J. Appl. Prob. 10, 824–836 (1973)
Printed in Israel
© *Applied Probability Trust* 1973

CHANNEL CAPACITIES FOR LIST CODES

RUDOLF AHLSWEDE, *Ohio State University, Columbus*

Abstract

In the present paper we demonstrate that the concept of a list code is from a mathematical point of view a more canonical notion than the classical code concept (list size one) in that it allows a unified treatment of various coding problems. In particular we determine for small list sizes the capacities of arbitrarily varying channels.

PROBABILISTIC CODING THEORY; ARBITRARILY VARYING CHANNELS; COMPOUND CHANNELS; CHANNEL CAPACITIES; ZERO ERROR CAPACITY; LIST CODES; FEEDBACK

1. Introduction

A. Preliminary remarks

List codes were first considered by Elias in [3]. They are a natural generalization of ordinary codes. Instead of making a single decision about which code word was sent, the decoder decides on a list of code words. Only if the transmitted code word is not on the list do we say that a decoding error has occurred. This decision scheme is of *practical* use whenever incorrect code words on the list can be recognized by internal evidence. This is, for example, sometimes possible in the case of the transmission of a text in an ordinary language, because languages have a certain degree of redundancy.

The concept of a list code has found an additional justification by providing a helpful tool for proving results for ordinary coding, (see, for instance, [8]).

Our present investigation originated from an analysis of a method to prove the coding theorem for certain channels with noiseless feedback ([4], [5]).

The method consists of three parts:

(1) a lemma on list codes (Lemma 1 in Section 2);

(2) a procedure to reduce the list codes iteratively to list codes of small list size;

(3) the reduction to list size 1, that is, to ordinary coding.

In [4] we developed the method for the discrete memoryless channel with noiseless feedback (d.m.c.f.) and in [5] we gave an extension to channels with arbitrarily varying channel probability functions with noiseless feedback (a.v.ch.f.). The method does not work for compound channels with noiseless feedback (c.ch.f.). This is somehow surprising, since the method seems to be very natural.

Received 27 November 1972.

Of course c.ch.f. can easily be treated otherwise, but there remains this formal inconsistency, which seems to indicate that something is not clearly understood. We therefore asked for the "well understood" method and for the coding problem for which this method is applicable without exceptions.

In the method described above, feedback is made use of only in Parts (2) and (3). In the present paper we provide a tool (Lemma 4 in Section 2), which makes it possible to make the reduction in (2) also without feedback. Thus we obtain a method to prove coding theorems for a large class of channels in the case of list decoding and this modified method is therefore just appropriate for list decoding. We shall treat compound channels, a.v.ch. and the zero error capacity problem. For ordinary coding the capacity of an a.v.ch. is known only for the case of a binary output alphabet [7] and a few other special cases. Our original method works in the case of feedback only for those channels for which the list code capacity (see (1.7)) equals the feedback capacity. For compound channels, for instance, this is in general not the case.

We limit ourselves throughout this paper to finite alphabets even though some of the results extend to infinite alphabets.

B. Definitions

I. *Channels, list codes and capacities*

Let $X = \{1, \cdots, a\}$ be the "input alphabet" and $Y = \{1, \cdots, b\}$ be the "output alphabet" of the channels we shall introduce below. Let $X^t = X$ and $Y^t = Y$ for $t = 1, 2, \cdots$. By $X_n = \prod_{t=1}^{n} X^t$ we denote the set of input n-sequences (words of length n) and by $Y_n = \prod_{t=1}^{n} Y^t$ we denote the set of output n-sequences.

Let $w(\cdot|\cdot)$ be an $a \times b$ stochastic matrix and let $\mathscr{D}$ be a discrete memoryless channel (d.m.c.) with transmission probabilities $P(\cdot|\cdot)$ defined by

$$P(y_n|x_n) = \prod_{t=1}^{n} w(y^t|x^t) \tag{1.1}$$

for every $x_n = (x^1, \cdots, x^n) \in X_n$ and every $y_n = (y^1, \cdots, y^n) \in Y_n$; $n = 1, 2, \cdots$.

Let S be any set and let $W = \{w(\cdot|\cdot|s) | s \in S\}$ be a set of stochastic $a \times b$ matrices. Set $S^t = S$ for $t = 1, 2, \cdots, n$. For every

$$s_n = (s^1, \cdots, s^n) \in \prod_{t=1}^{n} S^t$$

we define $P(\cdot|\cdot|s_n)$ by

$$P(y_n|x_n|s_n) = \prod_{t=1}^{n} w(y^t|x^t|s^t) \tag{1.2}$$

for every $x_n \in X_n$ and every $y_n \in Y_n$.

