

# The Weak Capacity of Averaged Channels

R. AHLWEDE

Received December 27, 1967

## Contents

1. Introduction . . . . .	61
2. Definitions . . . . .	62
3. A Channel without Strong Capacity . . . . .	63
4. The Weak Capacity of an Averaged Discrete Channel. . . . .	66
5. The Weak Capacity of an Averaged Semicontinuous Channel . . . . .	67
6. Nonstationary Averaged Channels . . . . .	68
7. Averages of Channels with Respect to General Probability Distributions . . . . .	69

*Summary.* Coding theorem and weak converse of the coding theorem are proved for averaged semicontinuous stationary channels and for almost periodic discrete channels, whose phases are statistically known. Explicit formulas for the capacities are given. The strong converses of the coding theorems do not hold.

## 1. Introduction

Averaged channels were introduced by JACOBS [5]. In this class of channels he found the first example of a channel for which the strong converse of the coding theorem does not hold, but the coding theorem and the weak converse of the coding theorem hold (unpublished). The proof uses STRASSEN's version of SHANNON's method of random codes. The disadvantage of this method is that the channel capacity cannot be computed. In [9] WOLFOWITZ gave simpler examples of channels without strong capacity and he gave the reason for which the strong converse of the coding theorem fails to hold, but he does not show if these channels have a weak capacity. In 3. we give a simple example of a channel without strong capacity, but with weak capacity. The proof of the coding theorem and its weak converse uses classical results for simultaneous channels [8], [3]. The main idea consists in a comparison of codes for an averaged channel with its corresponding simultaneous channel. This leads to a method which is representative for the more complicated proofs in par. 3-par. 6, where we deal with coding theorem and weak converse for averaged channels under different assumptions on the time structure (stationary, almost periodic, nonstationary) and the output alphabet (finite, nonfinite) of the channel. It is remarkable that we need the strong converse for simultaneous channels for the proof of weak converses for averaged channels. This demonstrates the importance of the concept of a strong converse — which is due to WOLFOWITZ [10], [8] — even to coding theorists who are only interested in weak converses. The results can be formulated roughly as follows:

The weak capacity of the averaged discrete stationary channel equals the strong capacity of the corresponding simultaneous channel (Theorem 2).

The weak capacity of the averaged semicontinuous stationary channel is in general greater than the weak capacity of the corresponding simultaneous channel (Theorem 3).

The weak capacity of an almost periodic discrete channel, whose phase is statistically known, is greater than or equal to the strong capacity of the corresponding simultaneous channel. Equality holds if each phase has positive probability (Corollary of Theorem 4).

In all these cases we average with respect to a discrete p. d. In par. 7 we prove the coding theorem and the weak converse for stationary discrete averaged channels, where the average is taken with respect to a general p. d. The proof is based on the proof for Theorem 3 and several new ideas.

## 2. Definitions

Let  $X^t = X = \{1, \dots, a\}$  for  $t = 1, 2, \dots$  and  $(X'^t, \mathcal{Q}'^t) = (X' \mathcal{Q}')$  for  $t = 1, 2, \dots$  where  $X'$  is an arbitrary set and  $\mathcal{Q}'$  is a  $\sigma$ -algebra of subsets in  $X'$ . Furthermore let  $S = \{s, \dots\}$  be a nonempty (index) - set and let  $F^t(\cdot | 1 | s), \dots, F^t(\cdot | a | s)$  be probability distributions (p. d.) on  $(X^t, \mathcal{Q}'^t)$  ( $t \in N, s \in S$ ). For each  $x_n = (x^1, \dots, x^n)$   $\cdot X_n = \prod_{t=1}^n X^t$  we define a p. d. on

$$\left( X'_n = \prod_{t=1}^n X'^t, \mathcal{Q}'_n = \prod_{t=1}^n \mathcal{Q}'^t \right) \text{ by } F_n(\cdot | x_n | s) = \prod_{t=1}^n F^t(\cdot | x^t | s).$$

The sequence of kernels  $(F_n(\cdot | \cdot | s))_{n=1, 2, \dots}$  forms a semicontinuous (in general nonstationary) channel without memory. [In case  $X'^t = X'$  is finite, the kernels  $F^t(\cdot | \cdot | s)$  are given by stochastic matrices  $w^t(k | i | s) = F^t(\{k\} | i | s)$  ( $i \in X, k \in X'$ ). We speak then of a discrete channel without memory.] Thus we have assigned to each  $s \in S$  a semicontinuous channel. If we are interested in the simultaneous behaviour of all these channels, then we call this indexed set of channels a simultaneous channel (semicontinuous, without memory). Common properties of the individual channels are assigned to the simultaneous channel: stationarity, almost periodicity, etc. as follows. The set  $\{F_n(\cdot | \cdot | s) | s \in S\}$  designated by  $S_n$ , is called simultaneous channel  $S_n$  in the discrete time-interval  $\langle 1, n \rangle$ .

A code  $(n, N, \lambda)$  for the simultaneous channel  $S_n$  is a series of pairs

$$\{(u_1, A'_1) \cdots (u_N, A'_N)\},$$

where  $u_i \in X_n, A_i \in \mathcal{Q}'_n$  for  $i = 1, \dots, N, A_i \cap A_j = \emptyset$  for  $i \neq j$  and  $F_n(A'_i | u_i | s) \geq 1 - \lambda$  ( $0 < \lambda < 1$ ) for  $i = 1, 2, \dots, N, s \in S$  (sometimes instead of the code  $(n, N, \lambda)$  we speak of a  $\lambda$ -code.  $N$  is the length of the code  $(n, N, \lambda)$ .  $N(n, \lambda)$  is the maximal length of a  $\lambda$ -code in  $\langle 1, n \rangle$ ).

