Z. Wahrscheinlichkeitstheorie verw. Geb. 25, 239–252 (1973)

Channels with Arbitrarily Varying Channel Probability Functions in the Presence of Noiseless Feedback

Rudolf Ahlswede

In this article we study a channel with arbitrarily varying channel probability functions in the presence of a noiseless feedback channel (a.v.ch.f.). We determine its capacity by proving a coding theorem and its strong converse. Our proof of the coding theorem is constructive; we give explicitly a coding scheme which performs at any rate below the capacity with an arbitrarily small decoding error probability. The proof makes use of a new method ([1]) to prove the coding theorem for discrete memoryless channels with noiseless feedback (d.m.c.f.). It was emphasized in [1] that the method is not based on random coding or maximal coding ideas, and it is this fact that makes it particularly suited for proving coding theorems for certain systems of channels with noiseless feedback.

As a consequence of our results we obtain a formula for the zero-error capacity of a d.m.c.f., which was conjectured by Shannon ([8], p. 19).

1. Introduction

Let $X=\{1, \dots, a\}$ and $Y=\{1, \dots, b\}$ be finite sets, which serve as input and output alphabets of the channels discribed below. Write $X^t=X$ and $Y^t=Y$ for $t=1, 2, \dots$. By $X_n=\prod_{t=1}^{n} X^t$ denote the set of input n-sequences (words of length n) and by $Y_n=\prod_{t=1}^{n} Y^t$ denote the set of output n-sequences. Let $w(\cdot|\cdot)$ be a stochastic $a \times b$-matrix. The transmission probabilities of a discrete memoryless channel (d.m.c.) $\mathscr{D}$ are defined by

$$(1.1) \qquad P(y_n|x_n)=\prod_{t=1}^{n} w(y^t|x^t) \qquad \text{for every } x_n=(x^1, \dots, x^n) \in X_n,$$

$y_n=(y^1, \dots, y^n) \in Y_n$ and every $n=1, 2, \dots$.

Let S be any set, and let $\mathfrak{C}=\{w(\cdot|\cdot|s)|s \in S\}$ be a set of stochastic $a \times b$-matrices $w(\cdot|\cdot|s)$. Set $S^t=S$ for $t=1, 2, \dots$. For every $s_n=(s^1, \dots, s^n) \in S_n=\prod_{t=1}^{n} S^t$ we define $P(\cdot|\cdot|s_n)$ by

$$(1.2) \qquad P(y_n|x_n|s_n)=\prod_{t=1}^{n} w(y^t|x^t|s^t) \qquad \text{for all } x_n \in X_n, \; y_n \in Y_n.$$

For every n; $n=1, 2, \dots$; set $\mathfrak{C}_n=\{P(\cdot|\cdot|s_n)|s_n \in S_n\}$. A channel with arbitrarily varying channel probability functions (a.v.ch.) $\mathfrak{A}$ is defined by the sequence $(\mathfrak{C}_n)_{n=1,2,\dots}$.

$\mathfrak{A}$ is of o-1-type, if $\mathfrak{C}$ contains only matrices, which have 0 and 1 as entries. We denote this channel by $\mathfrak{A}_1$.

Suppose that sender and receiver want to communicate over $\mathfrak{A}$ without knowing which $P(\cdot|\cdot|s_n)$ will govern the transmission of any word sent. A code (n, N, λ) for this situation is a system $\{(u_i, A_i)|i=1, \dots, N\}$, where $u_i \in X_n$, $A_i \subset Y_n$ for $i=1, 2, \dots, N$; $A_i \cap A_j = \emptyset$ for $i \neq j$ and $P(A_i|u_i|s_n) \geqq 1-\lambda$ for $i=1, \dots, N$ and all $s_n \in S_n$. A number C is called the capacity of the channel $\mathfrak{A}$, if for any $\varepsilon > 0$ and any λ, $0<\lambda<1$, there exists a code $(n, e^{(C-\varepsilon)n}, \lambda)$ and there does not exist a code $(n, e^{(C+\varepsilon)n}, \lambda)$ for all sufficiently large n. In case $b=2$ a formula for C is known ([4]). For $b \geqq 3$ a formula for the capacity, which makes it in principle possible to compute its value within any desired accuracy, is still unknown. It was shown in [2] that the problem to determine the capacity C_1 of $\mathfrak{A}_1$ is equivalent to the problem to find a computable formula for the zero-error capacity ([8]) C_0 of a discrete memoryless channel (d.m.c.). This problem is of graph theoretic nature and still unsolved.

We introduce now an a.v.ch. with noiseless feedback (a.v.ch.f.) which we denote by $\mathfrak{A}_f$. By this we mean that in addition to $\mathfrak{A}$ there exists a return channel which sends back from the receiving point to the transmitting point the element of Y actually received. It is assumed that this information is received at the transmitting point before the next letter is sent, and can therefore be used for choosing the next letter to be sent. The assumption of noiseless feedback is certainly quite restrictive for a real communication situation, but mathematically it should be considered as a step forward that one can prove theorems about a.v.ch. under this assumption. Shannon gave in [8] for a.d.m.c. with noiseless feedback (d.m.c.f.) $\mathscr{D}_f$ a formula for its zero-error capacity C_{0f}. This result encouraged us in finding a formula for the capacity of $\mathfrak{A}_f$. However the approach taken by Shannon in [8] does not extend to $\mathfrak{A}_f$.

Henceforth, when we talk about feedback we shall always mean noiseless feedback.

We describe now the encoding for $\mathscr{D}_f$ and $\mathfrak{A}_f$. Suppose there is given a finite set of messages $M=\{1, \dots, N\}$ one of which will be presented to the sender for transmission. Message $m \in M$ is encoded by an encoding (vector valued) function

$$f_n(m) = [f_m^1, f_m^2(Z^1), \dots, f_m^t(Z^1, \dots, Z^{t-1}), \dots, f_m^n(Z^1, \dots, Z^{n-1})], \tag{1.3}$$

where f_m^t is defined on Y_{t-1} for $t>1$ and takes values in X^t, and $Z^1, Z^2, \dots, Z^{t-1}$ are the chance received elements of Y (known to the sender before he sends $f_m^t(Z^1, \dots, Z^{t-1})$); f_m^1 is an element of X^1.

The distribution of the random variables Z^t; $t=1, 2, \dots, n$; is determined by $f_m^1, \dots, f_m^{t-1}$ and w (resp. s_n). We denote the probability of receiving $y_n \in X_n$, if m is thus encoded, by $P(y_n|f_n(m))$ (resp. $P(y_n|f_n(m)|s_n)$, $s_n \in S_n$).

An (n, N) code is a system

$$\{(f_n(m), A_m)|m \in M\}, \quad \text{where} \quad A_m \subset Y_n \quad \text{for } m \in M; \quad A_m \cap A_{m'} = \emptyset \tag{1.4}$$

for $m \neq m'$, and the $f_n(m)$ are defined as in (1.3).

The (n, N) code $\{(f_n(m), A_m)|m \in M\}$ is an $(n, N, 0)$ code for $\mathscr{D}_f$, if

$$P(A_m|f_n(m)) = 1 \quad \text{for } m \in M, \tag{1.5}$$

and an (n, N, λ) code for $\mathfrak{A}_f$, if

$$P(A_m|f_n(m)|s_n) \geqq 1-\lambda \quad \text{for all } m \in M \text{ and } s_n \in S_n. \tag{1.6}$$

(1.7) A number C_{0f} is called the zero-error capacity of $\mathscr{D}_f$ if for any $\varepsilon>0$ and some $n_0(\varepsilon)$ there exists a code $(n, e^{(C_{0f}-\varepsilon)n}, 0)$, and if for no n there exists a code $(n, e^{(C_{0f}+\varepsilon)n}, 0)$.

