COLLOQUIA MATHEMATICA SOCIETATIS JÁNOS BOLYAI
16. TOPICS IN INFORMATION THEORY, KESZTHELY (HUNGARY), 1975.

TWO CONTRIBUTIONS TO INFORMATION THEORY

R. AHLSWEDE* — P. GÁCS

1. INTRODUCTION

In the first part of the paper we give a common generalization of two fundamental theorems of information theory: the noiseless coding theorem and the coding theorem for noisy channels.

In order to state the result we need some definitions.

Let $\mathfrak{X}$ and $\mathfrak{Y}$ be finite sets, which serve as input resp. output alphabet of a discrete memoryless channel (DMC) with transmission matrix $(W(y|x))_{x\in\mathfrak{X},\, y\in\mathfrak{Y}}$. $\mathfrak{X}^n = \prod_1^n \mathfrak{X}$ and $\mathfrak{Y}^n = \prod_1^n \mathfrak{Y}$ are the input resp. output words of length n. $(W(v^n|u^n))_{u^n\in\mathfrak{X}^n,\, v\in\mathfrak{Y}^n}$ is the n-th extension of the channel. $\mathfrak{X}^0$ and $\mathfrak{Y}^0$ are sets containing only the "empty word" and $\bar{\mathfrak{X}} = \bigcup_{n=0}^{\infty} \mathfrak{X}^n$, $\bar{\mathfrak{Y}} = \bigcup_{n=0}^{\infty} \mathfrak{Y}^n$.

*Research of this author was supported by the Deutsche Forschungsgemeinschaft.

We introduce now the notion of a praefix code for the DMC. $l^N = (l_1, \ldots, l_N)$ denotes a vector with natural numbers as components. A (l^N, λ)-PC (praefix code) for the DMC is a system of pairs $\{(u_i, D_i): 1 \leqslant i \leqslant N\}$ with

(a) $u_i \in \mathcal{X}^{l_i}$, $D_i \subset \mathcal{Y}^{l_i}$ for $i = 1, \ldots, N$,

(b) no element in $D = \bigcup_{i=1}^{N} D_i$ is praefix of another element in D,

(c) $\max_i W(D_i^c \mid u_i) \leqslant \lambda$.

In Theorem 1 of Section 2 we generalize Kraft's inequality to (l^N, λ)-PC's.

Assume now that there is given a probability distribution (p.d.) $\mathsf{P} = (P_1, \ldots, P_N)$ on the set of messages $\mathcal{M} = \{1, \ldots, N\}$. Then the *average length* of the (l^N, λ)-PC equals $L = \sum_{i=1}^{N} P_i l_i$. Denote by $L(\mathsf{P}, \lambda)$ the smallest average length of PC's with N code words and error probability λ. Our Theorem 2 in Section 3 states that

$$L(\mathsf{P}, \lambda) = \frac{H(\mathsf{P})}{C} + O(\sqrt{H(\mathsf{P})}),$$

where H is the entropy function and C is the capacity of the DMC.

Praefix codes are special cases of sequential coding schemes (SCS) for DMC's with complete feedback. If $L_f(\mathsf{P}, \lambda)$ denotes the smallest average length of SCS's for $(\mathsf{P}, \mathcal{M})$ with maximal error probability λ, then for any $\delta > 0$ there exists a λ, $0 < \lambda < 1$, such that

$$L_f(\mathsf{P}, \lambda) \geqslant (1 - \delta)\ \frac{H(\mathsf{P})}{C} + O(\sqrt{H(\mathsf{P})})\ .$$

This weak converse type estimate is the content of Theorem 3 in Section 4.

We give now a brief sketch of the reasoning which led to these results.

Originally we were interested in the following *search problem:* Suppose we are given a set $Z = \{z_1, \ldots, z_N\}$ of objects and exactly one element

z_i of Z is defective (wrong, unknown) with probability P_i $\left(\sum_i P_i = 1\right)$. We want to identify the unknown element z_i by testing some subsets A of Z whether A contains z_i or not. The result of a test is right and false with probabilities β and $1-\beta$, respectively, and these cases are independent of the other tests. Our goal is to minimize the *average* number of tests needed to identify the unknown element with a probability larger than a prescribed $1-\lambda$ $(0<\lambda<1)$.

In case $P_i = \frac{1}{N}$ for $i = 1, \ldots, N$ Rényi [1] considered the quantity $L_{\max}(\mathsf{P}, \lambda)$ = the minimal number of tests *maximally* needed in order to identify the unknown element with probability larger than $1-\lambda$. He showed that one needs about $\frac{\log N}{C}$ tests, where C is the capacity of the binary symmetric channel (BSC) with crossover probability $1-\beta$.

For general a priori distributions $\mathsf{P} = (P_1, \ldots, P_N)$ one gets a meaningful new problem only if one tries to minimize the *expected* number of tests. It was noticed by Sobel ([2]) and also by others (see [3]) that in the noiseless case $(\beta = 1)$ this problem is equivalent to encoding a source $\mathsf{P} = (P_1, \ldots, P_N)$ by a praefix code of minimal average length. The noiseless coding theorem says that this length is about $\frac{H(P)}{\log 2}$. (The same is true for the more general uniquely decipherable codes.)

This and Rényi's result *suggest the conjecture* that in our more general search problem we should manage with about $\frac{H(\mathsf{P})}{C}$ tests in average.

Rényi's proof of his result is an existence proof based on a standard random argument and gives no hint for a proof of the conjecture. However, his problem can be formulated in a language more familiar to channel coding theoretists: Given the number N of messages how large a block length n is needed to transmit those messages with small maximal error probability over a BSC in case of complete feedback? In this formulation the problem was solved already by Shannon ([4]) even for general DMC's: $n \sim \frac{\log N}{C}$.

Moreover, the construction of [5] gives an explicit asymptotically optimal fixed blocklength coding scheme and therefore also an asymptotically optimal search strategy for Rényi's problem.

Our search problem for general a priori distribution P is simply a coding problem for BSC's with complete feedback, if we permit sequential coding strategies (SCS) and measure the *code length* by

$$L = \sum_{i=1}^{N} P_i \mathsf{E}L_i,$$

where $\mathsf{E}L_i$ is the expected number of symbols needed to encode the i-th message. Passing from the BSC to general DMC's one might conjecture that

$$L \approx \frac{H(P)}{C}.$$

Since on the other hand feedback does not increase the capacity of a DMC one is led to conjecture this relationship also for praefix codes.