For every n; $n = 1, 2, \cdots$, define $\mathfrak{A}_n$ by

(1.3) $$\mathfrak{A}_n = \{P(\cdot|\cdot|s_n) | s_n \in S_n\}$$

and $\mathscr{C}_n$ by

(1.4) $$\mathscr{C}_n = \{P(\cdot|\cdot|s_n) | s_n = (s, \cdots, s),\ s \in S\}.$$

A channel with arbitrarily varying channel probability functions (a.v.ch.) $\mathfrak{A}$ is defined by the sequence $(\mathfrak{A}_n)_{n=1,2,\cdots}$ and a compound channel (c.ch.) $\mathscr{C}$ is defined by the sequence $(\mathscr{C}_n)_{n=1,2,\cdots}$.

(1.5) Let L be a positive integer. A list code (n, N, L) is a system

$$\{(u_i, A_i) | i = 1, \cdots, N\},$$

where $u_i \in X_n$, $A_i \subset Y_n$ and $\sum_{i=1}^{N} 1_{A_i}(y_n) \leqq L$ for all $y_n \in Y_n$. 1_B denotes the indicator function of a set B.

(1.6) A code (n, N, L) is a λ-code (n, N, L, λ)

(a) for the d.m.c. $\mathscr{D}$, if $P(A_i|u_i) \geqq 1 - \lambda$ for $i = 1, \cdots, N$;

(b) for the a.v.ch. $\mathfrak{A}$, if $P(A_i|u_i|s_n) \geqq 1 - \lambda$ for $i = 1, \cdots, N$ and all $s_n \in S_n$;

(c) for the c.ch. $\mathscr{C}$, if $P(A_i|u_i|s_n) \geqq 1 - \lambda$ for $i = 1, \cdots, N$ and all $P(\cdot|\cdot|s_n) \in \mathscr{C}_n$.

(1.7) We call a number $\mathrm{K}_l \geqq 0$ the list code capacity or l-capacity of a channel, if

(a) for any $\varepsilon > 0$, $\delta > 0$ and λ, $0 < \lambda < 1$, there exists a λ-code $(n, e^{n(K_l - \delta)}, e^{\varepsilon n}, \lambda)$ for all sufficiently large n, and

(b) for any $\delta > 0$ and λ, $0 < \lambda < 1$, there exists no ε, $0 < \varepsilon < \delta$, such that there exists a λ-code $(n, e^{n(K_l + \delta)}, e^{cn}, \lambda)$ for all sufficiently large n.

(1.8) We call a number $D_{0l} \geqq 0$ the list code zero error capacity of a d.m.c. $\mathscr{D}$, if D_{0l} satisfies (1.7) for $\lambda = 0$.

Similar to the feedback case, [5], we determine first the l-capacity A_l for the a.v.ch. $\mathfrak{A}$. The result for D_{0l} follows by specialization. D_{0l} is by definition the maximal rate which can be achieved on a d.m.c. with 0 error probability for list codes of not exponentially increasing list length .Actually we show that one can achieve D_{0l} for codes of a list length smaller than n, the word length.

II. *Entropy and rate functions*

(1.9) The entropy of a probability vector $p = (p_1, \cdots, p_c)$ is defined to be

$$H(p) = -\sum_{i=1}^{c} p_i \log p_i.$$

(1.10) The "rate" for the probability vector π on X and matrix $w(\cdot|\cdot)$ is

$$R(\pi, w(\cdot\,|\,\cdot)) = H(q(w)) - \sum_i \pi_i H(w(\cdot\,|\,i)),$$

where $q(w) = \pi w(\cdot\,|\,\cdot)$.