(1.8) A number C_f is called the capacity of $\mathfrak{A}_f$, if for any $\varepsilon>0$ and any λ, $0<\lambda<1$, the following is true for all n sufficiently large: There exists a code $(n, e^{(C_f-\varepsilon)n}, \lambda)$ and there does not exist a code $(n, e^{(C_f+\varepsilon)n}, \lambda)$.

Denote C_f of $\mathfrak{A}_{1f}$ by C_{1f}.

We introduce now several channels which are related to channel $\mathfrak{A}$.

For $i\in X$ let $T(i)$ be the convex closed hull of the set of probability distributions (p.d.) $\{w(\cdot|i|s)|s\in S\}$.

Denote by $\bar{\mathfrak{C}}$ the closed convex hull of $\mathfrak{C}$ and by $\bar{\bar{\mathfrak{C}}}$ the row-convex hull of $\mathfrak{C}$, that is

$$\bar{\bar{\mathfrak{C}}}=\{w(\cdot|\cdot)|w(\cdot|\cdot)\in T(i) \text{ for } i\in X\}. \tag{1.9}$$

Define $\mathfrak{C}^e$ by

$$\mathfrak{C}^e=\{w(\cdot|\cdot)|\text{for every } i\in X \text{ there exists an } s\in S\colon\ w(\ |i)=w(\ |i|s)\}. \tag{1.10}$$

For $n=1,2,\dots$ define $\bar{\mathfrak{C}}_n$, $\bar{\bar{\mathfrak{C}}}_n$ and $\mathfrak{C}_n^e$ analogously to $\mathfrak{C}_n$ and index sets $\bar{S}_n=\prod_1^n \bar{S}$, $\bar{\bar{S}}_n=\prod_1^n \bar{\bar{S}}$ and $S_n^e=\prod_1^n S^e$.

Set $\bar{\mathfrak{A}}=(\bar{\mathfrak{C}}_n)_{n=1,2,\dots}$; $\bar{\bar{\mathfrak{A}}}=(\bar{\bar{\mathfrak{C}}}_n)_{n=1,2,\dots}$; $\mathfrak{A}^e=(\mathfrak{C}_n^e)_{n=1,2,\dots}$.

In case of feedback we write $\bar{\mathfrak{A}}_f$, $\bar{\bar{\mathfrak{A}}}_f$ and $\mathfrak{A}_f^e$ or $\bar{\mathfrak{A}}_{1f}$, $\bar{\bar{\mathfrak{A}}}_{1f}$ and $\mathfrak{A}_{1f}^e$, if $\mathfrak{C}$ contains only 0-1-matrices. Denote the corresponding capacities – in case they exist – by $\bar{C}_f$, $\bar{\bar{C}}_f$, C_f^e, $\bar{C}_{1f}$ $\bar{\bar{C}}_{1f}$ and C_{1f}^e.

We say that a channel has a positive rate, if for a positive number R and for any λ, $0<\lambda<1$, there exists a code (n, e^{Rn}, λ) for all sufficiently large n. One easily verifies that $\mathfrak{C}^e=\bar{\bar{\mathfrak{C}}}$ and hence also that $\mathfrak{A}_f^e=\bar{\bar{\mathfrak{A}}}_f$. It follows from Lemma 2 in section 2 that the capacities for $\mathfrak{A}_f$ and $\bar{\mathfrak{A}}_f$ (resp. $\mathfrak{A}_f^e$ and $\bar{\bar{\mathfrak{A}}}_f$) are equal. One can limit oneself therefore to the study of the channels $\bar{\mathfrak{A}}_f$ and $\bar{\bar{\mathfrak{A}}}_f$. If we choose $\mathfrak{C}$ such that $\mathfrak{C}^e=\mathfrak{C}$, then $\bar{\mathfrak{A}}_f=\bar{\mathfrak{A}}_f^e=\bar{\bar{\mathfrak{A}}}_f$. A channel $\bar{\bar{\mathfrak{A}}}_f$ is therefore a special channel of type $\bar{\mathfrak{A}}_f$. Example 2 in section 3 shows that $\bar{\mathfrak{A}}_f$ can have positive rate and still $\bar{\bar{\mathfrak{A}}}_f$ has capacity 0. (It was shown in [5], Lemma 3, that $\mathfrak{A}$, $\bar{\mathfrak{A}}$ and $\bar{\bar{\mathfrak{A}}}$ have the same capacities.) In section 3 the channel $\bar{\bar{\mathfrak{A}}}_f$ is treated and we explain there how the limitation to this channel can be motivated from a practical point of view. The other channel concepts introduced above are needed in section 4 only.

2. Auxiliary Results

In order to state and prove several lemmas used in the later sections we need the following list of definitions.

For a p.d. π on X and a stochastic $a\times b$-matrix define a p.d. q on Y by

$$q_j=\sum_i \pi_i\, w(j|i), \qquad j\in Y. \tag{2.1}$$

We also shall write $q(w)$ instead of q, if we want to indicate the dependence on w.

Define w^* as any stochastic $b\times a$-matrix which satisfies

$$w^*(i|j)\, q_j=\pi_i\, w(j|i), \qquad i\in X,\ j\in Y. \tag{2.2}$$

Let l be a positive integer:

(2.3) For $u \in X_l$, $v \in Y_l$, $i \in X$ and $j \in Y$ let $N(i|u)$ (resp. $N(j|v)$) count how often i (resp. j) occurs as a component of u (resp. v) and let $N(i, j|u, v)$ count the number of components in which u has an i and v has a j.

(2.4) Set $Q(u, i) = \{t | u^t = i\}$ for $u = (u^1, \ldots, u^l) \in X_l$ and $i \in X$.

(2.5) For $v \in Y_l$ define a p.d. $\tilde{q}$ on Y by $\tilde{q}_j = N(j|v)\, l^{-1}$ for $j \in Y$.

We define now the following sets:

(2.6) $$X_l(\pi) = \{x_l | x_l \in X_l, |\pi_i l - N(i|x_l)| \leqq 1 \text{ for } i \in X\},$$

(2.7) $$Y_l(u, \varepsilon, s_l) = \{v | v \in Y_l, |N(i, j|u, v) - \sum_{t \in Q(u, i)} w(j|i|s^t)| \leqq \varepsilon l \text{ for } i \in X, j \in Y\},$$

where $u \in X_l(\pi)$, $s_l = (s^1, \ldots, s^l) \in \overline{\overline{S}}_l$ and $\varepsilon > 0$.

(2.8) $$Y_l(u, \varepsilon) = \bigcup_{s_l \in \overline{\overline{S}}_l} Y_l(u, \varepsilon, s_l),$$

(2.9) $$Y_l(u, \varepsilon, w) = \{v | v \in Y_l, |N(i, j|u, v) - N(i|u)\, w(j|i)| \leqq \varepsilon l \text{ for } i \in X, \ j \in Y\},$$

(2.10) $$\overline{\overline{\mathfrak{C}}}(\varepsilon, \pi, \tilde{q}, l) = \{w | w \in \overline{\overline{\mathfrak{C}}}, |q_j(w) - \tilde{q}_j| \leqq a(\varepsilon + l^{-1}) \text{ for } j \in Y\},$$

(2.11) $$X_l(v, \varepsilon, \pi, w) = \{u | u \in X_l(\pi), |N(i, j|u, v) - w^*(i|j)\, N(j|v)| \leqq 2a(\varepsilon l + 1) \text{ for } i \in X, j \in Y\},$$

(2.12) $$X_l(v, \varepsilon, \pi) = \bigcup_{w \in \overline{\overline{\mathfrak{C}}}(\varepsilon, \pi, \tilde{q}, l)} X_l(v, \varepsilon, \pi, w).$$

Finally we define functions H, R and $\overline{\overline{K}}$ by

(2.13) $$H(p) = -\sum_{i=1}^{c} p_i \log p_i \quad \text{for a p.d. } p = (p_1, \ldots, p_c).$$

(2.14) $$R(\pi, w) = H(\pi) - \sum_j q_j H(w^*(\cdot|j))$$

and

(2.15) $$\overline{\overline{K}} = \max_{\pi} \min_{w \in \overline{\overline{\mathfrak{C}}}} R(\pi, w).$$

We come now to the lemmas. The most important one of them is Lemma 1. For it's proof we need 4 propositions.