From a practical point of view it seems more appropriate to use the *average error concept*

$$\bar{\lambda} = 1 - \sum_{i=1}^{N} P_i W(D_i | u_i).$$

This changes the problem essentially. Results for this case are included in the forthcoming [6].

In the second part of the paper we determine the capacity region for multiple-access channels without synchronization (UMC) and thus give a generalization of the results of [7], [8], and [9]. In [7] Dobrushin introduced one-way channels without synchronization at the receiver and he gave a formula for the capacity in terms of a limiting expression. In [8] a computable formula for the capacity was obtained. The approach taken there is linked to the maximal error concept and is not adaptable to UMC's in case of average error, the only error concept for which the capacity region of the MC is known (see [9], [10]). Here we proceed by a rather simple reduction to the synchronized case via list decoding. Our proof is even in case of one-way channels much simpler than the previous one. It also

applies to degraded broadcast channels without synchronization. The reader familiar with those channels readily can see how our present approach can be applied. The strong converse for unsynchronized one-way channels can also be proved this way with the help of Margulis' theorem ([11], see also [12]). This argument seems not to apply if average errors are genuinely used.

2. A GENERALIZATION OF KRAFT'S INEQUALITY

Theorem 1.

(1) *Let* $\{(u_i, D_i)\colon 1 \leqslant i \leqslant N\}$ *be an* (l^N, λ)-PC *for the* DMC w, *then for all* λ $(0 < \lambda < 1)$ *and* γ $(0 < \gamma < 1)$

$$\sum_{i=1}^{N} e^{-Cl_i - \sqrt{\frac{d l_i}{(1-\lambda)\gamma}} + \log\{(1-\lambda)(1-\gamma)\}} \leqslant 1, \tag{2.1}$$

where d *depends only on* w.

(2) *One can give explicitly a function* $K(\lambda)$ *(see proof) such that the following is true: if*

$$\sum_{i=1}^{N} e^{-Cl_i + K(\lambda)\sqrt{l_i}} \leqslant 1 \tag{2.2}$$

holds for $l^N = (l_1, \ldots, l_N)$ *then there exists an* (l^N, λ)-PC.

Proof.

(1) Let $\bar{q}$ be the unique output distribution on $\mathcal{Y}$ which maximizes

$$\sum_{x,y} p(x)\, w(y \mid x) \log \frac{w(y \mid x)}{q(y)} \qquad (q = pw),$$

and let $\bar{q}^n = \bar{q} \times \ldots \times \bar{q}$. It is well-known (see [13]) that

$$\sum_{y} w(y \mid x) \log \frac{w(y \mid x)}{\bar{q}(y)} \leqslant C \qquad (x \in \mathcal{X}) \tag{2.3}$$

and that for all $u \in \mathcal{X}^n$

$$(2.4) \qquad \mathsf{E}_{W(\cdot|u)} \log \frac{W(\cdot|u)}{\bar{q}^n(\cdot)} \leqslant Cn$$

$$(2.5) \qquad \mathrm{Var}_{W(\cdot|u)} \log \frac{W(\cdot|u)}{\bar{q}^n(\cdot)} \leqslant nd,$$

where

$$d = \max_{x \in \mathfrak{X}} \mathrm{Var}_{w(\cdot|x)} \log \frac{w(\cdot|x)}{\bar{q}(\cdot)}$$

and $d = 0$, if the channel is noiseless.

We show now by a standard argument (originally due to Kemperman [14]) that for $u \in \mathfrak{X}^n$, $D \subset \mathfrak{Y}^n$ with $W(D|u) \geqslant 1 - \lambda$

$$(2.6) \qquad \bar{q}^n(D) \geqslant (1 - \lambda)(1 - \gamma) e^{-Cn - \sqrt{\frac{nd}{(1-\lambda)\gamma}}}.$$

To see this, set

$$B = \left\{ y^n \in \mathfrak{Y}^n : \log \frac{W(y^n|u)}{\bar{q}^n(y^n)} \geqslant Cn + \sqrt{\frac{nd}{(1-\lambda)\gamma}} \right\}$$

and derive from here

$$e^{Cn + \sqrt{\frac{nd}{(1-\lambda)\gamma}}} \cdot \bar{q}^n(D) \geqslant W(D \cap B^c | u) =$$

$$= W(D|u) - W(B|u) \geqslant 1 - \lambda - (1 - \lambda)\gamma,$$

where we have used in the last step Chebyshev's inequality and (2.4) and (2.5).

Set now $l^* = \max_{i=1,\ldots,N} l_i$ and define $D_i^* = D_i \times \mathfrak{Y}^{l^* - l_i}$ for $i = 1, \ldots, N$. Those sets are again disjoint and $\bar{q}^{l^*}(D_i^*) = \bar{q}^{l_i}(D_i)$. Therefore,

$$1 \geqslant \sum_{i=1}^{N} \bar{q}^{l^*}(D_i^*) = \sum_{i=1}^{N} \bar{q}^{l_i}(D_i) \geqslant$$

$$\geqslant (1-\lambda)(1-\gamma)\sum_{i=1}^{N} e^{-Cl_i-\sqrt{\frac{d}{(1-\lambda)\gamma}}l_i},$$

which was to be proved.

In the noiseless case $d=0$ and $\lambda=0$. By choosing γ arbitrarily small we get the classical Kraft-inequality.

(2) Our proof is based on a generalization of Feinstein's maximal coding method [15].

Let p be a p.d. on $\mathfrak{X}$ and $q=pw$ the corresponding p.d. on $\mathfrak{Y}$. We make use of the notions and simple properties of typical sequences and of generated sequences, both defined within $c(\lambda)\sqrt{n}$ deviation (see [13], Ch. 3). Denote by $\mathfrak{X}^n(p)$ the set of typical n-sequences and by $\mathfrak{Y}^n(p,w,u)$ the elements of $\mathfrak{Y}^n$ generated by $u\in\mathfrak{X}^n(p)$.