(1.11) For π and $w(\cdot\,|\,\cdot)$ define a $b \times a$ stochastic matrix $w^*(\cdot\,|\,\cdot)$ by

$$w^*(i\,|\,j) = \pi_i w(j\,|\,i)/q_j(w); j = 1, \cdots, b; i = 1, \cdots, a.$$

It is well known and easy to verify that

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

The ordinary capacity D of the d.m.c. $\mathscr{D}$ is given by

(1.13) $$D = \max_\pi R(\pi, w(\cdot\,|\,\cdot)).$$

(1.14) For $i \in X$ denote the closed convex hull of the set of probability vectors $\{w(\cdot\,|\,i)\,|\,w \in W\}$ by $\bar{\bar{W}}(i)$ and set

$$\bar{\bar{W}} = \{w(\cdot\,|\,\cdot)\,|\,w(\cdot\,|\,i) \in \bar{\bar{W}}(i) \text{ for all } i \in X\}.$$

$\bar{\bar{W}}$ is called the row convex closure of the set of matrices W.

We define two quantities C and A by

(1.15) $$C = \max_\pi \inf_{w \in W} R(\pi, w)$$

and

(1.16) $$A = \max_\pi \min_{w \in \bar{\bar{W}}} R(\pi, w).$$

C is the ordinary capacity of the c.ch. $\mathscr{C}$ and A is the capacity of the a.v.ch. $\mathfrak{A}$ in the case of noiseless feedback, provided the capacity is positive (see [5]).

2. Auxiliary results

First, we restate results, which were obtained in [4] and [5], in terms of list codes.

Lemma 1. Let $\mathscr{D}$ be a d.m.c., π a probability distribution (p.d.) on X, l a positive integer and ε a positive number. One can construct a list code (l, N, L, λ) for $\mathscr{D}$, such that:

(a) $$N \geqq \exp\{H(\pi)l - f(\pi, a)\log l\};$$

(b) $$L \leqq \exp\left\{\sum_{j=1}^{b} q_j(w)\, H(w^*(\cdot\,|\,j))\, l + g(\varepsilon)l\right\};$$

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

The functions $f(\pi, a)$, $g(\varepsilon)$ and $E(\varepsilon, \pi, w)$ can be given explicitly. $E(\varepsilon, \pi, w)$ is positive and $\lim_{\varepsilon \to 0} g(\varepsilon) = 0$. (Compare Lemma 1 and (2.11) of [4] and the error calculation given there.)

In [5], Lemma 1, we obtained a generalization, which we state as Lemma 2.

Lemma 2. Let $\mathfrak{A}$ be an a.v.ch., π a p.d. on X, l a positive integer and ε a positive number. One can construct a list code (l, N, L, λ) for $\mathfrak{A}$, such that:

(a) $$N \geqq \exp\{H(\pi)l - f(\pi, a)\log l\};$$

(b) $$L \leqq \exp\left\{\max_{w \in \bar{\bar{W}}} \sum_{j=1}^{b} q_j(w) H(w^*(\cdot \mid j))l + \bar{\bar{g}}(\varepsilon)l\right\};$$

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

$\bar{\bar{g}}(\varepsilon)$ can be given explicitly and $\lim_{\varepsilon \to 0} \bar{\bar{g}}(\varepsilon) = 0$;

$$E(\varepsilon, \pi) = \min_{w \in \bar{\bar{W}}} E(\varepsilon, \pi, w) > 0.$$

The methods in [5] yield a similar result for the c.ch. Lemma 2 can be expressed for this channel, if one changes (b) into

(2.1) $$L \leqq \exp\left\{\sup_{w \in \bar{W}} \sum_{j=1}^{b} q_j(w) H(w^*(\cdot \mid j))l + \bar{\bar{g}}(\varepsilon)l\right\}.$$

For an understanding of our later arguments, familiarity with the proofs of Lemmas 1 and 2 is not necessary.

Lemma 3. Let $\mathfrak{D}$ be a d.m.c. with transmission matrix w and let W_0 be a set of stochastic $0-1$ matrices given by

$$W_0 = \{\bar{w} \mid \bar{w} \text{ stochastic } 0-1 \text{ matrix}, \ \bar{w}(j \mid i) = 0 \text{ if } w(j \mid i) = 0\}.$$

For the a.v.ch. $\mathfrak{A}_0$—determined by W_0—the following statements hold:

(a) an (n, N, L, λ) code for $\mathfrak{A}_0$ is an $(n, N, L, 0)$ code for $\mathfrak{A}_0$;

(b) an $(n, N, L, 0)$ code for $\mathfrak{A}_0$ is an $(n, N, L, 0)$ code for $\mathfrak{D}$, and conversely.

Proof. (a) follows from the fact that W_0 contains only $0-1$ matrices. (b) is a consequence of the fact that w can be written as a convex combination of matrices in W_0. (cf. Lemma in [6]. There we used ordinary codes, but the same argument extends to list codes.)