Let  $\mathfrak{D}_n = \langle D_1, \dots, D_b \rangle$  be a partition of  $X'_n$  in finitely many elements of  $\mathcal{Q}'_n$  and let  $\mathfrak{B}_n$  be the set of all such finite partitions.  $\mathfrak{P}_n$  is the set of all p. d. on  $(X_n, \mathcal{Q}_n)$ , where  $\mathcal{Q}_n$  is the  $\sigma$ -algebra of all subsets of  $X_n$ .

$$R(p_n, \mathfrak{D}_n, s) = \sum_{i=1, \dots, b} \sum_{x_n \in X_n} p_n(x_n) F_n(D_i | x_n | s) \log \frac{F_n(D_i | x_n | s)}{\sum_{y_n \in X_n} p_n(y_n) F_n(D_i | y_n | s)}$$

is the (channel)-rate for  $p_n, \mathfrak{D}_n, F_n(\cdot|\cdot|s)$ .

Kesten [6] has shown by proving the coding theorem and the weak converse that

$$C_1 = \lim_{n \rightarrow \infty} \frac{1}{n} \sup_{p_n \in \mathfrak{P}_n} \sup_{\mathfrak{D}_n \in \mathfrak{B}_n} \inf_{s \in S} R(p_n, \mathfrak{D}_n, s)$$

is the weak capacity of the stationary semicontinuous memoryless channel.

$$C = \max_{p^1} \inf_{s \in S} R(p^1, F(\cdot|\cdot|s)) = \max_{p^1} \inf_{s \in S} \sum_{X^1} p^1(x^1) \int \log \frac{dF(\cdot|x^1|s)}{\sum_{Y^1 \in X^1} p^1(y^1) dF(\cdot|y^1|s)} dF(\cdot|x^1|s)$$

is the strong capacity of the stationary discrete memoryless simultaneous channel. This was proved by WOLFOWITZ [8]. KEMPERMAN gave an example of a semicontinuous simultaneous channel with  $C_1 < C$  (published in [6]). However,  $C$  is the weak capacity of the averaged semicontinuous channel, defined by:

$$P_n(\cdot|x_n) = \sum_{s \in S} q_s F_n(\cdot|x_n|s),$$

where  $q$  is a p.d. on a countable index set  $S$  and  $(F_n(\cdot|\cdot|s))$ ,  $n = 1, \dots$ , are semicontinuous stationary simultaneous channels without memory (Theorem 3).

### 3. A Channel without Strong Capacity

Given  $X = X' = \langle 1, \dots, a \rangle$  and the stochastic matrices  $w(\cdot|\cdot|1)$ ,  $w(\cdot|\cdot|2)$  with  $a$  rows and columns. For  $s = 1, 2$ , we define the discrete memoryless channel (d.m.c.)  $(P_n(\cdot|\cdot|s))$   $n = 1, 2, \dots$  by

$$P_n(x'_n|x_n|s) = \prod_{t=1}^n w(x'^t|x^t|s) \quad \text{for all } x_n \in X_n, x'_n \in X'_n, \quad n = 1, 2, \dots$$

and the averaged channel (a.c.)  $(P_n(\cdot|\cdot))$   $n = 1, 2, \dots$  by

$$P_n(x'_n|x_n) = \frac{1}{2} P_n(x'_n|x_n|1) + \frac{1}{2} P_n(x'_n|x_n|2) \quad (x_n \in X_n, x'_n \in X'_n, n = 1, 2, \dots)$$

**Theorem 1** (Coding theorem and weak converse). *Let*

$$\begin{aligned} C &= \max_p \inf_{s=1,2} R(p, w(\cdot|\cdot|s)) \\ &= \text{strong capacity of the s.c. } (S_n) \quad n = 1, 2, \dots \end{aligned}$$

For the maximal code length  $N(n, \lambda)$  of the a.c.  $(P_n)$   $n = 1, \dots$  the following estimates hold:

a) Given  $0 < \lambda < 1$ ,  $\delta > 0$ , then there exists an  $n_0 = n_0(\lambda, \delta)$  such that

$$N(n, \lambda) > e^{(C+\delta)n} \quad \text{for } n \geq n_0.$$

b) Given  $\delta > 0$ , then there exists a  $\lambda$  and an  $n_0 = n_0(\lambda, \delta)$  such that

$$N(n, \lambda) < e^{(C+\delta)n} \quad \text{for } n \geq n_0.$$

*Proof of part a)* A  $\lambda$ -code for the simultaneous channel

$$S_n = \{P_n(\cdot|\cdot|s) | s = 1, 2\} \quad \text{is a } \lambda\text{-code for } P_n(\cdot|\cdot).$$

b) Choose  $\lambda < \frac{1}{2}$ . For a  $\lambda$ -code  $\{(u_i, A'_i) | i = 1, \dots, N\}$  of  $P_n$  we have

$$P_n(A'_i|u_i) \geq 1 - \lambda, \quad i = 1, \dots, N$$

and therefore

$$P_n(A'_i | u_i | s) \geq 1 - \lambda - \frac{1}{2} = 1 - \lambda' > 0 \quad (s = 1, 2; i = 1, \dots, N).$$

A  $\lambda$ -code for  $P_n$  is a  $\lambda'$ -code for  $S_n$ , if  $\lambda < \frac{1}{2}$ .

By the strong converse for  $S_n$  we have  $N(n, \lambda) \leq e^{Cn+k(\lambda')\sqrt{n}}$ , where  $k(\lambda')$  is a known function [10]. This proves b).

**Remark 1.** The strong converse of the coding theorem gives, if it holds, an estimate of the following type:

Given  $0 < \lambda < 1$ ,  $\delta > 0$ , then there exists an  $n_0(\lambda, \delta)$  such that for  $n \geq n_0$

$$N(n, \lambda) < e^{(C+\delta)n}$$

holds, where  $C$  is the maximal constant for which estimation a) holds. (For nonstationary channels we have instead of  $C$  a capacity function  $C_n$ .)