Then for some known functions $c_1(\lambda), c_2(\lambda), c_3(\lambda)$ and $c_4(\lambda)$:

$$|\mathfrak{Y}^n(p,w,u)| \leqslant e^{\sum_x p(x)H(w(\cdot|x))n+c_1(\lambda)\sqrt{n}} \tag{2.7}$$

$$q^n(v) \leqslant e^{-H(q)n+c_2(\lambda)\sqrt{n}} \quad \text{for} \quad v\in\mathfrak{Y}^n(p,w,u) \tag{2.8}$$

and

$$W(\mathfrak{Y}^n(p,w,u)|u) \geqslant 1-c_3(\lambda) \tag{2.9}$$

where $1-c_3(\lambda) > 1-\frac{\lambda}{2}$ if $c(\lambda)$ is large enough,

$$p^n(\mathfrak{X}^n(p)) \geqslant 1-c_4(\lambda) > 0. \tag{2.10}$$

Without loss of generality we can assume that $l_1 \leqslant l_2 \leqslant \ldots \leqslant l_N$. For a t $(1\leqslant t\leqslant N)$ and $l^t=(l_1,\ldots,l_t)$ let now $\{(u_i,D_i)\colon 1\leqslant i\leqslant t\}$ be a (l^t,λ)-PC with the following properties:

(a) $u_i\in\mathfrak{X}^{l_i}(p)$

(b) $D_i=\mathfrak{Y}^{l_i}(p,w,u_i)-\bigcup_{k=1}^{i-1} D_k\times\mathfrak{Y}^{l_i-l_k}$

(c) $W(D_i \mid u_i) \geqslant 1 - \lambda$

(d) The code is maximal in the sense that it is impossible to add another (u_{t+1}, D_{t+1}) without violating (a), (b) or (c).

It follows from (2.9) that such a code exists.

If $t < N$, then for every $u \in \mathfrak{X}^{l_{t+1}}(p)$ either $u \notin \{u_1, \ldots, u_t\}$ and for $D = \bigcup_{i=1}^{t} D_i \times \mathfrak{X}^{l_{t+1}-l_i}$

$$W(\mathfrak{Y}^{l_{t+1}}(p, w, u) \cap D \mid) > \frac{\lambda}{2}, \tag{2.11}$$

because otherwise the code could be prolonged or $u \in \{u_1, \ldots, u_t\}$ and then

$$\mathfrak{Y}^{l_{t+1}}(p, w, u) \subset D \tag{2.12}$$

by construction.

It follows from (2.9) that (2.11) holds also in this case. (2.10) and (2.11) imply that

$$\begin{aligned} q^{l_{t+1}}(D) &= \sum_{u \in \mathfrak{X}^{l_{t+1}}} p^{l_{t+1}}(u) W(D \mid u) \geqslant \\ &\geqslant \frac{\lambda}{2}(1 - c_4(\lambda)) = c_5(\lambda). \end{aligned} \tag{2.13}$$

On the other hand it follows from (2.7) and (2.8) that

$$\begin{aligned} q^{l_{t+1}}(D) &= \sum_{i=1}^{t} q^{l_{t+1}}(D_i \times \mathfrak{Y}^{l_{t+1}-l_i}) \leqslant \\ &\leqslant \sum_{i=1}^{t} e^{\sum_x p(x) H(w(\cdot \mid x)) l_i + c_1(\lambda)\sqrt{l_i} - H(q) l_i + c_2(\lambda)\sqrt{l_i}} \end{aligned} \tag{2.14}$$

By choosing for p a maximizing input distribution $(pw = \bar{q})$, we obtain

$$q^{l_{t+1}}(D) \leqslant \sum_{i=1}^{t} e^{-Cl_i + c_6(\lambda)\sqrt{l_i}} \tag{2.15}$$

where $c_6(\lambda) = c_1(\lambda) + c_2(\lambda)$.

This and (2.13) imply

$$(2.16) \qquad 0 < c_5(\lambda) \leqslant \sum_{i=1}^{t} e^{-Cl_i + c_6(\lambda)\sqrt{l_i}}$$

or

$$1 \leqslant \sum_{i=1}^{t} e^{-Cl_i + (c_6(\lambda) - \log c_5(\lambda))\sqrt{l_i}}.$$

This is a contradiction to our assumption $t < N$, if

$$K(\lambda) = c_6(\lambda) - \log c_5(\lambda),$$

because the terms in the sum of (2.2) are positive.

Q.E.D.

3. THE CODING THEOREM FOR PRAEFIX CODES

Assume now that a set $\mathfrak{M} = \{1, \ldots, N\}$ of messages is given with probability P_i that i is to be sent over the channel. We are interested in the quantity

$$(3.1) \qquad L = L(\mathsf{P}, \lambda) = \min_{(l^N, \lambda)\text{-PC's}} \sum_{i=1}^{N} P_i l_i.$$

Theorem 2 (coding theorem and strong converse). *For any fixed* λ $(0 < \lambda < 1)$

$$L(\mathsf{P}, \lambda) = \frac{H(\mathsf{P})}{C} + O(\sqrt{H(\mathsf{P})}).$$

Proof. From Theorem 1 we get for

$$T = \max\left(K(\lambda), \sqrt{\frac{d}{(1-\lambda)\gamma}} - \log(1-\lambda)(1-\gamma)\right)$$

the estimates

$$(3.2) \qquad \min\left\{\sum_{i=1}^{N} P_i l_i : \sum_{i=1}^{N} e^{-Cl_i - T\sqrt{l_i}} \leqslant 1\right\} \leqslant$$

$$\leqslant L \leqslant \min\Big\{\sum_{i=1}^{N} P_i l_i \colon \sum_{i=1}^{N} e^{-Cl_i + T\sqrt{l_i}} \leqslant 1\Big\}.$$

First we derive an upper bound on L.

