels*

$$\mathfrak{R}_1 \neq \bar{\mathfrak{R}}_1 \quad \text{(the regions) can occur.} \tag{8.3}$$

This has recently been confirmed by examples ([22]). Whereas $\bar{\mathfrak{R}}_1$ can be nicely described in terms of information measures no such characterisation now exists or is even to be expected to exist at all for $\mathfrak{R}_1$. It seems likely that completely new information measures have to be invented and this must be very hard, because a determination of $\mathfrak{R}_1$ for instance for the multiple-access channel already would imply in the deterministic case the asymptotic solution of a whole spectrum of combinatorial extremal problems. In the study of multi-user coding problems it has become clear that the theory of coding as founded by Shannon is intimately connected with the average error concept. Theorem 1 confirms this view.

d) "In Many Data-Transmission Systems, the Probabilities with Which Messages are to be Used are Either Unknown or Unmeaningful" ([24], p. 14)

Taking this seriously and also what we said in c) one should worry about the use of building a theory of multi-user communication. Indeed, we had such doubts for quite a while. Therefore we consider it as a gift from the gods that Theorem 2 holds. Randomisation in the encoding can be used successfully if the message statistic is unknown!

e) Comments about the Proofs of Theorems 1, 2, 3

The approach chosen in this paper makes use of the Random code theorem of [1]. There it is proved by a game theoretic argument. Since the Shannon

random coding method consists in specifying a certain K_3-code it is natural to base a proof directly on it. This was done in [5] with the decoding idea of the "ideal observer". Then the novelty of our approach simply consists in applying Shannon's method *several times* and to use the concatenation argument. The simple probabilistic fact which makes every thing work is that the probability with which the mean $n^{-1}\sum_{i=1}^{n} X_i$ of independent, identically distributed RV's exceeds any number larger than the common expected value, decreases exponentially with n. In our first attack on the problems we startet out with the list reduction lemma of [11] or [12] for maximal errors. Selecting out of a list code a subcode at random and passing to average errors one can also prove Theorem 3, but not more. At this stage the english translation of the Dobrushin-Stambler paper [13] reached us and we were stimulated by the fact that they could prove $\bar{C}_1 = C_R$ for certain channels, which was somewhat surprising because of Example 1 (from [1]) known to us. Taking all auxiliary results into consideration our first proof of Theorem 3 is somewhat more complicated than the one presented here and we therefore did not include it. Let us just announce the following part of some independent interest:

Selecting a code with positive rate at random for a DMC fails to lead to a code with error probability bounded by a constant λ with double exponentially small probability.

f) Problems

Among the references we have listed a number of papers ([23, 27, 28, 25, 26]) dealing with arbitrarily varying *sources* from several points of view.

The more recent work on correlated sources ([29–32]) is intimately connected with channel coding techniques. In source coding maximal errors don't occur and we feel that the foundations now laid should suffice to deal successfully with coding problems for arbitrarily varying correlated sources (without or with side information). Similar extensions for multi-way channels should now be within reach. For recent surveys on this topic see [33] and [34].

References

1. Blackwell, D., Breimann, L., Thomasian, A.J.: The capacities of certain channel classes under random coding. Ann. Math. Statist. **31**, 558–567 (1960)
2. Kiefer, J., Wolfowitz, J.: Channels with arbitrarily varying channel probability functions. Information and Control **5**, 44–54 (1962)
3. Dobrushin, R.L.: Unified information-transmission schemes for discrete memoryless channels and messages with independent components. Dokl. Akad. Nauk. SSSR 148, No. **6**, 1245–1248 (1963)
4. Dobrushin, R.L.: Unified information-transmission schemes: the general case. Dokl. Akad. Nauk SSSR 149, No. **1**, 16–19 (1963)
5. Stiglitz, I.G.: Coding for a class of unknown channels. IEEE Trans. Information Theory, IT-12, No. **2**, 189–195 (1966)
6. Ahlswede, R., Wolfowitz, J.: Correlated decoding for channels with arbitrarily varying channel probability functions. Information and Control, **14**, No. 5, 457–473 (1969)
7. Ahlswede, R., Wolfowitz, J.: The capacity of a channel with arbitrarily varying channel probability functions and binary output alphabet. Z. Wahrscheinlichkeitstheorie verw. Gebiete, **15**, 186–194 (1970)