Choose  $w(\cdot | \cdot | s)$  such that  $(P_n(\cdot | \cdot | 1))$   $n = 1, 2, \dots$  has capacity 0 and  $(P_n(\cdot | \cdot | 2))$   $n = 1, 2, \dots$  has capacity  $C_2 > 0$ . Then a fortiori  $C = 0$ . A  $\lambda$ -code for  $\frac{1}{2}P_n(\cdot | \cdot | 2)$  is a  $\lambda$ -code  $P_n$ . Choose  $\lambda > \frac{1}{2}$  and let  $N_2(n, \lambda)$  be the maximal code length for  $P_n(\cdot | \cdot | 2)$ . Thus we have  $N(n, \lambda) \geq N_2\left(n, \frac{1+\lambda}{2}\right) > e^{C_2-k'(\lambda)\sqrt{n}}$  for all  $n$ . Thus the strong converse does not hold.

**Remark 2.** A weaker form of the strong converse can be introduced; namely, there exists an  $\alpha > 0$ , such that for all  $\lambda$  with  $0 < \lambda < \alpha$ ,  $\delta > 0$  there exists an  $n_0(\lambda, \delta)$  with the property:

$$N(n, \lambda) < e^{(C+\delta)n}; \quad n \geq n_0.$$

For the channel we discussed above this estimate holds with  $\alpha = \frac{1}{2}$ . However, for the channels considered in 4 even this weaker estimate does not hold.

**Remark 3.** For several stationary channels (e.g. discrete channel without memory, discrete channel with finite memory)  $\lim_{n \rightarrow \infty} \frac{1}{n} R(p_\infty, F_n)$  exists for each stationary p.d.  $p_\infty$  on

$$(X_\infty, \mathcal{Q}_\infty) = \left( \prod_{t=1}^{\infty} X^t, \prod_{t=1}^{\infty} \mathcal{Q}^t \right) \quad \text{and} \quad C_{\text{stat.}} \stackrel{\text{def}}{=} \sup_{p_\infty} \lim_{n \rightarrow \infty} \frac{1}{n} R(p_\infty, F_n)$$

turns out to be the (weak or strong) channel capacity  $C$ .

The proofs of the coding theorem work mostly with ergodic stationary p.d. In the first step it is shown that the joint source-channel distribution is ergodic under suitable conditions on the channel. Then McMILLAN's theorem gives that the information function  $\frac{1}{n} I(p_\infty, F_n)$  converges to

$$\lim_{n \rightarrow \infty} \frac{1}{n} R(p_\infty, F_n) \quad (L^1 \text{ or with probability one}).$$

In the last step an application of FEINSTEIN's maximal code theorem leads to the coding theorem. The channel defined in 1) has the remarkable property

that

$$\begin{aligned}
 (*) \quad C &< \max_{\substack{p \times p \times \dots \times p \\ \text{independent} \\ \text{sources}}} \lim_{n \rightarrow \infty} \frac{1}{n} R(p \times \dots \times p, P_n) \\
 &\leq \max_{\substack{p_{\text{ergodic}} \\ \text{stat.}}} \lim_{n \rightarrow \infty} \frac{1}{n} R(p_{\infty}, R_n) \stackrel{\text{def}}{=} C_{\text{erg.}} \leq \max_{p_{\text{stat.}}} \lim_{n \rightarrow \infty} \frac{1}{n} R(p_{\infty}, P_n) \stackrel{\text{def}}{=} C_{\text{stat.}}
 \end{aligned}$$

Therefore the usual method of proving a coding theorem is not applicable.

Some authors already speak of channel capacity if  $C_{\text{stat.}}$  exists — without having proved the coding theorem and a converse. (\*) shows that this is not admissible. Let us now proof (\*). It is known, that the mean entropy  $\bar{H}$  of a stationary p.d. on the product measure-space has the linearity property:

(\*\*) Given a p.d.  $q_1, \dots, q_k$  and stationary p.d.  $p^1, \dots, p^k$  on  $(X_{\infty}, \Omega_{\infty})$ , then

$$\bar{H} \left( \sum_{i=1}^k q_i p_i \right) = \sum_{i=1}^k q_i \bar{H}(P^i).$$

We now choose  $p_n = p \times \dots \times p$  for  $n = 1, 2, \dots$  and define

$$\begin{aligned}
 p'_n(x'_n | s) &= \sum_{x_n \in X_n} p_n(x_n) P_n(x'_n | x_n | s) \quad s = 1, 2, \dots; x'_n \in X'_n \\
 \tilde{p}_n(x'_n | x_n | s) &= p_n(x_n) P_n(x'_n | x_n | s) \quad s = 1, 2, \dots; x_n \in X_n; x'_n \in X'_n \\
 p'_n(x'_n) &= \sum_{s=1}^2 \frac{1}{2} p'_n(x'_n | s) \\
 \tilde{p}_n(x'_n, x_n) &= \sum_{s=1}^2 \frac{1}{2} \tilde{p}_n(x'_n, x_n | s).
 \end{aligned}$$

From  $R(p_n, P_n) = H(p_n) + H(p'_n) - H(\tilde{p}_n)$  and the above definitions, we conclude that

$$\lim_{n \rightarrow \infty} \frac{1}{n} R(p_n, P_n) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{s=1}^2 \frac{1}{2} [H(p_n) + H(p'_n(\cdot | s)) - H(\tilde{p}_n(\cdot | s))].$$

Using (\*\*) and the independence we get

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \frac{1}{n} R(p_n, P_n) &= \sum_{s=1}^2 \frac{1}{2} [H(p_1) + H(p'_1(\cdot | s)) - H(\tilde{p}_1(\cdot | s))] \\
 &= \sum_{s=1}^2 \frac{1}{2} R(p_1, w(\cdot | s))
 \end{aligned}$$

and therefore

$$\max_{p_n = p \times p \times \dots \times p} \lim_{n \rightarrow \infty} \frac{1}{n} R(p_n, P_n) = \max_{|p} \frac{1}{2} R(p, w(\cdot | 2)) = \frac{1}{2} C_2 > C.$$

In [4] HU GUO DING proved that the coding theorem and the weak converse hold if and only if the channel is “information stable” (Def. given in [4]). Information stability, however, is difficult to verify for our channel and it is even more difficult to find a formula for the channel capacity by this method.