Choose minimal integers l_i with

$$P_i \geqslant e^{-Cl_i + T\sqrt{l_i}}; \qquad (i = 1, \ldots, N). \tag{3.3}$$

Then clearly,

$$1 = \sum_i P_i \geqslant \sum_i e^{-Cl_i + T\sqrt{l_i}}.$$

The larger of the two solutions of the equation

$$\log P_i = -Cx + T\sqrt{x}$$

is

$$\begin{aligned} x &= -\frac{\log P_i}{C} + \frac{T^2}{2C^2} + \sqrt{\Big(\frac{T^2}{2C^2}\Big)^2 - \frac{T^2}{C^3}\log P_i} \leqslant \\ &\leqslant -\frac{\log P_i}{C} + \frac{T^2}{C^2} + \frac{T}{C^{\frac{3}{2}}}\sqrt{-\log P_i}. \end{aligned} \tag{3.4}$$

Therefore,

$$l_i \leqslant x + 1 \leqslant -\frac{\log P_i}{C} + \frac{T}{C^{\frac{3}{2}}}\sqrt{-\log P_i} + \Big(1 + \frac{T^2}{C^2}\Big)$$

and hence

$$L \leqslant \sum_i P_i l_i \leqslant \frac{H(P)}{C} + \frac{T}{C^{\frac{3}{2}}}\sum_i P_i\sqrt{-\log P_i} + \Big(1 + \frac{T^2}{C^2}\Big).$$

Since $\sqrt{\ }$ is a concave function we finally get

$$L \leqslant \frac{H(\mathsf{P})}{C} + O(\sqrt{H(\mathsf{P})}). \tag{3.5}$$

Let now $l^N = (l_1, \ldots, l_N)$ be such that

$$\sum_i e^{-Cl_i - T\sqrt{l_i}} \leqslant 1 \tag{3.6}$$

and

$$(3.7) \qquad L \geqslant \sum_{i=1}^{N} P_i l_i.$$

We can write

$$\sum_i P_i l_i = \frac{1}{C} \sum_i - P_i \log e^{-Cl_i - T\sqrt{l_i}} - \frac{T}{C} \sum_i P_i \sqrt{l_i}$$

and hence again by the concavity of $\sqrt{\ }$

$$(3.8) \qquad L + \frac{T}{C}\sqrt{L} \geqslant \frac{1}{C} \sum_i - P_i \log e^{-Cl_i - T\sqrt{l_i}}.$$

Define

$$Q_i = \frac{e^{-Cl_i - T\sqrt{l_i}}}{\sum_i e^{-Cl_i - T\sqrt{l_i}}}$$

and use the inequality

$$- \sum_i P_i \log P_i \leqslant - \sum_i P_i \log Q_i$$

in order to derive from (3.8)

$$(3.9) \qquad L + \frac{T}{C}\sqrt{L} \geqslant \frac{H(\mathsf{P})}{C} - \frac{1}{C} \sum_i P_i \log \sum_i e^{-Cl_i - T\sqrt{l_i}} \geqslant \frac{H(\mathsf{P})}{C},$$

because of (3.6). An easy calculation yields finally

$$L \geqslant \frac{H(\mathsf{P})}{C} - O(\sqrt{H(\mathsf{P})}).$$

Q.E.D.

4. A LOWER BOUND ON THE MINIMAL AVERAGE LENGTH $L_f(\mathsf{P}, \lambda)$

In case of complete feedback the sender has the possibility to encode a message by an encoding (vector-valued) function, that is, he chooses the symbols to be send depending on the letters received in so far:

$$f = [f_1, f_2(Y_1), \ldots, f_k(Y_1, \ldots, Y_{k-1}), \ldots]$$

where $f_1 \in \mathfrak{X}$, $f_k \colon \mathfrak{Y}^{k-1} \to \mathfrak{X}$ for $k = 2, 3, \ldots$.

The distribution of Y_k depends on $Y_1, \ldots, Y_{k-1}$ and w.

An $(\bar{l}^N, \lambda)$-SCS, is a system $\{(f_i, D_i)\colon 1 \leqslant i \leqslant N\}$, where for $i = 1, \ldots, N$

(a) $D_i \subset \bar{\mathfrak{Y}}$,

(b) no sequence in $D = \bigcup_{i=1}^{N} D_i$ is praefix of another sequence in D, in particular, the D_i's are disjoint,

(c) $f_i = [f_{1i}, f_{2i}(Y_1), \ldots, f_{t(Y_1, \ldots, Y_{t-1})i}(Y_1, \ldots, Y_{t-1})]$ where t is the largest integer such that $Y_1, \ldots, Y_{t-1}$ is a *non-trivial* praefix of an element in D_i,

(d) $W(D_i | f_i) \geqslant 1 - \lambda$,

(e) $\bar{l}_i = \sum_{l=1}^{\infty} l W(\mathfrak{Y}^l | f_i)$.

In order to derive a lower bound on $L_f(\mathsf{P}, \lambda)$ we change a given $(\bar{l}^N, \lambda)$-SCS in such a way that within every new decoding set all sequences have the same length – a property which PC's have by definition. At the same time we keep control over the error probabilities.

Let $l_i = l_i(\epsilon) = [\bar{l}_i(1 + \epsilon)]$, that is, the smallest integer larger than $\bar{l}_i(1 + \epsilon)$ and let

$$B_i(\epsilon) = \bigcup_{l=1}^{l_i} \mathfrak{Y}^l.$$

Then

$$W(B_i(\epsilon)^c | f_i) l_i \leqslant \bar{l}_i \tag{4.1}$$

and

$$W(B_i(\epsilon)^c | f_i) \leqslant \frac{1}{1 + c}. \tag{4.2}$$

For ϵ such that $\lambda(\epsilon) = \lambda + \frac{1}{1+\epsilon} < 1$ we have then

(4.3) $\quad W(D_i \cap B_i(\epsilon) | f_i) \geqslant 1 - \lambda(\epsilon) > 0.$

Define now

(4.4) $\quad D_i^* = \{y^{l_i}: \exists$ praefix of y^{l_i} in $D_i \cap B_i(\epsilon)\}$

and for $y^{l_i} \in D_i^*$

(4.5) $\quad f_i^*(y^{l_i}) = [f_{i1}, f_{i2}(y_1), \ldots, f_{is}(y_1, \ldots, y_{s-1}), f_{is+1}^*, \ldots, f_{il_i}^*]$

where $f_{is+1}^*, \ldots, f_{il_i}$ are arbitrary elements of $\mathfrak{X}$ and $(y_1, \ldots, y_s)$ is the praefix of y^{l_i} contained in $D_i \cap B_i(\epsilon)$. Again we have