Proposition 1.

$$\exp\{H(\pi)\, l - c(\pi) \log l\} \leqq |X_l(\pi)| \leqq \exp\{H(\pi)\, l + c(\pi) \log l\}, \quad \textit{for } l = 1, 2, \ldots.$$

$c(\pi)$ is a function, which can be given explicitly.

This follows immediately from definition (2.6) and Stirling's formula.

Proposition 2. *One can give explicitly a function $E(\varepsilon, w) > 0$ which is continuous in w, such that for $u \in X_l$, $s_l \in \overline{\overline{S}}_l$ and $l = 1, 2, \ldots$:*

$$P\big(Y_l(u,\varepsilon,s_l)\,|\,u\,|\,s_l\big)\geqq 1-\exp\left\{-\sum_{t=1}^{l} E\big(\varepsilon, w(\cdot\,|\cdot|\,s^t)\big)\right\}\geqq 1-\exp\{-E(\varepsilon)\,l\},$$

where $E(\varepsilon)=\min\limits_{w\in\bar{\bar{\mathfrak{C}}}} E(\varepsilon,w)>0$.

It follows from definition (2.7) and Chebyshev's inequality that one can construct a function $E(\varepsilon,w)$ with the desired properties. $E(\varepsilon)$ is positive, because $\bar{\bar{\mathfrak{C}}}$ is compact and $E(\varepsilon,w)$ is continuous and positive.

Proposition 3. *For* $w\in\bar{\bar{\mathfrak{C}}}$, $v\in Y_l$ *and* $\tilde q$ *as in* (2.5):

a) $|X_l(v,\varepsilon,\pi,w)|\leqq\exp\{\sum_j \tilde q_j H(w^*(\cdot|j))\,l+g(\varepsilon)\,l\}$,

b) $|X_l(v,\varepsilon,\pi)|\leqq\exp\{\max\limits_{w\in\bar{\bar{\mathfrak{C}}}}\sum_j q_j(w)\,H(w^*(\cdot|j))\,l+\bar g(\varepsilon)\,l\}$ *for* $l\geqq c_0(\varepsilon)$.

$g(\varepsilon)$, $\bar g(\varepsilon)$ *and* $c_0(\varepsilon)$ *are known functions and* $\lim\limits_{\varepsilon\to 0} g(\varepsilon)=\lim\limits_{\varepsilon\to 0}\bar g(\varepsilon)=0$.

Proof. Part a) follows from (2.11) and Chebyshev's inequality. (Compare lemma 2.1.6. of [9]. The only difference between that lemma and part a) of our proposition is that we use w^* instead of w.) We proof now part b). The set $\bar{\bar{\mathfrak{C}}}(\varepsilon,\pi,\tilde q,l)$ can be partioned into disjoint sets $\bar{\bar{\mathfrak{C}}}(1),\dots,\bar{\bar{\mathfrak{C}}}(L)$ in such a way that for 2 matrices in $\bar{\bar{\mathfrak{C}}}(\rho)$; $\rho=1,\dots,L$; the corresponding stared matrices differ componentwise by less than ε and such that $L\leqq(1/\varepsilon)^{ab}$. Let w_ρ be an element of $\bar{\bar{\mathfrak{C}}}(\rho)$ and let w_ρ^* correspond to w_ρ, then

$$\text{(2.16)}\qquad \bigcup_{w\in\bar{\bar{\mathfrak{C}}}(\rho)} X_l(v,\varepsilon,\pi,w)\subset\{u\,|\,u\in X_l(\pi),\,|N(i,j\,|\,u,v)-w_\rho^*(i|j)\,N(j\,|\,v)|\leqq(2a+1)(\varepsilon\,l+1)\text{ for } i\in X, j\in Y\}.$$

Part a) yields

$$\text{(2.17)}\qquad \Big|\bigcup_{w\in\bar{\bar{\mathfrak{C}}}(\rho)} X_l(v,\varepsilon,\pi)\Big|\leqq\exp\{\sum_j \tilde q_j H\big(w_\rho^*(\cdot|j)\big)\,l+g_\rho^*(\varepsilon)\,l\}$$

where $g_\rho^*(\varepsilon)$ is a known function and $\lim\limits_{\varepsilon\to 0} g_\rho^*(\varepsilon)=0$.

Since

$$|X_l(v,\varepsilon,\pi)|\leqq(1/\varepsilon)^{ab}\max_\rho\Big|\bigcup_{w\in\bar{\bar{\mathfrak{C}}}(\rho)} X_l(v,\varepsilon,\pi,w)\Big|$$

the statement follows from (2.17) and definition (2.10).

Proposition 4. *If* $u\in X_l(\pi)$ *and* $v\in Y_l(u,\varepsilon)$, *then*

a) $v\in\bigcup\limits_{w\in\bar{\bar{\mathfrak{C}}}(\varepsilon,\pi,\tilde q,l)} Y_l(u,\varepsilon,w)$

and

b) $u\in X_l(v,\varepsilon,\pi)$.

Proof. $v\in Y_l(u,\varepsilon)$ and (2.8) imply $v\in Y_l(u,\varepsilon,s_l)$ for some $s_l\in\bar{\bar{S}}_l$. Introduce a matrix $\tilde w(\cdot|\cdot)\in\bar{\bar{\mathfrak{C}}}$ by

$$\text{(2.18)}\qquad \tilde w(j\,|\,i)=N(i\,|\,u)^{-1}\sum_{t\in Q(u,i)} w(j\,|\,i\,|\,s^t)\qquad\text{for } i\in X,\ j\in Y.$$

From (2.7), (2.9) and (2.18) one obtains

$$\text{(2.19)}\qquad Y_l(u,\varepsilon,s_l)=Y_l(u,\varepsilon,\tilde w)$$

Since $u\in X_l(\pi)$ one can conclude that

$$\text{(2.20)}\quad Y_l(u,\varepsilon,s_l)\subset\{v^*|\;|N(i,j\,|\,u,v^*)-\tilde w(j\,|\,i)\,\pi_i\,l|\leqq\varepsilon\,l+1\ \text{ for } i\in X, j\in Y\}.$$

Since $N(j|v)=\sum_i N(i,j|u,v)$ for $j\in Y$, we obtain from (2.20) that

$$(2.21)\qquad |N(j|v)-q_j(\tilde{w})\,l|\leqq a(\varepsilon\, l+1)\quad \text{for } j\in Y.$$

This and the definition of $\tilde{q}$ yield

$$(2.22)\qquad |\tilde{q}_j-q_j(\tilde{w})|\leqq a(\varepsilon+l^{-1})\quad \text{for } j\in Y$$

and therefore we have $\tilde{w}\in\bar{\bar{\mathfrak{C}}}(\varepsilon,\pi,\tilde{q},l)$, which proves part a).