8. Shannon, C.E.: The zero error capacity of a noisy channel. IRE Trans. Information Theory IT-2, 8–19 (1956)
9. 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. Ann. Math. Statist. **41**, 1027–1033 (1970)
10. Elias, P.: List decoding for noisy channels. Technical Report 335, Research Laboratory of Electronics, MIT, Cambridge, Mass (1955)
11. Ahlswede, R.: Channels with arbitrarily varying channel probability functions in the presence of noiseless feedback. Z. Wahrscheinlichkeitstheorie verw. Gebiete, **25**, 3, 239–252 (1973)
12. Ahlswede, R.: Channel capacities for list codes. J. Appl. Probability **10**, 824–836 (1973)
13. Dobrushin, R.L., Stambler, S.Z.: Coding theorems for classes of arbitrarily varying discrete memoryless channels. Problems of Information Transmission, Vol. **11**, No. 2, 3–22 (1975)
14. Blackwell, D., Breimann, L., Thomasian, A.G.: The capacity of a class of channels. Ann. Math. Statist., **30**, 1229–1241 (1959)
15. Wolfowitz, J.: Simultaneous channels. Arch. Rational Mech. Anal. **4**, 371–386 (1960)
16. Ahlswede, R.: Certain results in coding theory for compound channels. Proc. Coll. on Information. Theory, Debrecen, Hungary, 35–60 (1967)
17. Wolfowitz, J.: Channels without capacity. Information and Control **6**, 49–54 (1963)
18. Ahlswede, R., Wolfowitz, J.: The structure of capacity functions for compound channels. Lecture Notes in Mathematics Probability and Information Theory: Proc. of the International Symposium at McMaster University. Canada, April 1968, No. 89, pp. 12–54. Berlin-Heidelberg-New York: Springer 1969
19. Shannon, C.E.: Two-way communication channels. Math. Statist. Probab. Proc. 4th Berkeley Sympos. Math. Statist. Probab. Univ. Calif. **1**, 611–644 (1961)
20. Ahlswede, R.: On two-way communication channels and a problem of Zarankiewicz. Trans. 6th Prague Confer. on Information Theory, Statistical Decision Functions and Random Processes, Prague, 23–37 (1971)
21. Ahlswede, R.: Multy-way communication channels. Second Internat. Sympos. Information Theory, Thakadsor, 1971, Publishing House of the Hungarian Ac. of Sc., 23–52 (1973)
22. Dueck, G.: Maximal error capacity regions are smaller than average error capacity regions for multi-user channels. Submitted to Problems of Control and Information Theory
23. Strassen, V.: Messfehler und Information. Wahrscheinlichkeitstheorie verw. Gebiete **2**, 273–305 (1964)
24. Gallager, R.G.: Information Theory and Reliable Communication. New York: Wiley 1968
25. Davisson, L.D.: Universal noiseless coding. IEEE Trans. Information Theory IT-19, 783–795 (1972)
26. Ziv, J.: Coding of sources with unknown statistics. IEEE Trans. Information Theory, IT-18, 384–388 (1972)
27. Dobrushin, R.L.: Unified schemes for optimal message quantization. Problems of cybernetics [in Russian], No. 22, Nauka Moscow, 107–156 (1970)
28. Berger, T.: The source coding game. IEEE Trans. Information Theory, IT-17 (1971)
29. Slepian, D., Wolf, J.K.: Noiseless coding of correlated information sources. IEEE Trans. Information Theory, Vol. IT-19, 471–479 (1973)
30. Ahlswede, R., Körner, J.: Source coding with sicle information and a converse for degraded broadcast channels. IEEE Trans. Information Theory IT-21, 629–637 (1975)
31. Wyner, A.D.: On source coding with side information at the decoder. IEEE Trans. Information Theory IT-21, 294–300 (1975)
32. Gallager, R.G.: Source coding with side information and universal coding. To appear in IEEE Trans. Information Theory
33. van der Meulen, E.C.: Multi-way channels. Survey article, IEEE Trans. Information Theory (1978)
34. Cover, T.M.: Some advances in broadcast channels. Chapter in Advances in Communication Systems, Vol. 4, Theory and Applications, Ed. by A. Viterbi. San Francisco: Academic Press 1975
35. Shannon, C.E.: A note on a partial ordering for communication channels. Information and Control **1**, 390–397 (1958)

Received November 22, 1976