(4.6) $\quad W(D_i^* | f_i^*) \geqslant 1 - \lambda(\epsilon).$

Now we are almost in the situation we discussed in Section 2, the only difference is that instead of code words we have constant length encoding functions. In Kemperman's ([14]) strong converse proof for fixed block length feedback schemes the analogue of (2.6) was established:

(4.7) $\quad \bar{q}^{l_i}(D) \geqslant e^{-Cl_i - g_1(\lambda, \epsilon)\sqrt{l_i}}.$

From here we can derive by the same arguments as the ones used in Section 2:

(4.8) $\quad \sum_{i=1}^{N} e^{-Cl_i - g_2(\lambda, \epsilon)\sqrt{l_i}} \leqslant 1.$

As in Section 3 we conclude that

(4.9) $\quad \sum_i' P_i l_i \geqslant \frac{H(\mathsf{P})}{C} - O(\sqrt{H(\mathsf{P})}).$

Since $\bar{l}_i \geqslant \frac{1}{1+\epsilon} l_i - 1$ and since by choosing λ arbitrarily small we can choose ϵ arbitrarily small, (4.9) implies

Theorem 3 (weak converse). *For given* $\delta > 0$ *there exists a* $\lambda(\delta)$ *such that*

$$L_f(\mathsf{P}, \lambda) \geqslant (1 - \delta)\left(\frac{H(\mathsf{P})}{C} - O(\sqrt{H(\mathsf{P})})\right)$$

(or such that $L_f(\mathsf{P}, \lambda) \geqslant (1 - \delta)\frac{H(\mathsf{P})}{C}$ *for* $H(\mathsf{P})$ *large).*

5. THE MULTIPLE-ACCESS CHANNEL WITHOUT SYNCHRONIZATION

Let $\mathfrak{X}, \mathfrak{Y}$ and $\mathfrak{Z}$ be finite sets, $\mathfrak{X}^n = \prod_1^n \mathfrak{X}$, $\mathfrak{Y}^n = \prod_1^n \mathfrak{Y}$ and $\mathfrak{Z}^n = \prod_1^n \mathfrak{Z}$. Furthermore, let

$$\bar{\mathfrak{Z}} = \bigcup_{l=0}^{\infty} \mathfrak{Z}^l, \tag{5.1}$$

where $\mathfrak{Z}^0$ constains only the "empty word".

$\bar{\mathfrak{Z}}$ is the set of all finite output sequences. If $x \in \mathfrak{X}$ and $y \in \mathfrak{Y}$ are sent by the two senders $w(\bar{z} \mid x, y)$ is the probability that the sequence $\bar{z} \in \bar{\mathfrak{Z}}$ is received by the receiver. Clearly, $\sum_{\bar{z} \in \bar{\mathfrak{Z}}} w(\bar{z} \mid x, y) = 1$ for every $x \in \mathfrak{X}$, $y \in \mathfrak{Y}$.

Notice that we permit the possibility of an erasure, that is, the receiver receives the "empty word".

We assume a memoryless character of the channel:

$$W(\bar{z}^n \mid x^n, y^n) = \prod_{t=1}^n w(\bar{z}_t \mid x_t, y_t)$$

for every $x^n = (x_1, \ldots, x_n) \in \mathfrak{X}^n$, $y^n = (y_1, \ldots, y_n) \in \mathfrak{Y}^n$ and $\bar{z}^n = (\bar{z}_1, \ldots, \bar{z}_n) \in \bar{\mathfrak{Z}}^n$.

We denote by $\bar{z}_1 \ldots \bar{z}_n$ an element of $\bar{\mathfrak{Z}}$ obtained by writing the terms of the sequence $\bar{z}_1, \ldots, \bar{z}_n$ consecutively in their natural coder. With this convention we define the transmission probabilities of an unsynchronized multiple-access channel (UMC) by

$$P(\bar{z} \mid x^n, y^n) = \sum_{\bar{z}_1 \ldots \bar{z}_n = \bar{z}} W(\bar{z}^n \mid x^n, y^n) \tag{5.2}$$

for every $x^n \in \mathfrak{X}^n$, $y^n \in \mathfrak{Y}^n$ and $\bar{z} \in \bar{\mathfrak{Z}}$ $(n = 1, 2, \ldots)$.

A (n, M, N, λ) *code* for the UMC is a system $\{(u_i, v_j, D_{ij})$: $1 \leqslant i \leqslant M$; $1 \leqslant j \leqslant N\}$ where $u_i \in \mathfrak{X}^n$, $v_j \in \mathfrak{Y}^n$, $D_{ij} \subset \bar{\mathfrak{Z}}$, $D_{ij} \cap D_{i'j'} = \phi$ for $(i, j) \neq \neq (i', j')$ and

$$(5.3) \qquad \frac{1}{MN} \sum_{i,j} P(D_{ij}^c \mid u_i, v_j) \leqslant \lambda.$$

A pair of non-negative real numbers (R_1, R_2) is called a pair of achievable rates, if for any λ $(0 < \lambda < 1)$, and any $\epsilon > 0$ there exists an (n, M, N, λ)-code with $\frac{1}{n} \log M \geqslant R_1 - \epsilon$ and $\frac{1}{n} \log N \geqslant R_2 - \epsilon$ for all sufficiently large n. The *capacity region* $\mathfrak{U}$ is simply the set of all pairs of achievable rates.

Remark. If we would exchange the phrase "for all sufficiently large n" by "for an infinite sequence of n's" we would get a capacity region $\mathfrak{U}^*$ which a priori might be larger. If this is not the case one says that a channel has a capacity. Actually the known results in multi-user communication are all such that $\mathfrak{U} = \mathfrak{U}^*$.

Even though our results hold in greater generality – in order to keep the arguments simple – we make the

(5.4) *Supposition*

If $w(\bar{z} \mid x, y) > 0$ for some $(x, y) \in \mathfrak{X} \times \mathfrak{Y}$ then $0 \leqslant l(\bar{z}) \leqslant B$, where B is a constant and $l(\bar{z})$ equals the number of components of $\bar{z}$, i.e., the length of the "letter" $\bar{z}$.