It is clear from the definition of $X_l(v,\varepsilon,\pi)$ and from a) that in order to establish b) it suffices to show that for any $w\in\bar{\bar{\mathfrak{C}}}$ and $u\in X_l(\pi)$ the following relation holds:

$$(2.23)\qquad v\in Y_l(u,\varepsilon,w)\quad \text{implies}\quad u\in X_l(v,\varepsilon,\pi,w).$$

From $u\in X_l(\pi)$ and $v\in Y_l(u,\varepsilon,w)$ we obtain for $i\in X, j\in Y$ that

$$|N(i,j|u,v)-\pi_i\, l\, w(j|i)|\leqq\varepsilon\, l+1,\qquad |N(j|v)-q_j\, l|\leqq a(\varepsilon\, l+1)$$

and finally that $|N(i,j|u,v)-N(j|v)\,w^*(i|j)|\leqq(a+1)(\varepsilon\, l+1)$. Hence, $u\in X_l(v,\varepsilon,\pi,w)$ and b) is proved.

The system $\left(X_l(\pi),\{X_l(v,\varepsilon,\pi)|\,v\in Y_l\}\right)$ can be interpreted as a list code for $\bar{\bar{\mathfrak{A}}}$. $X_l(\pi)$ is the set of code words and $X_l(v,\varepsilon,\pi)$ is the list of code words the receiver decides upon, if he has received v. For any list code denote by N its length, by L its maximal list size and by λ its maximal error probability for channel $\bar{\bar{\mathfrak{A}}}$. In this case: $N=|X_l(\pi)|$, $L=\max_v |X(v,\varepsilon,\pi)|$.

Lemma 1. *One can give explicitly a function $c_1(\varepsilon)$ such that for $l\geqq c_1(\varepsilon)$ the list code $\left(X_l(\pi),\{X_l(v,\varepsilon,\pi)|\,v\in Y_l\}\right)$ for $\bar{\bar{\mathfrak{A}}}$ has the following properties:*

a) $\exp\{H(\pi)\,l+g_1(\varepsilon)\,l\}\geqq N\geqq\exp\{H(\pi)\,l-g_1(\varepsilon)\,l\}$,

b) $L\leqq\exp\{\max_{w\in\bar{\bar{\mathfrak{C}}}}\sum_j q_j(w)\,H\left(w^*(\cdot|j)\right)l+g_1(\varepsilon)\,l\}$,

c) $\lambda\leqq\exp\{-E(\varepsilon)\,l\}$.

$E(\varepsilon)$ and $g_1(\varepsilon)$ are known positive functions and $\lim_{\varepsilon\to 0} g_1(\varepsilon)=0$.

Proof. a) is clear from Proposition 1, and b) follows from Proposition 3, b). c) is a consequence of Proposition 2 and Proposition 4, b).

For the channel $\mathfrak{A}_{1f}$ we can choose S as a finite set. Define $\bar{w}(\cdot|\cdot)$ by

(2.24) $\bar{w}(j|i)=|S|^{-1}\sum_{s\in S} w(j|i|s)$ for $i\in X$; $j\in Y$; and define a d.m.c. $\bar{\mathscr{D}}$ as in (1.1).

Denote this channel by $\bar{\mathscr{D}}_f$ in case of feedback.

Lemma 2. *For any λ, $0\leqq\lambda<1$, we have:*

a) *an (n,N,λ) code for $\mathfrak{A}_f$ is an (n,N,λ) code for $\bar{\mathfrak{A}}_f$, and conversely,*

b) *an (n,N,λ) code for $\mathfrak{A}_{1f}$ is an (n,N,o) code for $\bar{\mathscr{D}}_f$, and conversely.*

Proof. Any element $w(\cdot|\cdot|\bar{s})\in\bar{\mathfrak{C}}$ can be approximated arbitrarily closely by expressions of the form

$$\sum_{s\in S} r(s|\bar{s})\,w(\cdot|\cdot|s),$$

where $r(\cdot|\bar{s})$ is a finite p.d. on S. Set

$$r(\cdot|\bar{s}_n)=\prod_{t=1}^{n} r(\cdot|\bar{s}^t)\qquad \text{for } \bar{s}_n=(\bar{s}^1,\dots,\bar{s}^n).$$

Any element $P(\cdot|\cdot\bar{s}_n)\in\bar{\mathfrak{C}}_n$ can be approximated arbitrarily closely by expressions of the form

$$\sum_{s_n\in S_n} r(s_n|\bar{s}_n)\,P(\cdot|\cdot|\bar{s}_n).$$

Therefore, for every $y_n\in Y_n$ and $m\in M$, $P(y_n|(f_m^1, f_m^2(y^1), \dots, f_m^n(y^1, \dots, y^{n-1}))|\,\bar{s}_n)$ can be approximated arbitrarily closely by expressions of the form

$$\sum_{s_n} r(s_n|\bar{s}_n)\,P(y_n(f_m^1, \dots, f_m^n(y^1, \dots, y^{n-1}))|s_n).$$

Hence, $P(A_m|f_n(m)|s_n)\geqq 1-\lambda$ for $m\in M$ and $s_n\in S_n$ implies

$$P(A_m|f_n(m)|\bar{s}_n)\geqq 1-\lambda \qquad \text{for } m\in M,$$

$\bar{s}_n\in\bar{S}_n$. The converse implication is obvious. This proves part a) of the lemma.

If $\{(f_n(m), A_m)|m=1, \dots, N\}$ is an (n, N, λ) code for $\mathfrak{A}_{1f}$, then

$$P(A_m|f_n(m)|s_n)\geqq 1-\lambda>0 \qquad \text{for } m\in M, \qquad s_n\in S_n.$$

Since $\mathfrak{C}$ contains only $0-1$-matrices, we conclude that $P(A_m|f_n(m)|s_n)=1$ for $m\in M$, $s_n\in S_n$. Since $\bar{w}(\cdot|\cdot)\in\bar{\mathfrak{C}}$, part a) implies $\bar{P}(A_m|f_n(m))=1$ for $m\in M$. The converse implication is immediate from the definition of $\bar{w}$. It was proved in [7] that $\mathfrak{A}$ (resp. $\bar{\mathfrak{A}}$ or $\bar{\bar{\mathfrak{A}}}$) has a positive rate if and only if the following condition (K.W.) holds:

there exists an $i\in X$ and an $i'\in X$ such that $T(i)\cap T(i')=\emptyset$. For a.v.ch.f. we have

Lemma 3. a) (K.W.) *is sufficient for* $\bar{\bar{\mathfrak{A}}}_f$ *to have a positive rate*[1].

b) (K.W) *is necessary and sufficient for* $\mathfrak{A}_{1f}$ (resp. $\bar{\mathfrak{A}}_{1f}$ *or* $\bar{\bar{\mathfrak{A}}}_{1f}$) *to have a positive rate.*

Proof. Part a) follows from the result quoted above. It remains to show that (K.W.) is necessary in case b). If (K.W.) does not hold, then any two row-vectors of $\bar{w}$ have a common support and hence the zero-error capacity C_{0f} of $\bar{\mathscr{D}}_f$ equals 0. (This was noticed in [8], p. 17.) It follows now from Lemma 2 that $\mathfrak{A}_{1f}$ and consequently also $\bar{\mathfrak{A}}_{1f}$ and $\bar{\bar{\mathfrak{A}}}_{1f}$ have capacity 0.