Denote the cardinality of a set G, say, by $|G|$.

Lemma 4. Let N, M, L and t be non-negative integers such that $M \cdot L \leqq tl$. Denote the set $\{1, \cdots, N\}$ by $\bar{N}$ and the set $\{1, \cdots, L\}$ by $\bar{L}$. For any system $\{T_1, \cdots, T_M\}$ of M subsets of $\bar{N}$, which satisfies $|T_j| \leqq \bar{L} (j = 1, \cdots, M)$, there exists a mapping Φ from $\bar{N}$ into $\bar{L}$ such that $|\Phi^{-1}(i) \cap T_j| < t$ for all $i \in \bar{L}$ and $j = 1, \cdots, M$.

Proof. For $A \subset \bar{N}$ denote by $\mathscr{F}(A)$ the set of all mappings from $\bar{N}$ into $\bar{L}$, which are constant on A. Set

$$\mathscr{F}_j^t = \bigcup_{A \subset T_j, |A| \geq t} \mathscr{F}(A)$$

and set

$$\mathscr{F}^t = \bigcup_{j=1}^{M} \mathscr{F}_j^t.$$

Denote the set of all mappings from $\bar{N}$ into $\bar{L}$ by $\mathscr{F}$. It suffices to prove that $|\mathscr{F}^t|\,|\mathscr{F}|^{-1} < 1$. Since $|\mathscr{F}| = L^N$ and

$$|\mathscr{F}^t| \leqq M \binom{L}{t} \cdot L \cdot L^{N-t},$$

we obtain that

$$|\mathscr{F}^t|\,|\mathscr{F}|^{-1} \leqq M \binom{L}{t} L^{1-t} \leqq ML/t! < 1,$$

by assumption.

(This Lemma will provide a substitute for the lack of feedback and will enable us to make an iterative list reduction.)

We give now the definitions needed in order to state the elementary Lemma 5. Let p be a non-negative number, let r be larger than p and let l be a positive integer. Set $l_1 = l$. Define now for every positive integer i an l_i inductively as the smallest integer for which

$$p^{l_1 - 1} \leqq r^{l_i}. \tag{2.2}$$

Obviously, $l_1 \geqq l_2 \geqq l_3 \geqq \cdots$. In the following we use "[ ]" as the smallest integer larger than the number in brackets.

Lemma 5. Set $Q = \log p(\log r)^{-1}$ and $Q^* = (1 - Q)^{-1}$. For

$$I = 1 + [-(\log Q)^{-1} \log l]$$

we have:

(a) $$l_I \leqq 1 + Q^*;$$

(b) $$\sum_{i=1}^{I} l_i \leqq lQ^* + IQ^*.$$

Proof. By (2.2) and the definition of Q we have $l_2 \leqq l_1 Q + 1$ and generally

$$l_i \leqq l_1 Q^{i-1} + Q^{i-2} + \cdots + 1$$

for $i = 1, 2, \cdots$. Hence,

$$l_I \leqq l Q^{I-1} + Q^* \leqq 1 + Q^*,$$

by definition of Q^* and I. Furthermore,

$$\sum_{i=1}^{I} l_i \leqq \sum_{i=1}^{I} \left(l Q^{i-1} + \sum_{j=0}^{i-2} Q^j \right) \leqq l Q^* + I Q^*.$$

3. Determination of the list code capacities A_l, D_{0l} and C_l

Let W_0 be as defined in Lemma 3 and use the operation "=" in the sense of Definition (1.14). Our goal is to prove the following theorem.

Theorem. The list code capacities are given by the formulae:

(a) $A_l = \max_{\pi} \min_{w \in \overline{W}} R(\pi, w)$ for the a.v.ch. $\mathfrak{A}$;

(b) $D_{0l} = \max_{\pi} \min_{w \in \overline{W}_0} R(\pi, w)$ for the d.m.c. $\mathscr{D}$;

(c) $C_l = \max_{\pi} \inf_{w \in W} R(\pi, w)$ for the c.ch. $\mathscr{C}$. The expression to the right is known to be the capacity for ordinary coding.

The capacities can actually be achieved with a list length *smaller than n*, the word length.