We shall treat the UMC by comparing it with an MC, which we now define. Let I be a positive integer and let

$$(\mathfrak{X}^I)_t = \prod_{s=I(t-1)+1}^{It} \mathfrak{X}_s \qquad (t = 1, 2, \ldots)$$

$$(\mathfrak{X}^I)^m = \prod_{t=1}^{m} (\mathfrak{X}^I)_t.$$

Define $(\mathfrak{Y}^I)_t$ and $(\mathfrak{Y}^I)^m$ analogously. We write for convenience $\tilde{\mathfrak{X}}$ instead of $\mathfrak{X}^I$, resp. $\tilde{\mathfrak{Y}}$ instead of $\mathfrak{Y}^I$, and we define $\tilde{\mathfrak{Z}}$ as

$$\tilde{\mathfrak{Z}} = \{\bar{z}\colon\ \bar{z} = \bar{z}_1 \dots \bar{z}_I,\ 0 \leqslant l(\bar{z}_s) \leqslant B \ \text{ for } \ s = 1, \dots, I\}.$$

The transmission matrix of the synchronized channel J^I is defined by

$$W(\bar{z} \mid \tilde{x}, \tilde{y}) = \sum_{\bar{z}_1 \dots \bar{z}_I = \bar{z}} W(\bar{z}_1, \dots, \bar{z}_I \mid \tilde{x}, \tilde{y}) \tag{5.5}$$

For $m = 1, 2, \dots$ and every $\tilde{x}^m = (\tilde{x}_1, \dots, \tilde{x}_m) \in (\mathfrak{X}^I)^m$, every $\tilde{y}^m = (\tilde{y}_1, \dots, \tilde{y}_m) \in (\mathfrak{Y}^I)^m$ and every $\bar{z}^m = (\bar{z}_1, \dots, \bar{z}_m) \in \bar{\mathfrak{Z}}^m$ the transmission probabilities of J^I are defined by

$$Q(\bar{z}^m \mid \tilde{x}^m, \tilde{y}^m) = \prod_{t=1}^{m} \tilde{w}(\bar{z}_t \mid \bar{x}_t, \bar{y}_t). \tag{5.6}$$

J^I is a (memoryless) MC with input alphabets $\tilde{\mathfrak{X}}$, $\tilde{\mathfrak{Y}}$ of sizes $|\mathfrak{X}|^I, |\mathfrak{Y}|^I$ and output alphabet $\tilde{\mathfrak{Z}} \subset \bar{\mathfrak{Z}}$ of size $A < |\mathfrak{Z}|^{(B+1)I}$.

6. AN AUXILIARY RESULT

We shall need in the next section a result which allows us to reduce list codes of small list size for the MC to codes of list size 1 without loosing too much in rate or error probability. For one-way channels the argument was used in [16]. The present generalization is straightforward, but we include the argument, because it is so brief. We give a general formulation for channels without time structure.

Lemma (list reduction). *Let* $\{(u_i, v_j, D_{ij})\colon\ 1 \leqslant i \leqslant M;\ 1 \leqslant j \leqslant N\}$ *be an* $(1, M, N, \lambda, L)$ *list code for the* MC w *with alphabets* $\mathfrak{X}, \mathfrak{Y}$ *and* $\mathfrak{Z}$. *Then for* $M^* \leqslant M$, $N^* \leqslant N$ *there exists a* $(1, M^*, N^*, \lambda^*, 1)$*-subcode* $\{(u_{i_s}, v_{j_t}, D_{i_s j_t})\colon\ 1 \leqslant s \leqslant M^*;\ 1 \leqslant t \leqslant N^*\}$ *with error probability*

$$\lambda^* \leqslant \lambda + \frac{3}{2}\left(M^* \frac{L}{M} + N^* \frac{L}{N}\right).$$

Proof. Let S_i $(i = 1, \dots, M^*)$ and T_j $(j = 1, \dots, N^*)$ be independent random variables with distributions

$$\mathsf{P}(S_i = k) = \frac{1}{M}; \quad \mathsf{P}(T_j = l) = \frac{1}{N}$$

(6.1)

$$(k = 1, 2, \ldots, M;\ l = 1, 2, \ldots, N).$$

With every outcome

$$\{(s_i, t_j):\ 1 \leqslant i \leqslant M^*;\ 1 \leqslant j \leqslant N^*\}$$

of $\{(S_i, T_j):\ 1 \leqslant i \leqslant M^*;\ 1 \leqslant j \leqslant N^*\}$ we associate codewords

$$\{(u_{s_i}, v_{t_j}):\ 1 \leqslant i \leqslant M^*;\ 1 \leqslant j \leqslant N^*\}$$

and decoding sets

$$\{F_{s_i t_j}:\ 1 \leqslant i \leqslant M^*;\ 1 \leqslant j \leqslant N^*\}$$

where

$$F_{s_i t_j} = D_{s_i t_j} - \bigcup_{(i',j') \neq (i,j)} D_{s_{i'} t_{j'}}. \tag{6.2}$$

For reasons of symmetry the expected average error probability of the ensemble of codes

$$\mathsf{E}\lambda(S_1, \ldots, S_{M^*};\ T_1, \ldots, T_{N^*}) =$$

$$= M^{*-1} N^{*-1} \sum_{i=1}^{M^*} \sum_{j=1}^{N^*} \mathsf{E} \sum_{z \in F^c_{S_i T_j}} w(z \mid u_{S_i}, v_{T_j})$$

equals

$$\mathsf{E} \sum_{z \in F^c_{S_1 T_1}} w(z \mid u_{S_1}, v_{T_1})$$

and this expression is upper bounded by

$$\mathsf{E} \sum_{z \in D^c_{S_1 T_1}} w(z \mid u_{S_1}, v_{T_1}) + \mathsf{E} \sum_{z \in D_{S_1 T_1} \cap F^c_{S_1 T_1}} w(z \mid u_{S_1}, v_{T_1}).$$

The first term is $\leqslant \lambda$.

Assume now that $(S_1, T_1) = (s_1, t_1)$ and that $z \in D_{s_1 t_1}$. The probability that z is element of $F^c_{s_1 t_1}$, that is, of $\bigcup_{(i,j) \neq (1,1)} D_{S_i T_j}$ is smaller than

$$\frac{L}{MN}(M^*-1)(N^*-1)+\frac{L}{M}(M^*-1)+\frac{L}{N}(N^*-1)\leqslant$$

$$\leqslant\frac{3}{2}\left(M^*\frac{L}{M}+N^*\frac{L}{N}\right).$$

The same bound holds for the whole second term.