Lemma 4 (see [4], Lemma 4).

$$\bar{\bar{K}}=\max_{\pi}\min_{w\in\bar{\bar{\mathfrak{C}}}} R(\pi, w)=\min_{w\in\bar{\bar{\mathfrak{C}}}}\max_{\pi} R(\pi, w).$$

Proof. It is known that $R(\pi, w)$ is concave in π for each w and convex in w for each π. $\bar{\bar{\mathfrak{C}}}$ and $\{\pi\}$ are norm compact convex sets and $R(\pi, w)$ is norm continuous in both variables. Therefore the minimax theorem is applicable and yields the equality.

3. The Capacity for $\mathfrak{A}_f$ and an Optimal Coding Scheme

In [1] we presented an optimal coding scheme for the d.m.c.f. The scheme is *not* sequential (encoding functions of variable length) and consists in an *iterative* procedure to reduce the list of possible messages on the receiver's side. The iteration is made possible because of the feedback. The present results for $\bar{\bar{\mathfrak{A}}}_f$

[1] U. Augustin has informed us about an example which shows that the condition is not necessary in this case.

concern again codes of fixed block length as defined in (1.4) and they are obtained by the very same iterative approach as described in [1]. The bounds on N, L and λ of Lemmas 1, 2, and 3 of [1] are now replaced by those in our Lemma 1. In order to make this paper a self-entity repition of parts of [1] cannot be avoided. Before we come to the coding scheme we derive first an upper bound on $\bar{\bar{C}}_f$.

Lemma 4 yields that $\bar{\bar{K}}$ equals $\min\limits_{w\in\bar{\bar{\mathfrak{C}}}} \max\limits_{\pi} R(\pi, w)$. Let w' be such that $\bar{\bar{K}} = \max\limits_{\pi} R(\pi, w')$, let $\mathscr{D}'_f$ be the d.m.c.f. corresponding to w', and denote it's capacity by C'_f. The strong converse of the coding theorem for $\mathscr{D}'_f$ (Kempermann [6] and Kesten (oral communication), published also in [9]) says that:

(3.1) for $\delta>0$ and any λ, $0\leqq\lambda<1$, there does not exist a code $(n, \exp\{(C'_f+\delta)\,n\}, \lambda)$ for all sufficiently large n.

Since $w'\in\bar{\bar{\mathfrak{C}}}$ and since $C'_f=\max\limits_{\pi} R(\pi, w')=\bar{\bar{K}}$, we obtain that

$$\bar{\bar{C}}_f \leqq \bar{\bar{K}}. \tag{3.2}$$

Assume now that $\bar{\bar{C}}_f>0$ and hence $\bar{\bar{K}}>0$. (Example 1 below shows that $\bar{\bar{K}}$ can be positive and still $\bar{\bar{C}}_f=0$.) Choose π such that $\bar{\bar{K}}=\min\limits_{w\in\bar{\bar{\mathfrak{C}}}} R(\pi, w)$. Abbreviate $H(\pi)$ as H and $\max\limits_{w\in\bar{\bar{\mathfrak{C}}}} \sum\limits_{j} q_j(w)\, H(w^*(\cdot|j))$ as $\overline{H}$. With this notation we can write $\bar{\bar{K}}$ as $H-\overline{H}$.

We describe now our coding scheme. Let r be a positive integer and let $M_1=\{1, \ldots, a^r\}$ be a set of $N=a^r$ messages. Choose l_1 as the smallest integer for which $|X_{l_1}(\pi)|\geqq a^r$. It follows from Lemma 1, a) that for $l_1\geqq c_1(\varepsilon)$:

$$H^{-1}\log a\cdot r+g_2(\varepsilon)\, r\geqq l_1\geqq H^{-1}\log a\cdot r-g_2(\varepsilon)\, r, \tag{3.3}$$

where $g_2(\varepsilon)$ can be given explicitly and $\lim\limits_{\varepsilon\to 0} g_2(\varepsilon)=0$.

We now map M_{1_l} one to one into $X_{l_1}(\pi)$ and call the image $\overline{X}_{l_1}(\pi)$. Let $u=(f_m^1, \ldots, f_m^{l_1})$ be the image of m, $m\in M_1$. For $m\in M_1$ and $t=1, 2, \ldots, l_1$ we now define $f_m^t(Z^1, \ldots, Z^{t-1})$ by

$$f_m^t(Z^1, \ldots, Z^{t-1})=f_m^t. \tag{3.4}$$

Suppose the sender is sending message m and he has already sent the letters $f_m^1, \ldots, f_m^{l_1}$. The receiver has received a sequence $v=(v^1, \ldots, v^{l_1})\in Y_{l_1}$ and decides on the list $M_2=X_{l_1}(v, \varepsilon, \pi)$. It follows from Lemma 1 that u is on this list with a probability $1-\lambda_1\geqq\exp\{-E(\varepsilon)\, l_1\}$ and that $|X_{l_1}(v, \varepsilon, \pi)\leqq\exp\{\overline{H}+g_1(\varepsilon)\, l_1\}$ for $l_1\geqq c_1(\varepsilon)$. The v received and therefore also the list M_2 is known to the sender, because we have feedback. If u is not on the list, then we count this as a decoding error and it is irrelevant how the sender continues the transmission (over the fixed block length n, to be determined later). If u is in M_2, then we iterate the procedure as follows. Let l_2 be the smallest integer such that $|X_{l_2}(\pi)|\geqq\exp\{\overline{H}+g_1(\varepsilon)\, l_1\}$. It follows from Lemma 1 that

$$\left(\frac{\overline{H}}{H}-g_3(\varepsilon)\right) l_1\leqq l_2\leqq\left(\frac{\overline{H}}{H}+g_3(\varepsilon)\right) l_1 \quad \text{for } l_1\geqq c_2(\varepsilon)\geqq c_1(\varepsilon), \tag{3.5}$$

where $g_3(\varepsilon)$ and $c_2(\varepsilon)$ are known functions and $\lim\limits_{\varepsilon\to 0} g_3(\varepsilon)=0$.

H is positive, because $\bar{\bar{K}}=H-\bar{H}>0$ by assumption. Moreover, $0\leqq\frac{\bar{H}}{H}<1$. We now map M_2 one to one into $X_{l_2}(\pi)$ and call the image $\bar{X}_{l_2}(\pi)$. This mapping depends on v, is otherwise arbitrary and is known to sender and receiver. Let $(f_m^{l_1+1},\dots,f_m^{l_1+l_2})$ be the image of $(f_m^1,\dots,f_m^{l_1})\in M_2$. For $m\in M$ and $t=l_1+1,\dots,l_1+l_2$ we define $f_m^t(Z^1,\dots,Z^{t-1})$ by

$$f_m(Z^1,\dots,Z^{t-1})=f_m^t. \tag{3.6}$$

After these l_2 letters have been sent we come up with a set M_3, defined analogously to M_2. For $l_2\geqq c_2(\varepsilon)$ the image of m is contained in M_3 with a probability

$$1-\lambda_2\geqq 1-\exp\{-E(\varepsilon)\,l_2\}.$$

Set $K(\varepsilon)=\frac{\bar{H}}{H}+g_3(\varepsilon)$ and $\bar{K}(\varepsilon)=\frac{\bar{H}}{H}-g_3(\varepsilon)$. By iterating the procedure for $s=3,4,\dots$ we obtain

$$\bar{K}(\varepsilon)\,l_{s-1}\leqq l_s\leqq K(\varepsilon)\,l_{s-1} \qquad \text{for all } s \text{ with } l_{s-1}\geqq c_2(\varepsilon). \tag{3.7}$$

Since $K(\varepsilon)<1$ for ε sufficiently small, we thus constantly reduce the number of possible messages on the receiver's side. However, the inequality. $l_{s-1}\geqq c_2(\varepsilon)$ imposes a bound on the number of steps we can take in the described way.