Remarks. (1) In the case of noiseless feedback one can reduce codes of list size n to ordinary codes provided that the feedback capacities are positive. Denoting these capacities by A_f, D_{0f} and C_f we thus obtain: $A_f \geqq A_l$ if $A_f > 0$; $D_{0f} \geqq D_{0l}$ if $D_{0f} > 0$ and $C_f \geqq C_l$ if $C_f > 0$. Comparing the Theorem with the results of [5] we can actually conclude that $A_f = A_l$ if $A_f > 0$ and $D_{0f} = D_{0l}$ if $D_{0f} > 0$. However, $A_f(D_{0f})$ may be 0 and $A_l(D_{0l})$ still be positive. In the case of list codes the formulae (a) and (b) for A_l and D_{0l} are valid independently of whether A_l or D_{0l} are positive or not. For the d.m.c. $\mathscr{D}$ given by the matrix

$$W = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix},$$

for instance, we have $D_{0l} = \log \frac{3}{2}$ and $D_{0f} = 0$.

(2) For averaged channels, [10], one can show that list coding provides no improvement over ordinary coding, whereas feedback leads in general to a larger capacity. This is the case whenever we have channels with *memory*. We propose therefore as a *problem* to determine feedback capacities and eventually also to construct coding schemes for a large class of channels with memory.

(3) One can introduce zero error list code capacities D_{0l}^L by requiring that the list size be smaller than a *constant* L. For $L = 1$ one obtains Shannon's zero error capacity, [1]. There are examples of channels for which $D_{0l}^2 > D_{0l}^1$.

(4) Forney has introduced in [9] a different notion of list codes. Instead of requiring that the list size is uniformly bounded by L (see Definition (1.5)) he assumes that the *average* list size converges to 1 as the block length n increases.

The average is taken over all messages, which are selected at random according to the uniform distribution, and over all received sequences, which occur at random according to the transmission probabilities. Let us denote by D_{0l}^1 the maximal rate, which can be achieved on the d.m.c. with those codes and with error probabilities zero. Furthermore, let $\theta = (\theta_{ji})$ $(i = 1, \cdots, a; j = 1, \cdots, b)$ be defined by

$$\theta_{ji} = \begin{cases} 0 \text{ if } w(j|i) = 0, \\ 1 \text{ if } w(j|i) > 0. \end{cases}$$

For $D_{0f} > 0$ Shannon proved in [1] that

$$D_{0f} = \max_{\pi}(-\log \max_{j} q_j(0)),$$

where $q_j(0) = \sum_{i=1}^{a} \pi_i \theta_{ji}$, and we showed in [5] that $D_{0f} = \max_{\pi} \min_{w \in W_0} R(\pi, w)$. From Equation (b) in the Theorem we obtain, therefore, that

$$D_{0l} = \max_{\pi}(-\log \max_{j} q_j(0)),$$

if $D_{0f} > 0$. One can actually show analytically that

$$\max_{\pi}(-\log \max_{j} q_j(0)) = \max_{\pi} \min_{w \in W_0} R(\pi, w)$$

always[1]. Forney proved—using methods which are completely different from ours—that

$$D_{0l}^1 \geqq \max_{\pi}\left(-\log \sum_{j=1}^{b} q_j q_j(0)\right)$$

and we can therefore conclude that $D_{0l}^1 \geqq D_{0l}$. D_{0l}^1 is in general larger than D_{0l}. This result is by no means trivial. On the one hand, Forney makes a weaker requirement than we do by considering an *average* list size rather than a uniform list size; on the other hand, he has a stronger condition on the list size than we have. It might be of some interest to decide whether Forney's result can be extended to the a.v.ch. (Forney also introduces for the d.m.c.f. a zero error capacity D_{0F}, say, by allowing a *sequential* coding procedure. He proves that

$$D_{0F} \geqq \max_{\pi}(-\sum_{j} q_j \log q_j(0)).$$

Since

$$\max_{\pi}(-\sum_{j} q_j \log_j(0)) \geqq \max_{\pi}(-\log \sum_{j} q_j q_j(0))$$

$$\geqq \max_{\pi}(-\log \max_{j} q_j(0)),$$

[1] We are grateful to P. Elias for drawing our attention to his publication [13]. In it he derives the formula $D_{0l} = \max_{\pi}(-\log \max_{j} q_j(0))$ using a very elegant argument. However, the same argument does not extend to a.v.ch.

and since in many cases one has strict inequalities, we can conclude that $D_{0F} > D_{0f}$ in those cases.)