Q.E.D.

7. THE CAPACITY REGION AS A LIMIT

Since single letter characterizations of the capacity region of a MC are known ([9], [10]), we know the capacity region $\mathfrak{C}^I$ of J^I, in particular we know that $\mathfrak{C}^I$ is a convex, compact set in E^2.

Define now

$$\mathfrak{C}=\bigcap_{I=1}^{\infty}\mathfrak{C}^I. \tag{7.1}$$

Clearly, $\mathfrak{C}$ is again convex and compact. For all n which are multiples of I: $n=s(n)I$ a code for $\mathfrak{U}$ induces canonically a code for J^I and therefore $\mathfrak{U}\subset\mathfrak{C}^I$ for all I. Hence.

$$\mathfrak{U}\subset\mathfrak{C}=\bigcap_{I=1}^{\infty}\mathfrak{C}^I. \tag{7.2}$$

Remark. In order to show that actually also $\mathfrak{U}^*\subset\mathfrak{C}$ holds one has to vorry about the cases where n is not multiple of I:

$$s(n)I\leqslant n\leqslant s(n)I+I',\ I'<I.$$

One could introduce a non-stationary MC, where every $(s(n)+1)$-st component is different from all the others. If n is large, the effect of this component is negligible as can be seen from [9].

The direct part: $\mathfrak{U}\supset\mathfrak{C}$.

Let now $I(n)$ be a function of n such that $I(n)=o(n)$ and $\lim_{n\to\infty}I(n)=\infty$. A more specific choice of $I(n)$ will be made below.

Let $s(n)$ be the unique number for which

(7.3) $$n' = s(n)I(n) \leqslant n < s(n)I(n) + I'(n), \qquad I'(n) < I(n).$$

Given an (n', M, N, λ)-code $\{(u_i^*, v_j^*, D_{ij}^*)\colon 1 \leqslant i \leqslant M;\ 1 \leqslant j \leqslant N\}$ for $J^{I(n)}$, we can assign to every D_{ij}^* a decoding set D_{ij} for $\mathfrak{U}$ by

$$D_{ij} = \{\bar{z} = \bar{z}_1 \ldots \bar{z}_{n'}\bar{z}_{n'+1} \ldots \bar{z}_n\colon$$
$$(\bar{z}_1 \ldots \bar{z}_{I(n)}, \bar{z}_{I(n)+1} \ldots \bar{z}_{2I(n)}, \ldots$$
$$\ldots, \bar{z}_{(s(n)-1)I(n)} \ldots \bar{z}_{n'}) \in D_{ij}^* \text{ and } \bar{z}_t \in \bar{\mathfrak{Z}},\ n'+1 \leqslant t \leqslant n\}$$

Also, we pass to a block length n code by defining for some $x \in \mathfrak{x}$, $y \in \mathfrak{y}$, $u_i = (u_{i1}, \ldots, u_{in})$, $v_j = (v_{j1}, \ldots, v_{jn})$ with

$$u_{it} = u_{it}^*$$

$$v_{jt} = v_{jt}^*$$

for $1 \leqslant t \leqslant n'$ and with

$$u_{it} = x$$

$$v_{jt} = y$$

for $n'+1 \leqslant t \leqslant n$.

$$\{(u_i, v_j, D_{ij})\colon 1 \leqslant i \leqslant M;\ 1 \leqslant j \leqslant N\}$$

is now a block length n *list code* for $\mathfrak{U}$ with error probability $\leqslant \lambda$ and *list size*

$$L(n) \leqslant \binom{n(B+1) + s(n)}{s(n)}.$$

If we choose now $I(n)$ such that

(7.4) $$L(n) \leqslant e^{n\delta_n}, \qquad \delta_n \to 0 \quad (n \to \infty)$$

then we can apply the *list reduction lemma* of Section 6 and get a list size 1 code for U of essentially the same rates. What are the achievable rates for $J^{I(n)}$? Notice that the variances of the information densities occurring

in the proof of the coding theorem for MC ([9], [10]) now depend on $I(n)$. However, they are bounded by

$$\log^2(|\mathfrak{X}|\,|\mathfrak{Y}|)^{I(n)}.$$

and we thus can achieve with error probability λ all rates in

$$\mathfrak{C}^{I(n)}[1 - n^{-1}c(\lambda)\sqrt{s(n)}\log^2(|\mathfrak{X}|\,|\mathfrak{Y}|)^{I(n)}]$$

where $c(\lambda)$ depends on λ only.

The term in brackets tends to 1 as $n \to \infty$ if

$$I(n)^2\sqrt{s(n)} = o(n). \tag{7.5}$$

Since $L(n) \leqslant \exp\{s(n)(\log n + \log(B+1)\}$, $I(n) = [n^{\frac{1}{4}}]$ satisfies both constraints: (7.4) and (7.5). This completes the proof of

Theorem 4.

$$\mathfrak{U} = \mathfrak{C}.$$

8. A "COMPUTABLE" FORMULA FOR THE CAPACITY REGION

Let $\mathfrak{C}^n$ and $\mathfrak{C}$ be defined as in Section 7. We know that $\mathfrak{C}^n \supset \mathfrak{C}$ and both sets are convex and compact in E. The quantities

$$d(\mathfrak{C}^n, \mathfrak{C}) = \max_{a\in\mathfrak{C}^n} \min_{b\in\mathfrak{C}} \rho(a, b), \tag{8.1}$$

where ρ is the euclidian distance, measures how well $\mathfrak{C}^n$ approximates $\mathfrak{C}$.

Theorem 5.

$$d(\mathfrak{C}^n, \mathfrak{C}) \leqslant T\frac{\log n}{n} \qquad (n = 2, 3, \ldots) \tag{8.2}$$

where $T \leqslant (1 + \log 2(B+2)) \sum_{l=0}^{\infty} \frac{1+l}{2^l}$ *(and* B *is defined in* (5.4)*).*

Remark. The estimate is – to within the constant T – the best possible as can be seen from the following *example:* let $\mathfrak{X} = \{0, 1\}$, $\mathfrak{Y} = \{0, 1\}$

and w such that $w(0|0,0) = w(0|0,1) = 1$, $w(00|1,0) = w(00|1,1) = 1$. Then $\mathfrak{C}^n = \{(a,0): a \leqslant n^{-1} \log(n+1)\}$ and $\mathfrak{C} = \{(0,0)\}$.