#### 4. The Weak Capacity of an Averaged Discrete Channel

Given a set of stochastic matrices

$$\left\{ (w(ij|s))_{\substack{i=1,\dots,a \\ j=1,\dots,a}} \mid s \in S = \{1, 2, \dots\}, \quad X = X' = \{1, \dots, a\} \right\}$$

and a p.d.  $q$  on  $S$ . We can assume without loss of generality that  $q_s > 0$  for all  $s \in S$ . Otherwise we would take instead of  $S$  a subset  $S'$  with  $q'_s > 0$ ,  $s' \in S'$  and  $\sum_{s' \in S'} q'_s = 1$ .

We define the individual discrete channels by

$$P_n(x'_n | x_n | s) = \prod_{t=1}^n w(x'^t | x^t | s) \quad x'_n \in X'_n, \quad x_n \in X_n, \quad s \in S, \quad n = 1, 2, \dots;$$

the averaged discrete channel by

$$P_n(x'_n | x_n) = \sum_{s \in S} q_s P_n(x'_n | x_n | s)$$

and the simultaneous channel by

$$S_n = \{P_n(\cdot | \cdot | s) \mid s \in S\} \quad n = 1, 2, \dots$$

We need the following:

**Lemma 1.** *If  $f_n$  is a decreasing sequence of continuous, convex, nonnegative functions defined on a simplex in  $R^n$ , then  $\lim_{n \rightarrow \infty} f_n = f$  is continuous and the convergence is uniform.*

*Sketch of the proof.* It is clear that the limit exists.  $f$  is convex and could have discontinuities only on the extreme points, but this does not happen. From DINI's theorem we conclude uniform convergence.

**Lemma 2.** a)  $f_k(p) = \inf_{s=1,\dots,k} R(p, F(\cdot | \cdot | s))$  converges uniformly in  $p$  to  $\inf_{s \in S} R(p, F(\cdot | \cdot | s))$ .

$$\begin{aligned} \text{b) } \lim_{k \rightarrow \infty} C_k &= \lim_{k \rightarrow \infty} \max_p \inf_{s=1,\dots,k} R(p, F(\cdot | \cdot | s)) \\ &= \max_p \inf_{s \in S} R(p, F(\cdot | \cdot | s)). \end{aligned}$$

*Proof.* The set of all p.d. on  $\{1, \dots, a\}$  forms a simplex.  $R(p, F(\cdot | \cdot | s))$  as a function of  $p$  is continuous, convex and non-negative.  $f_k(p)$  satisfies the hypotheses of Lemma 1. This gives us a) and as an easy consequence b) (cf. [8]).

We can now formulate:

**Theorem 2** (*Coding theorem and weak converse for the discrete averaged channel*  $(P_n)_{n=1,2,\dots}$ ). *Let  $C = \max_p \inf_{s \in S} R(p, w(\cdot | \cdot | s))$ . Then the following estimates hold:*

a) *Given  $0 < \lambda < 1$ ,  $\delta > 0$ , then there exists an  $n_0 = n_0(\lambda, \delta)$ , such that*

$$N(n, \lambda) > e^{(C-\delta)n} \quad \text{for } n \geq n_0.$$

b) *Given  $\delta > 0$ , there exists a  $\lambda > 0$  and an  $n_0 = n_0(\lambda, \delta)$ , such that*

$$N(n, \lambda) < e^{(C+\delta)n} \quad \text{for } n \geq n_0.$$

*Proof.* Part a). A  $\lambda$ -code for the simultaneous channel  $S_n$  is a  $\lambda$ -code for  $P_n$ . The statement follows from the coding theorem for simultaneous stationary channels without memory [8].

Part b). Define

$$\varepsilon_k = \sum_{\alpha=k+1} q_\alpha, \quad \eta_k = \inf_{\alpha=1, \dots, k} q_\alpha > 0.$$

(0) For the given  $\delta > 0$  choose  $k$  such that  $|C_k - C| \leq \delta/2$ , then choose  $\lambda = \eta_k/2$  and  $\lambda' = \varepsilon_k + \frac{1}{2} \eta_k$ .

A  $\lambda$ -code  $\{(u_i, A'_i) | i = 1, \dots, N\}$  for  $P_n$  is a  $\lambda'$ -code for  $\sum_{s=1}^k q_s P_n(\cdot | \cdot | s)$ , since

$$(1) \quad \sum_{s=1}^k q_s P_n(A'_i | u_i | s) \geq P_n(A'_i | u_i) - \varepsilon_k \geq 1 - (\frac{1}{2} \eta_k + \varepsilon_k).$$

But

$$(2) \quad \sum_{s=1}^k q_s P_n(A'_i | u_i | s) - q'_s P_n(A'_i | u_i | s') \leq 1 - \varepsilon_k - \eta_k \quad \text{for } s' = 1, \dots, k.$$

From (1) and (2) we have

$$q_s P_n(A'_i | u_i | s) \geq \frac{1}{2} \eta_k \quad \text{for } s = 1, \dots, k, \quad i = 1, \dots, N$$

and therefore

$$P_n(A'_i | u_i | s) \geq \frac{1}{2} \eta_k \quad \text{for } s = 1, \dots, k, \quad i = 1, \dots, N.$$

Now we apply the *strong* converse of the coding theorem for simultaneous discrete channels without memory and conclude that

$$(3) \quad N(n, \lambda) < e^{n C_k} + k(\eta_k) \sqrt{n} \quad \text{for all } n.$$

Statement b) follows from (0) and (3).