Let D be the smallest integer such that

$$l_D<c_2(\varepsilon)\leqq l_{D-1}. \tag{3.8}$$

Since $\bar{K}(\varepsilon)^{D-1}\,l_1\leqq l_D<c_2(\varepsilon)\leqq l_{D-1}\leqq K(\varepsilon)^{D-2}\,l_1$, we obtain

$$(D-1)\log\bar{K}(\varepsilon)+\log l_1<\log c_2(\varepsilon)\leqq(D-2)\log K(\varepsilon)+\log l_1 \tag{3.9}$$

and from the last inequality that

$$D\leqq g_4(\varepsilon)\log l_1 \qquad \text{for } l_1\geqq c_3(\varepsilon)\geqq c_2(\varepsilon), \tag{3.10}$$

where $g_4(\varepsilon)$ and $c_3(\varepsilon)$ can be given explicitly.

If we would follow the scheme up to $s=D$, then we would be left with fewer than a^{l_D} messages on the receiver's side. Later we shall discuss how to seperate the message m sent from a "small" set of messages. Presently we are concerned about the error probabilities $\lambda_s(s=1,\dots,D)$ involved in the scheme. Since the l_s's are decreasing the error probabilities λ_s increase with s. In order to keep them small two changes are necessary in the scheme above. First of all we want to exclude that λ_s exceeds $\frac{1}{2}$. Therefore we define for any constant β, $0<\beta<\frac{1}{2}$, an integer D_1 as the largest integer smaller than D for which

$$\exp-\{E(\varepsilon)\,l_{D_1}\}\leqq\beta \tag{3.11}$$

and we restrict s to the set $\{1,\dots,D_1\}$. We assert

$$l_{D_1}\leqq L(\varepsilon)=\min\bigl(|\log\beta|\,[E(\varepsilon)\,\bar{K}(\varepsilon)]^{-1},c_2(\varepsilon)\bigr). \tag{3.12}$$

The inequality clearly holds for $D_1=D$ because of (3.8). For $D_1<D$ we conclude from (3.11) that $l_{D_1}\geqq|\log\beta|\,E(\varepsilon)^{-1}>l_{D_1}+1$.

This and $l_{D_1}+1\geqq\bar{K}(\varepsilon)\,l_{D_1}$ imply the inequality in this case.

Secondly, in order to keep the error probability of the scheme—which is bounded by the sum of the error probabilities at each step—small, we iterate only $d=d(l_1)$ times, where d is the largest integer such that

$$l_d \geqq l_1^{\frac{1}{2}} \quad \text{for } l_1 \geqq c_4(\varepsilon)=\max\left(c_3^2(\varepsilon), L^2(\varepsilon)\right). \tag{3.13}$$

This definition and (3.12) imply that $d \leqq D_1$. The decoding error probability after d steps is bounded by $\sum_{s=1}^{d} \lambda_s$, which is smaller than $d \exp\{-E(\varepsilon)\, l_d\}$. This, $d \leqq D_1$ and (3.13) imply

$$\sum_{s=1}^{d} \lambda_s \leqq D_1 \cdot \exp\{-E(\varepsilon)\, l_1^{\frac{1}{2}}\} \quad \text{for } l_1 \geqq c_4(\varepsilon). \tag{3.14}$$

For the remaining steps $(s=d+1, \ldots, D_1)$ we have by definition of d:

$$l_s < l_1^{\frac{1}{2}}. \tag{3.15}$$

We achieve small error probabilities for these steps by *repeating* each step $[l_1^{\frac{1}{4}}]$ times. To be more specific, let us assume that at the step $s=d+1$ the sender has sent the sequence $u^*=(f_m^{l_1+\cdots+l_d+1}, \ldots, f_m^{l_1+\cdots+l_d+l_{d+1}})$ and the receiver has decided according to the scheme on $M_{d+2}=M_{d+2}(1)$ as list of possible messages. Now the sender sends the *same* u^* again and he keeps doing this $[l_1^{\frac{1}{4}}]$ times. At instant v; $v=1, \ldots, [l_1^{\frac{1}{4}}]$; one obtains a list of messages $M_{d+2}(v)$, say. All messages, which occur on more than half of the lists, shall form the final list $\overline{M}_{d+2}$ at step $s=d+1$. Thus

$$\overline{M}_{d+2}=\{u \mid u \in M_{d+2}(v) \text{ for more than } \tfrac{1}{2}[l_1^{\frac{1}{4}}] \text{ of the } v_s'\}. \tag{3.16}$$

Obviously,

$$|\overline{M}_{d+2}| \leqq 2 \max_{v} |M_{d+2}(v)|. \tag{3.17}$$

For any v; $v=1, \ldots, [l_1^{\frac{1}{4}}]$; u^* is contained in $M_{d+2}(v)$ with a probability α greater than $1-\lambda_{d+1} \geqq 1-\beta > \frac{1}{2}$. Since the channel is memoryless, we obtain that u^* is in $\overline{M}_{d+2}$ with a probability $1-\bar{\lambda}_{d+1}$, where

$$1-\bar{\lambda}_{d+1} \geqq \sum_{v=[\frac{1}{2} l_1^{\frac{1}{4}}]}^{[l_1^{\frac{1}{4}}]} \binom{[l_1^{\frac{1}{4}}]}{v} \alpha^v (1-\alpha)^{[l_1^{\frac{1}{4}}]-v} \geqq 1-\exp\{-H(\alpha, 1-\alpha)\, l_1^{\frac{1}{4}}\}. \tag{3.18}$$

We apply now the same procedure to the steps $s=d+2$, $s=d+3, \ldots, s=D_1$ and thus finally come up with a list $\overline{M}_{D_1}$, where

$$|\overline{M}_{D_1}| \leqq a^{l_{D_1}}. \tag{3.19}$$

In so far we have used only $\overline{\overline{K}}>0$. In order to "seperate" message m from the remaining elements in $\overline{M}_{D_1}$ we need *now* the assumption that $\overline{\overline{C}}_f>0$. This assumption implies that there exists a $\left(l_0(\varepsilon), a^{l_{D_1}}, \alpha_1\right)$ code for $\overline{\overline{\mathfrak{A}}}_f$, where $\alpha_1<\frac{1}{2}$. If we send every codeword of this code $[l_1^{\frac{1}{4}}]$ times, then we decrease the error probability to $\alpha_2 \leqq \exp\{-H(\alpha_1, 1-\alpha_1)\, l_1^{\frac{1}{4}}\}$. This concatenated code can be used to reduce $\overline{M}_{D_1}$ to one element.

The probability λ that this is not an image of message m satisfies

$$\begin{aligned}(3.20)\qquad \lambda < \sum_{s=1}^{d} \lambda_s + \sum_{s=d+1}^{D_1} \lambda_s + \alpha_2 \leqq D_1 \exp\{-E(\varepsilon)\, l_1^{\frac{1}{2}}\} \\ + D_1 \exp\{-H(\alpha, 1-\alpha)\, l_1^{\frac{1}{4}}\} + \exp\{-H(\alpha_1, 1-\alpha_1)\, l_1^{\frac{1}{4}}\}.\end{aligned}$$

The *total* number n of letters sent is less than

$$l_1(1+K(\varepsilon)+K^2(\varepsilon)+\cdots)+l_1^{\frac{1}{2}}\, l_1^{\frac{1}{4}}\, g_4(\varepsilon)\log l_1 + l_0(\varepsilon)\, l_1^{\frac{1}{4}}$$

and therefore

$$(3.21)\qquad n \leqq l_1(1+g_5(\varepsilon))(1-K(\varepsilon))^{-1} \qquad \text{for } l_1 \geqq c_5(\varepsilon),$$

where $g_5(\varepsilon)$ and $c_5(\varepsilon)$ are known functions and $\lim_{\varepsilon\to 0} g_5(\varepsilon)=0$.