Proof. Now we need for the first time that $\mathfrak{C}^n$ is characterized in terms of information quantities: mutual information and conditional mutual information (see [9] or [10]). This is the reason why $\mathfrak{C}^n$ and by the Theorem also $\mathfrak{C}$ can be numerically determined within any predescribed $f(Z)$ clearly

$$|H(f(Z)) - H(Z)| = H(Z|f(Z)) \leqslant \log \max_r |\{z: f(z) = r\}|$$

and therefore also for three finite-valued random variables X, Y, Z:

(8.4) $$|I(X \wedge Z|Y) - I(X \wedge f(Z)|Y)|$$

and

$$|I(X \wedge Z) - I(X \wedge f(Z))|$$

do not exceed

$$\log \max_r |\{z: f(z) = r\}|.$$

By the data processing theorem

(8.5) $$2n\mathfrak{C}^n \supset 2n\mathfrak{C}^{2n}$$

and

(8.6) $$n\mathfrak{C}^n \supset m\mathfrak{C}^m \quad \text{for} \quad n \geqslant m.$$

These two inclusions imply that for any $\epsilon > 0$ $\exists n(\epsilon)$:

(8.7) $$d(\mathfrak{C}^n, \mathfrak{C}) \leqslant \epsilon \quad \text{for all} \quad n \geqslant n(\epsilon).$$

(8.4) implies that

(8.8) $$d(2n\mathfrak{C}^n, 2n\mathfrak{C}^{2n}) \leqslant \log \binom{2n(B+1)+1}{1} \leqslant \log n + \log 2(B+2)$$

and hence

(8.9) $$d(\mathfrak{C}^n, \mathfrak{C}^{2n}) \leqslant \frac{\log n}{2n} + \frac{\log 2(B+2)}{2n}.$$

For any *real* number $h \geqslant 1$ for which 2^h is an *even* integer we therefore have

$$d(\mathfrak{C}^{2^{h-1}}, \mathfrak{C}^{2^h}) \leqslant \frac{h}{2^h} T^*, \tag{8.10}$$

where $T^* = 1 + \log 2(B + 2)$. This and (8.7) imply that

$$d(\mathfrak{C}^{2^{h-1}}, \mathfrak{C}) \leqslant \sum_{l=0}^{\infty} d(\mathfrak{C}^{2^{h+l-1}}, \mathfrak{C}^{2^{h+l}}) \leqslant$$

$$\leqslant T^* \sum_{l=0}^{\infty} \frac{h+l}{2^{h+l}} = T^* \frac{h}{2^h} \sum_{l=0}^{\infty} \frac{1 + \frac{l}{h}}{2^l} \leqslant T \frac{h}{2^h} \tag{8.11}$$

where $T = T^* \sum_{l=0}^{\infty} \frac{1+l}{2^l}$. Any natural number n can be written as $n = 2^{h-1}$ $(h \geqslant 1)$. Therefore

$$d(\mathfrak{C}^n, \mathfrak{C}) \leqslant T \frac{\log 2n}{2n} \leqslant T \frac{\log n}{n} \quad \text{for} \quad n \geqslant 2.$$

Q.E.D.

REFERENCES

[1] A. Rényi, On a problem of information theory, *Publ. Math. Inst. Hungar. Acad. Sci.*, 6 (1961), 505-516.

[2] M. Sobel, Group testing to classify efficiently ali defectives in a binomial sample, *Information and Decision Processes,* (ed. R.E. Machol), McGraw-Hill, 1960, 127-161.

[3] G. Katona, Combinatorial search problems, Notes of a lecture held at the International Centre of Mechanical Sciences, Udine, Italy, 1972.

[4] C.E. Shannon, Zero-error capacity of noisy channels, *IEEE Trans. on Inf. Theory,* 2 (1956), 8-19.

[5] R. Ahlswede, A constructive proof of the coding theorem of discrete memoryless channels in case of complete feedback, *Sixth Prague Conf. on Inf. Th.*, (1971), 1-22.

[6] R. Ahlswede, *Information Theory,* (book in preparation).

[7] R.L. Dobrushin, Shannon's theorem for channels with synchronization errors, *Problemy Peredači Informacii,* 3 (1967), 18-36, *Problems of Inf. Transmission,* 3 (1967), 11-26.

[8] R. Ahlswede – J. Wolfowitz, Channels without synchronization, *Adv. Appl. Prob.*, 3 (1971), 383-403.

[9] R. Ahlswede, Multi-way communication channels, *Second International Symposium on Inf. Th.*, Tsahkadsor, Armenian SSR, 1971, Publishing House of the Hungarian Academy of Sciences, 1973, 23-52.

[10] R. Ahlswede, The capacity region of a channel with two senders and two receivers, *Annals of Prob. Theory,* 2 (1974), 805-814.

[11] G.A. Margulis, Probabilistic properties of graphs with large connectivity, *Probl. Peredači Informacii,* 10 (1974), 101-108.

[12] R. Ahlswede – P. Gács – J. Körner, Bounds on conditional probabilities with applications in multi-user communication, *Z. Wahrscheinlichkeitstheorie u. verw. Geb.*, 34 (1976), 151-177.

[13] J. Wolfowitz, *Coding Theorems of Information Theory,* Springer Verlag, Berlin – Heidelberg – New York, 1961 and 1964.

[14] J.H.B. Kemperman, *Studies in coding theory,* Mimeographed lecture notes, 1961.

[15] A. Feinstein, A new basic theorem of information theory, *IRE Trans. PGIT,* 1954, 2-22.

[16] R. Ahlswede – G. Dueck, Every bad code has a good subcode: a local converse to the coding theorem, *Z. Wahrscheinlichkeitstheorie u. verw. Geb.*, 34 (1976), 179-182.

R. Ahlswede

Fakultät für Mathematik der Universität Bielefeld Kurt-Schumacher-Str.6, 4800 Bielefeld 1.

P. Gács

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