Remarks. 1. The proof uses essentially the strong converse for simultaneous discrete channels without memory. FANO's Lemma would lead to the estimate

$$N(n, \lambda) < e^{(n C_k + 1) \eta_k^{-1}}, \quad \text{but } \lim_{k \rightarrow \infty} \frac{C_k}{\eta_k} = \infty.$$

2. From the proof of the theorem it follows that the weak capacity remains unchanged if we average with respect to  $q^*$ , where  $q^*$  is equivalent to  $q$ .

### 5. The Weak Capacity of an Averaged Semicontinuous Channel

We return to the semicontinuous case as described in § 2.

**Theorem 3** (*Coding theorem and weak converse for the semicontinuous averaged channel  $(P_n)_{n=1,2,\dots}$* ). Let  $C = \max_p \inf_{s \in S} R(p, F(\cdot | \cdot | s))$ , then the following estimates hold:

a) Given  $0 < \lambda < 1$ ,  $\delta > 0$ , then there exists an  $n_0 = n_0(\lambda, \delta)$ , such that

$$N(n, \lambda) > e^{(C-\delta)n} \quad \text{for } n \geq n_0.$$

b) Given  $\delta > 0$ , there exists a  $\lambda > 0$  and an  $n_0 = n_0(\lambda, \delta)$ , such that

$$N(n, \lambda) < e^{(C+\delta)n} \quad \text{for } n \geq n_0.$$

*Proof.* a) Given  $0 < \lambda < 1$ ,  $\delta > 0$ . Choose  $k$  such that  $\varepsilon_k < \lambda$  and define  $\lambda_k := \frac{\lambda - \varepsilon_k}{1 - \varepsilon_k} < 1$ . A  $\lambda_k$ -code for  ${}_k S_n = \{P_n(\cdot | \cdot | s) | s = 1, \dots, k\}$  is a  $\lambda$ -code for  $P_n$ . The coding theorem for  ${}_k S_n$  gives

$$e^{C_k n - k(\lambda_k) \sqrt{n}} \leq N_k(n, \lambda_k) \leq N(n, \lambda)$$

and therefore  $N(n, \lambda) > e^{n(C - \delta)}$  for  $n$  sufficiently large.

b) If we use Lemma 2 and the strong converse of the coding theorem for  ${}_k S_n$ , the proof of Theorem 2 b) carries over verbatim.

**Remark.** An example of KEMPERMAN (published in [6]) shows that there are semicontinuous channels with  $C > C_1$ .  $C_1$  is the (weak) capacity of  $(S_n)_{n=1,2,\dots}$ ;  $C$  is the weak capacity of  $(P_n)_{n=1,2,\dots}$ . Therefore we can give the intuitive interpretation: The statistical knowledge of the individual channels which govern the transmission increases the weak capacity.

## 6. Nonstationary Averaged Channels

Given stochastic  $a \times a$ -matrices  $w^t(\cdot | \cdot | s)$ ,  $s \in S$ ,  $t = 1, 2, \dots$ , we define the nonstationary simultaneous channel (cf. [1])

$$\begin{aligned} (S_n)_{n=1,2,\dots} &= \{P_n(\cdot | \cdot | s) | P_n(x'_n | x_n | s) \\ &= \prod_{t=1}^n w^t(x^t | x^t | s), x_n \in X_n, x'_n \in X'_n, s \in S\} \quad n = 1, 2, \dots \end{aligned}$$

Let  $q = (q_1, q_2, \dots)$  be a discrete p.d. on the countable set  $S$  and  $q_s > 0$  for all  $s \in S$ .

Define

$$C_k(n) = \max_{p_n} \inf_{s=1,\dots,k} R(p_n, P_n(\cdot | \cdot | s)) \quad \text{and} \quad C(n) = \inf_{k=1,2,\dots} C_k(n).$$

**Theorem 4.** *If for the nonstationary averaged channel  $(P_n)_{n=1,2,\dots}$  the condition: (1) For each  $\delta_1 > 0$  there exists a  $k$  and an  $n_1(\delta_1, k)$  such that*

$$|C(n) - C_k(n)| < \delta_1 n \quad \text{for} \quad n \geq n_1(\delta_1, k)$$

*holds, then we have the estimates*

a) *Given  $0 < \lambda < 1$ ,  $\delta > 0$ , then there exists an  $n_0 = n_0(\lambda, \delta)$ , such that*

$$N(n, \lambda) > e^{C(n) - \delta n} \quad \text{for} \quad n \geq n_0.$$

b) *Given  $\delta > 0$ , then there exists a  $\lambda > 0$  and an  $n_0 = n_0(\lambda, \delta)$ , such that*

$$N(n, \lambda) < e^{C(n) + \delta n} \quad \text{for} \quad n \geq n_0$$

(cf. [1] Einleitung).

*Proof.* a) Given  $0 < \lambda < 1$ ,  $\delta > 0$ , choose  $k$  such that  $\lambda_k := \frac{\lambda}{1 - \varepsilon_k} < 1$ . A  $\lambda_k$ -code for  ${}_k S_n = \{P_n(\cdot | \cdot | s) | s = 1, \dots, k\}$  is a  $\lambda$ -code for  $P_n$ . The coding theorem for  ${}_k S_n$  gives

$$e^{C_k(n) - \delta n} \leq N_k(n, \lambda_k) \leq N(n, \lambda) \quad \text{for} \quad n$$

sufficiently large (Satz 2 in [1], Chapter III).



b) Using (1) and the strong converse of the coding theorem for  $kS_n$  (Satz 3 in [1], Chapter III) the proof of Theorem 2b) carries over verbatim.