Consequently, $l_1 \geqq (1-K(\varepsilon)(1+g_5(\varepsilon))^{-1} n$. This, $K(\varepsilon)=\dfrac{\bar{H}}{H}+g_3(\varepsilon)$, and (3.3) imply

$$(3.22)\qquad r \geqq (H^{-1}\log a + g_2(\varepsilon))^{-1}\, l_1 \geqq \log^{-1} a\,(H-\bar{H}-g_6(\varepsilon))\, n,$$

where $g_6(\varepsilon)$ is a known function and $\lim_{\varepsilon\to 0} g_6(\varepsilon)=0$.

Since $N=a^r$ and since $\bar{\bar{K}}=H-H$, we finally obtain

$$(3.23)\qquad N=\exp\{r\cdot\log a \geqq \exp\{\bar{\bar{K}}\, n - g_6(\varepsilon)\, n\}$$

for $n \geqq c_6(\varepsilon)$, a known function.

Assuming that condition (K.W.) holds one can easily *construct* a code $(l_0(\varepsilon), a^{l_{D_1}}, \alpha_1)$ for $\bar{\bar{\mathfrak{A}}}$ and hence for $\bar{\bar{\mathfrak{A}}}_f$. Thus the final step of our coding procedure is also constructive. We summarize the results in the following theorem.

Theorem. *Suppose that $T(i)\cap T(i')=\emptyset$ for some $i, i'\in X$, then*

a) *The capacity $\bar{\bar{C}}_f$ of $\bar{\bar{\mathfrak{A}}}_f$ is positive and equals $\bar{\bar{K}}$.*

b) *Given R, $0<R<\bar{\bar{K}}$, then one can compute an $E(R)$ and an $n_0(R)$ such that for every $n\geqq n_0(R)$ one can give explicitly a code of length $N=e^{Rn}$ such that the decoding error probability λ is smaller than* $\exp\{-E(R)\, n^{\frac{1}{4}}\}$.

The following two examples supplement the results.

Example 1. Let $w(\cdot|\cdot|1)=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}$, $w(\cdot|\cdot|2)=\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}$ and let $\mathfrak{C}=\{w(\cdot|\cdot|s)\,|\,s=1,2\}$

One easily verifies that in this case $\bar{\bar{K}}=\log\frac{3}{2}>0$. However, it follows from Lemma 3, b) that $\bar{\bar{C}}_f=0$. If we replace $w(\cdot|\cdot|2)$ by $\begin{pmatrix}\varepsilon&1-2\varepsilon&\varepsilon\\0&0&1\\1&0&0\end{pmatrix}$, then $T(1)\cap T(2)=\emptyset$ and the theorem yields for the capacity ${}_{\varepsilon}\bar{\bar{C}}_f$ of the new channel: ${}_{\varepsilon}\bar{\bar{C}}_f>0$ and $\lim_{\varepsilon\to 0}{}_{\varepsilon}\bar{\bar{C}}_f=\log\frac{3}{2}$. This shows that $\bar{\bar{C}}_f$ is discontinuous as function of the matrices (in canonical topologies).

Example 2. Let $w(\cdot|\cdot|1)=\begin{pmatrix}1 & 0\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$, $w(\cdot|\cdot|2)=\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ 0 & 1\end{pmatrix}$, $S=\{1,2\}$ and let $\mathfrak{C}=\{w(\cdot|\cdot|s)|s\in S\}$. Since $\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}\in\bar{\bar{\mathfrak{C}}}$, $\bar{C}_f=0$. Let $n=2$ and define two encoding functions $f_2(1)=[f_1^1, f_1^2\,(Z^1)]$ and $f_2(2)=[f_2^1, f_2^2(Z^1)]$ by

$$f_1^1=f_1^2(0)=0, \quad f_1^2(1)=1; \quad f_2^1=f_2^2(0)=1, \quad f_2^2(1)=0.$$

We describe the transmission probabilities for all $s_2=(s^1,s^2)\in S_2$

a) $s_2=(1,1)$

$$\begin{array}{c|cccc} & 00 & 01 & 10 & 11\\ \hline f_2(1) & 1 & 0 & 0 & 0\\ f_2(2) & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0\end{array}$$

b) $s_2=(1,2)$

$$\begin{array}{c|cccc} & 00 & 01 & 10 & 11\\ \hline f_2(1) & \frac{1}{2} & \frac{1}{2} & 0 & 0\\ f_2(2) & 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4}\end{array}$$

c) $s_2=(2,1)$

$$\begin{array}{c|cccc} & 00 & 01 & 10 & 11\\ \hline f_2(1) & \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4}\\ f_2(2) & 0 & 0 & 1 & 0\end{array}$$

d) $s_2=(2,2)$

$$\begin{array}{c|cccc} & 00 & 01 & 10 & 11\\ \hline f_2(1) & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2}\\ f_2(2) & 0 & 0 & \frac{1}{2} & \frac{1}{2}\end{array}$$

We show that $T(f_2(1))\cap T(f_2(2))=\emptyset$.

A vector in $T(f_2(1))$ is in it's first component $\geqq\frac{1}{4}$, equality holds only for the vector $(\frac{1}{4},\frac{1}{4},0,\frac{1}{2})$. The only vector in $T(f_2(2))$, which has a first component not smaller than $\frac{1}{4}$ is $(\frac{1}{4},\frac{1}{4},\frac{1}{2},0)$. But the 2 vectors are different. $\bar{\mathfrak{A}}_f$ has a positive rate. $\bar{\mathfrak{A}}$ has capacity 0, because $T(1)\cap T(2)\neq\emptyset$. Thus, feedback increases the capacity and $\bar{\bar{C}}_f$ and $\bar{C}_f$ are not equal.

Remark. We provide some justification for limiting ourselves to the channel $\bar{\bar{\mathfrak{A}}}_f$. A.v.ch. are a model for a transmission system which has several states and varies arbitrarily from one state to another. 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 a.v.ch.-model. There are two essentially different ways in which the system can operate:

1) The sequence of states $s_n=(s^1,\ldots,s^n)$ is selected arbitrarily but independent of the messages to be transmitted and the letters to be sent.

2) At every instant t s^t may depend on all letters sent up to $t-1$ and eventually also on the letter to be sent at instant t.

In the second case we have an unrestricted variability of states and it seems to us that this is the case closer to applications. It was shown in [5] that in case of maximal errors and no feedback the coding problems for the two cases are mathematically equivalent. In case of feedback those problems are no longer equivalent. $\bar{\mathfrak{A}}_f$ is the appropriate model for situation 1). In situation 2) s^t may now—because of the feedback—depend not only on the message to be sent, but also on the letters *received* up to $t-1$. Let us denote this channel by $\bar{\bar{\bar{\mathfrak{A}}}}_f$, without having stated the transmission functions formally. $\bar{\bar{\bar{\mathfrak{A}}}}_f$ is different from $\bar{\bar{\mathfrak{A}}}_f$. In the later channel the s_n may depend on the messages, but not on the letters actually sent. This channel is simply of type $\bar{\mathfrak{A}}_f$ with an enlarged class of matrices. Every (n,N,λ) code for $\bar{\bar{\bar{\mathfrak{A}}}}_f$ is certainly an (n,N,λ) code for $\bar{\bar{\mathfrak{A}}}_f$. The converse is not true as can be seen from footnote 1.) and (3.24) below. However, our coding scheme works for $\bar{\bar{\mathfrak{A}}}_f$ as well, because Lemma 1 is independent of feedback and still applies. Therefore condition (K.W.) is also sufficient for $\bar{\bar{\mathfrak{A}}}_f$ to have positive rate. Moreover we have:

(3.24) (K.W.) is *necessary* for $\bar{\bar{\mathfrak{A}}}_f$ to have a positive rate[2]. This can be seen as follows.