**Example.** Almost periodic discrete averaged channels. Let  $(w^t(\cdot \cdot))_{t=1, 2, \dots}$  be an almost periodic sequence of stochastic matrices (cf. [1], Chapter II), then we can define the simultaneous almost periodic channel

$$(S_n)_{n=1, 2, \dots} = \{P_n(\cdot | \cdot | s) | P_n(x'_n | x_n | s) \\ = \prod_{t=1}^n w^{t+s}(x'^t | x^t), x_n \in X_n, x'_n \in X'_n, s = 0, 1, \dots\} \quad n = 1, 2, \dots$$

and the averaged almost periodic channel

$$(P_n(\cdot \cdot))_{n=1, 2, \dots} = \left( \sum_{s=0} q_s P_n(\cdot | \cdot | s) \right) \quad n = 1, 2, \dots$$

From Theorem 4 we conclude the

**Corollary.** For the almost periodic averaged channel  $(P_n)_{n=1, 2, \dots}$  the coding theorem and the weak converse hold with

$$C = \lim_{n \rightarrow \infty} \frac{1}{n} \max_{(p^1, \dots, p^n) = p_n} \inf_{s=0, 1, \dots} R(p_n | P_n(\cdot | \cdot | s)) = \lim_{n \rightarrow \infty} \frac{1}{n} C_n.$$

*Proof.* We have to show that (1) is satisfied. But this follows from the almost periodicity and the normcontinuity of  $R(p, w)$  in  $w$  as is shown in [1], Chapter II, p. 2. The Bedingung 1, there is exactly the same as (1). It follows from the definition of  $C_n$  that  $\frac{1}{n} C_n$  is monotone nondecreasing and  $\frac{1}{n} C_n \leq \log a$ ; hence,  $\lim_{n \rightarrow \infty} \frac{1}{n} C_n = C$ . The capacity function is constant.

**Remarks.** 1.  $C$  can be greater than

$$\max_p \inf_{t=1, 2, \dots} R(p, w^t(\cdot \cdot)).$$

**Example.** Choose two stochastic matrices  $w(\cdot | \cdot | 1), w(\cdot | \cdot | 2)$  with

$$R(p, w(\cdot | \cdot | 1)) \equiv 0 = C_1 \quad \text{and} \quad \max_p R(p, w(\cdot | \cdot | 2)) = C_2 > 0.$$

Define

$$w^{2s}(\cdot | \cdot) = w(\cdot | \cdot | 2) \\ w^{2s-1}(\cdot | \cdot) = w(\cdot | \cdot | 1) \quad \text{for } s = 1, 2, \dots$$

Then  $(w^t(\cdot \cdot))_{t=1, 2, \dots}$  is a periodic sequence of stochastic matrices. The corresponding simultaneous periodic channel  $(S_n)_{n=1, 2, \dots}$  has the capacity

$$C = \frac{1}{2} C_2 > \max_p \inf_{t=1, 2, \dots} R(p, w^t(\cdot \cdot)) = 0.$$

2. The corollary says: if we know the phase of an almost periodic channel statistically and each phase has positive probability, then the (weak) capacity of this channel equals the (strong) capacity for  $(S_n)_{n=1, 2, \dots}$ .

[Coding theorem and strong converse for  $(S_n)_{n=1, 2, \dots}$  were proved in [1].] The statistical knowledge of the phase increases the maximal code length in such a way that instead of the strong converse only a weak converse holds.

If  $q_s$  is not positive for all  $s$ , then the capacity of the averaged channel can of course be greater than  $C$ .

Example. Choose  $w(\cdot|\cdot|1)$ ,  $w(\cdot|\cdot|2)$  such that

$$R(p, w(\cdot|\cdot|1)) \quad \text{and} \quad R(p, w(\cdot|\cdot|2))$$

have their maximum for different arguments  $p_1$ ,  $p_2$  and

$$R(p_1, w(\cdot|\cdot|1)) > R(p_1, w(\cdot|\cdot|2))$$

$$R(p_2, w(\cdot|\cdot|2)) > R(p_2, w(\cdot|\cdot|1)).$$

For  $q_1 = 1$  the averaged channel is the periodic channel  $(w^t(\cdot|\cdot))_{t=1, 2, \dots}$  and has a capacity greater than the capacity of the corresponding simultaneous channel.

### 7. Averages of Channels with Respect to General Probability Distributions

Until now we considered averages with discrete probability distributions. What happens, if we take averages with respect to nondiscrete p.d., for instance, the Lebesguemeasure on  $[0, 1]$ ?

I. Let us look at a discrete averaged stationary channel with  $S = [0, 1]$ ,  $q =$  Lebesguemeasure on  $S$ .

Define

$$P_n(\cdot|\cdot) = \int_{[0,1]} P_n(\cdot|\cdot|s) q(ds).$$

Of course,  $w(i|j|s)$  has to be measurable in  $s$  for  $i, j = 1, \dots, a$ . In this case our method from 4. is not applicable as can be seen by the following example:

$$w(i|j|s) = 1 \quad i = 1, \dots, a, \quad s \in (0, 1]$$

$$w(i|j|0) = 1/a \quad i, j = 1, \dots, a.$$

Then  $(P_n)_{n=1, 2, \dots}$  has strong capacity  $\log a$ , because  $N(n, \lambda) = e^{n \log a}$ ; but  $(S_n)_{n=1, 2, \dots}$  has capacity 0, because  $w(\cdot|\cdot|0)$  has capacity 0. We have to give another approach.

II. Let  $(S, \mathcal{C}, q)$  be a normed measure space,  $S$  is index-set as usual. Divide  $[0, 1]$  in disjoint intervals of length  $\beta$ .  $s_1$  and  $s_2$  are  $\beta$ -equivalent, if  $w(i|j|s_1)$  and  $w(i|j|s_2)$  are in the same interval for all  $i, j = 1, \dots, a$ . This equivalence relation leads to a partition of  $S$  in at most  $(1/\beta)^{a^2}$  measurable sets

$${}_l S (l = 1, \dots, L(\beta)) \leq (1/\beta)^{a^2}$$

and therefore to a partition of  $S_n$  in the sets

$${}_l S_n = \{P_n(\cdot|\cdot|s) | s \in {}_l S\}, \quad l = 1, \dots, L(\beta).$$

For  $0 < \alpha < 1$  define

$$(1) \quad C(\alpha, \beta) = \max_p \sup_{\substack{l_1, \dots, l_k \leq L(\beta) \\ q\left(\bigcup_{i=1}^k {}_i S\right) \geq 1-\alpha}} \inf_{s \in \bigcup_{i=1}^k {}_i S} R(p, w(\cdot|\cdot|s))$$

Instead of  $\max \sup$  we can write  $\max \max$ , because we vary over a finite set of index-constellations.

Furthermore we define

$$C(\alpha) = \max_p \sup_{\{S' \mid S' \subset S, q(S') \geq 1 - \alpha\}} \inf_{s \in S'} R(p, w(\cdot \mid \cdot \mid s)) \quad \text{and} \quad C = \inf_{\alpha > 0} C(\alpha).$$

It follows from the definitions that

$$(2) \quad C(\alpha, \beta) \leq C(\alpha) \quad \text{for all } \beta.$$

**Theorem 5.** *For the general stationary discrete averaged channel*

$$(P_n(\cdot \mid \cdot))_{n=1,2,\dots} = \left( \int_S P_n(\cdot \mid \cdot \mid s) q(ds) \right)_{n=1,2,\dots}$$

the following estimates hold with  $C = \inf_{\alpha > 0} C(\alpha)$

a) Given  $0 < \lambda < 1$ ,  $\delta > 0$ , then there exists an  $n_0(\lambda, \delta)$  such that:

$$N(n, \lambda) > e^{Cn - \delta n} \quad \text{for } n \geq n_0.$$

b) Given  $\delta > 0$ , then there exists a  $\lambda > 0$  and an  $n_0 = n_0(\lambda, \delta)$ , such that:

$$N(n, \lambda) < e^{Cn + \delta n} \quad \text{for } n \geq n_0.$$

*Proof.* a) Given  $\lambda, \delta > 0$ , choose  $\alpha < \lambda$  and  $S'$  such that  $q(S') \geq 1 - \alpha$  and

$$\left| \max_p \inf_{s \in S'} R(p, w(\cdot \mid \cdot \mid s)) - C(\alpha) \right| \leq \delta/2.$$

Define  $\lambda' = \frac{\lambda - \alpha}{1 - \alpha}$ . A  $\lambda'$ -code for  $S'_n$  is a  $\lambda$ -code for  $P_n$ , because  $(1 - \lambda')(1 - \alpha) = 1 - (1 - \alpha)\lambda' - \alpha = 1 - \lambda$ . Hence

$$N(n, \lambda) \geq N'(n, \lambda') \geq e^{(C(\alpha) - (\delta/2))n - K(\lambda')\sqrt{n}} \geq e^{(C - \delta)n}$$

for  $n$  sufficiently large.

b) First of all choose  $\alpha$  such that  $|C(\alpha) - C| \leq \delta/2$ . Let  ${}_{i_1}S \cdots {}_{i_k}S$  be a family of sets such that the maximum is attained in (1), then  $q\left(\bigcup_{i=1}^k {}_{i_i}S\right)$  has to be greater than  $1 - \alpha$ . We define  $1 - \varepsilon(\alpha, \beta) = q({}_{i_1}S \cup \cdots \cup {}_{i_k}S)$ .

$\varepsilon(\alpha, \beta)$  is by definition smaller or equal than  $\alpha$ . Define now

$$\eta(\alpha, \beta) = \inf_{i=1,\dots,k} q({}_{i_i}S) > 0$$

and choose

$$\lambda = \frac{\eta(\alpha, \beta)}{2}, \quad \lambda' = \varepsilon(\alpha, \beta) + \frac{1}{2}\eta(\alpha, \beta).$$

Then a  $\lambda$ -code for  $P_n$  is a  $\lambda'$ -code for

$$\int_{{}_{i_1}S \cup \cdots \cup {}_{i_k}S} P_n(\cdot \mid \cdot \mid s) q(ds),$$

since

$$(2) \quad \int_{{}_{i_1}S \cup \cdots \cup {}_{i_k}S} P_n(A'_i \mid u_i \mid s) \geq P_n(A'_i \mid u_i) - \varepsilon(\alpha, \beta) \geq 1 - (\frac{1}{2}\eta(\alpha, \beta) + \varepsilon(\alpha, \beta)).$$

But from (2) and the definitions given above, it follows that

$$(3) \quad \int_{{}_{i_1}S \cup \cdots \cup {}_{i_k}S} P_n(A'_i \mid u_i \mid s) q(ds) - \int_{{}_{i_j}S} P_n(A'_i \mid u_i \mid s) q(ds) \leq 1 - \varepsilon(\alpha, \beta) - \eta(\alpha, \beta) \quad \text{for } i = 1, \dots, N; j = 1, \dots, k.$$

From (2), (3) we have

$$(4) \quad \int_{\mathcal{U}^S} P_n(A'_i | u_i | s) q(ds) \geq \frac{1}{2} \eta(\alpha, \beta) \quad \text{for } j = 1, \dots, k; i = 1, \dots, N.$$

We need the trivial

**Combinatorial Lemma.** *Let  $B_i$ ,  $i = 1, \dots, I$  be measurable sets with  $q(B_i) \geq \gamma > 0$  ( $i = 1, \dots, I$ ). If we define  $m$  as the maximal number of sets  $B_i$  with a common element, then the estimate  $m \geq \gamma I$  holds.*

Denote by  $D_{ji}$  the set

$$\{s | P_n(A'_j | u_j | s) \geq \frac{1}{4} \eta(\alpha, \beta), s \in \mathcal{U}^S\} \quad (j = 1, \dots, N; i = 1, \dots, k).$$

It follows from (4) that

$$q(D_{ji}) \geq \frac{1}{4} \eta(\alpha, \beta) \quad (j = 1, \dots, N; i = 1, \dots, k).$$