Suppose that for every $i, i' \in X$: $T(i) \cap T(i') \neq \emptyset$. Then for every i, i' there is a $s(i, i') \in S^e \subset \bar{\bar{S}}$ such that $w(\cdot|i|s(i, i')) \equiv w(\cdot|i'|s(i, i'))$. Let

$$f_n(m) = (f_m^1, \ldots, f_m^n(Z^1, \ldots, Z^{n-1})) \quad \text{and} \quad f_n(m') = (f_{m'}^1, \ldots, f_{m'}^n(Z^1, \ldots, Z^{n-1}))$$

be any two encoding functions of any code. Choose s^1 such that $w(\cdot|f_m^1|s^1) \equiv w(\cdot|f_{m'}^1|s^1)$ and define s^t, $t = 2, \ldots, n$, inductively as follows. Suppose any sequence $y^1, \ldots, y^t$ has been received and $f_m^{t+1}(y^1, \ldots, y^t) = x^{t+1}$, $f_{m'}^{t+1}(y^1, \ldots, y^t) = x'^{t+1}$. Then set $s^{t+1} = s(x^{t+1}, x'^{t+1})$. Clearly, the code's probability of error cannot be made smaller than $\frac{1}{2}$.

4. The Proof of a Conjecture of Shannon

Let $\mathscr{D}_f$ be a d.m.c.f. given by a stochastic $a \times b$-matrix w. We denote it's zero-error capacity by $C_{0f}(w)$. The following result is due the Shannon ([8], Theorem 7).

Theorem S. *For $j \in Y$ define $S_j = \{i | i \in X, w(j|i) > 0\}$ and set $\pi_0 = \min\limits_{\pi} \max\limits_{j} \sum\limits_{i \in S_j} \pi_i$. Then*

a) $C_{0_f}(w) = \log \pi_0^{-1}$, *if* $C_{0_f}(w) > 0$,

b) $C_{0_f}(w) = 0$ *if and only if no 2 row vectors in w have disjoint support.*

Define a set of matrices $V(w)$ by

(4.1) $V(w) = \{w' | w'$ *stochastic, for any pair* (i, j): $w'(j|i) = 0$ *if* $w(j|i) = 0\}$ *and set* $C_{\min} = \max\limits_{\pi} \min\limits_{w' \in V(w)} R(\pi, w')$.

Shannon conjectured ([8], p. 19) *that*

$$C_{0_f}(w) = C_{\min} \quad \textit{if} \quad C_{0_f}(w) > 0. \tag{4.2}$$

The similarity between the formula for $C_{\min}$ and the formula for $\bar{K}$ [see (2.15)] is apparent. Define $\mathfrak{C} = \mathfrak{C}(w)$ by

(4.3) $\mathfrak{C} = \{w' | w'$ stochastic 0-1-matrix, for any (i, j): $w'(j|i) = 0$ if $w(j|i) = 0\}$.

One easily verifies that

$$\bar{\bar{\mathfrak{C}}} = V(w). \tag{4.4}$$

Therefore the conjecture (4.2) *can be restated as*

$$\mathfrak{C}_{0_f}(w) = \max_{\pi} \min_{w' \in \bar{\bar{\mathfrak{C}}}} R(\pi, w'), \quad \textit{if} \quad C_{0_f}(w) > 0. \tag{4.5}$$

We prove now (4.5). The result is an immediate consequence of our Theorem and Lemma 2. Two matrices w and $\tilde{w}$ are said to be adjacent, if for any (i, j): $w(j(i) > 0$ when and only when $\tilde{w}(j|i) > 0$. It is easy to see that d.m.c.f.'s which correspond to adjacent matrices have the same zero-error capacities, that is

[2] This observation was made by the reviewer, to whom our thanks are due.

$C_{0_f}(w)=C_{0_f}(\tilde{w})$ (see [8]). For $\mathfrak{C}$ as in (4.3) define $\bar{\bar{\mathfrak{C}}}$, $\mathfrak{C}^e$, S^e, $\mathfrak{A}^e_{1f}$, $\bar{\bar{\mathfrak{A}}}_{1f}$, C^e_{1f} and $\bar{\bar{C}}_{1f}$ as in section 1. S^e is finite and $\bar{\mathfrak{C}}^e=\bar{\bar{\mathfrak{C}}}$.

Let $w^e(\cdot|\cdot)$ be a stochastic matrix given by

$$w^e(j|i)=|S^e|^{-1}\sum_{s\in S^e} w(j|i|s) \quad \text{for } i\in X,\ j\in Y. \tag{4.6}$$

It follows from the definition of $\mathfrak{C}^e$ and from (4.6) that w and w^e are adjacent.

Therefore

$$C_{0_f}(w)=C_{0_f}(w^e). \tag{4.7}$$

It follows from Lemma 2, b) that

$$C_{0_f}(w^e)=C^e_{1f}. \tag{4.8}$$

Since $\bar{\mathfrak{C}}^e=\bar{\bar{\mathfrak{C}}}$, we obtain from Lemma 2, a) that

$$C^e_{1f}=\bar{\bar{C}}_{1f}. \tag{4.9}$$

(4.7), (4.8) and (4.9) yield

$$C_{0_f}(w)=\bar{\bar{C}}_{1_f}. \tag{4.10}$$

(4.5) follows now from (4.10) and the Theorem. We thus have proved the conjecture.

References

1. Ahlswede, R.: A constructive proof of the coding theorem for discrete memoryless channels with noiseless feedback. (To appear in the Transaction of the Sixth Prague Conf. on Information Theory, Statistical Decision Functions and Random Processes.)
2. 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. Statistics, **41**, 3, 1027–1033 (1970).
3. Ahlswede, R.: The capacity of a channel with arbitrarily varying Gaussian channel probability functions. (To appear in the Transactions of the Sixth Prague Conference on Inf. Th., Stat. Dec. Fct's and Rand. Proc.)
4. Ahlswede, R., Wolfowitz, J.: The capacity of a channel with arbitrarily varying channel probability functions and binary output alphabet. Z. f. Wahrscheinlichkeitstheorie verw. Geb. **15**, 186–194 (1970).
5. Ahlswede, R., Wolfowitz, J.: Correlated decoding for channel with arbitrarily varying channel probability functions. Inform. and Control **14**, 451–473 (1969).
6. Kempermann, J.H.B.: Strong converses for a general memoryless channel with feedback. (To appear in the Trans. of the Sixth Prague Conf. on Inf. Th., Stat. Dec. Fct's and Rand Proc.)
7. Kiefer, J., Wolfowitz, J.: Channels with arbitrarily varying channel probability functions. Inform. and Control **5**, 44–54 (1962).
8. Shannon, C.E.: The zero-error capacity of a noisy channel. IRE Trans. Inform. Theory, IT-**2**, 8–19 (1956).
9. Wolfowitz, J.: Coding Theorems of Information Theory. Heidelberg: Springer, Second ed., 1964.

R. Ahlswede
Ohio State University
Department of Mathematics
Columbus, Ohio 43210
USA

(Received May 4, 1971/in revised form October 20, 1972)