The sets  $D_{11}, D_{21}, \dots, D_{N1}$  satisfy the hypothesis of the Lemma. Hence, there exists an element  $s_1 \in \mathcal{U}^S$  which is contained in at least  $\frac{1}{4} \eta(\alpha, \beta) N$  of these sets. That means there exists a subcode of length  $\frac{1}{4} \eta(\alpha, \beta) N$  of the code

$$\{(u_i, A'_i) | i = 1, \dots, N\}$$

such that

$$P_n(\bar{A}'_{i1} | \bar{u}'_{i1} | s_1) \geq \frac{1}{4} \eta(\alpha, \beta) \quad \text{for } i_1 = 1, \dots, [\frac{1}{4} \eta(\alpha, \beta) N].$$

Apply now the same arguments to  $D_{12}, \dots, D_{N2}$ . Thus we find a subcode of our subcode which is now a simultaneous code for  $P_n(\cdot | \cdot | s_1)$  and  $P_n(\cdot | \cdot | s_2)$  of length greater than  $\frac{1}{4} \eta(\alpha, \beta) \cdot \frac{1}{4} \eta(\alpha, \beta) \cdot N$ . Proceeding in the same way we have after  $k \leq L(\beta)$  steps a subcode of length  $N^*$  greater than  $(\frac{1}{4} \eta(\alpha, \beta))^{L(\beta)} \cdot N$  with

$$P_n(A_j^* | u_j^* | s_i) \geq \frac{1}{4} \eta(\alpha, \beta), \quad j = 1, \dots, N^*; i = 1, \dots, k.$$

From the strong converse of the coding theorem for simultaneous channels and the normcontinuity of  $R(p, w)$  in  $w$  uniformly in  $p$  we conclude:

$$N(\frac{1}{4} \eta(\alpha, \beta))^{L(\beta)} \leq N^* \leq \exp C(\alpha, \beta) n + f(\beta) n + K(\eta) \sqrt{n} \quad \text{where } \lim_{\beta \rightarrow 0} f(\beta) = 0.$$

Using (2) we have

$$N(n, \lambda) \leq \exp C(\alpha) n + f(\beta) n + K(\eta) \sqrt{n} + L(\beta) \log \frac{1}{4} \eta(\alpha, \beta).$$

Choose now  $\beta$  such that  $f(\beta) \leq \delta/4$  and use  $|C(\alpha) - C| \leq \delta/2$ . Then we have  $N(n, \lambda) \leq e^{Cn + \delta n}$  for  $n$  sufficiently large. That proves b).

**Remark 1.** Theorem 5 can be extended to the semicontinuous case, if

$$\begin{aligned} C(\alpha) &= \max_p \sup_{\{S' | q(S') \geq 1-\alpha\}} \inf_{s \in S'} R(p, F(\cdot | \cdot | s)) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \max_n \sup_{p_n} \sup_{\{S' | q(S') \geq 1-\alpha\}} \sup_{\mathcal{D}_n} \inf_{s \in S'} R(p_n, \mathcal{D}_n, P_n(\cdot | \cdot | s)) \end{aligned}$$

for all  $\alpha > 0$ . Part a) follows then from the coding theorem for simultaneous semicontinuous channels [6]. For the proof of part b) we use that for an arbitrarily set of channel kernels  $\{F(\cdot | \cdot | s) | s \in \mathcal{S}\}$  the corresponding set of information functions  $\{R(p, F(\cdot | \cdot | s)) | s \in \mathcal{S}\}$  is totally bounded in the norm of uniform convergence. (This is a consequence of Hilfssatz 1 in [1], Chapter I, par. 4.) Hence,

we can find for given  $\beta > 0$  a family of sets  $\{jS^* | j = 1, \dots, L(\beta)\}$  such that for  $s_1, s_2 \in jS^*$

$$\sup_p |R(p, F(\cdot | \cdot | s_1)) - R(p, F(\cdot | \cdot | s_2))| \leq \beta \quad (j = 1, \dots, L(\beta)).$$

If we redefine the  $jS$ , which we used in the proof of Theorem 5, as follows:

$$jS := jS^*$$

then, the proof of part b) carries over to the semicontinuous case.

**Remark 2.** The extension of Theorem 5 to the nonstationary case seems to be difficult. It could be of interest for the "arbitrarily varying channel"-problem [7].

The author wishes to thank Professor JACOBS for putting at his disposal an unpublished manuscript concerned with averaged channels. Furthermore, he wishes to acknowledge a helpful remark of Professor BAJANSKI.

### References

1. AHLWEDE, R.: Beiträge zur Shannonschen Informationstheorie im Falle nichtstationärer Kanäle. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **10**, 1—42 (1968).
2. — Zur starken Umkehrung des Codingtheorems für halbstetige Simultankanäle. unpublished.
3. BLACKWELL, D., L. BREIMANN, and I. THOMASIAN: The capacity of a class of channels. *Ann. math. Statistics* **30**, 1229—1241 (1960).
4. HU GUO DING: On the informationstability of a sequence of channels. *Theor. Probab. Appl.* **7**, 258—269 (1962).
5. JACOBS, K.: Almost periodic channels. *Colloquium on Combinatorial Methods in Probability Theory*. Aarhus 1962.
6. KESTEN, H.: Some remarks on the capacity of compound channels in the semicontinuous case. *Inform. and Control* **4**, 169—184 (1961).
7. KIEFER, J., and J. WOLFOWITZ: Channels with arbitrarily varying channel. *Probability functions*. *Inform. and Control* **5**, 44—54 (1962).
8. WOLFOWITZ, J.: Simultaneous channels. *Arch. rat. Mech. Analysis* **4**, 371—386 (1960).
9. — Channels without capacity. *Inform. and Control* **6**, 49—54 (1963).
10. — The coding of messages subject to chance errors. *Illinois J. Math.* **1**, 591—606 (1957).

Professor R. AHLWEDE  
The Ohio State University  
Department of Mathematics  
231 W. 18th Avenue  
Columbus, Ohio 43210, USA