(5) So far no results are known about the zero error capacity in the case of a *noisy* feedback channel. Any progress on this problem could also be important for Shannon's zero error capacity problem.

Proof of the Theorem

In order to prove that the quantities listed in (a), (b) and (c) are the capacities in question, we have to prove coding theorems and converses of the coding theorems. We begin with the coding theorems.

A. *The coding scheme.* For $\varepsilon > 0$ let $m_0(\varepsilon)$ be the smallest integer for which $\varepsilon m_0(\varepsilon) \geqq f(\pi, a) \log m_0(\varepsilon)$. We abbreviate $H(\pi)\cdot\varepsilon$ by H and we let $\bar{H}$ stand for

$$\max_{w \in \bar{\bar{W}}} \sum_{j=1}^{b} q_j(w) H(w^*(\cdot \mid j)) + \bar{\bar{g}}(\varepsilon)$$

or for

$$\max_{w \in \bar{\bar{W}}_0} \sum_{j=1}^{b} q_j(w) H(w^*(\cdot \mid j)) + \bar{\bar{g}}(\varepsilon)$$

or for

$$\sup_{W \in \mathcal{W}} \sum_{j=1}^{b} q_j(w) H(w^*(\cdot \mid j)) + \bar{\bar{g}}(\varepsilon)$$

depending on whether we treat the a.v.ch., the d.m.c. with zero error or the c.ch. We shall also write E instead of $E(\varepsilon, \pi)$. With this convention we can say—because of Lemma 2, (2.1) and Lemma 3—that for all channels considered there exists an (m, N_m, L_m, λ_m)-list code

$$\{(u_m(i), A_m(i)) \mid i = 1, \cdots, N_m\}$$

such that

$$N_m = [e^{\Pi m}],\ L_m = [e^{\Pi m}],\ \lambda_m \leqq e^{-Em} \text{ for } m > m_0(\varepsilon). \tag{3.1}$$

For every $y_m \in Y_m$ there exists a list of code words $T(y_m) \subset \{u_m(i) \mid i = 1, \cdots, N_m\}$ into which y_m is decoded. $\{T(y_m) \mid y_m \in Y_m\}$ is the system of possible lists and $|T(y_m)| \leqq L_m$.

Starting for a fixed l with an $(l, N, L, \lambda) = (l, N_l, L_l, \lambda_l)$ list code

$$\{(u_l(i), A_l(i)) \mid i = 1, \cdots, N_l\}$$

for which (3.1) holds we now reduce the list size iteratively by a repeated application of Lemma 4 and (3.1). Set $l_1 = l$ and define a decreasing sequence of integers $l_1, l_2, \cdots$ as in (2.2) with $p = e^H$ and $r = e^{\Pi}$.

We now describe the first step of our reduction. Apply Lemma 4 with

$$N = [e^{Hl}],\ L = [e^{Hl}],\ M = |Y_l| = b^l,$$

to the set $\{u_l(i) \mid i = 1, \cdots, N_l\}$ and to the system of sets $\{T(y_l) \mid y_l \in Y_l\}$. For a suitable constant $l^*(b, \bar{H})$ we have

(3.2) $$b^l e^{Hl} \leqq l!, \text{ if } l \geqq l^*(b, \bar{H}).$$

Thus for $t = l$ the conclusion of Lemma 4 holds, that is, there exists a mapping from $\{u_l(i) \mid i = 1, \cdots, N_l\}$ into any set with at least L elements such that this mapping assumes on $T(y_l)$, $y_l \in Y_l$, the same value at most l times. Hence,

(3.3) there exists a mapping Φ_{l_1}, say, from $\{u_l(i) \mid i = 1, \cdots, N_l\}$ into $\{u_{l_2}(i) \mid i = 1, \cdots, N_{l_2}\}$, which assumes on $T(y_{l_1})$, $y_{l_1} \in Y_{l_1}$, the same value at most l times. Generally, for any s with

(3.4) $l_{s+1} \geqq \max(m_0(\varepsilon),\ l^*(b, \bar{H}))$ there exists—because of (3.1) and Lemma 4—a mapping Φ_{l_s} from $\{u_{l_s}(i) \mid i = 1, \cdots, N_{l_s}\}$ into $\{u_{l_{s+1}}(i) \mid i = 1, \cdots, N_{l_{s+1}}\}$, which assumes on $T(y_{l_s})$, $y_{l_s} \in Y_{l_s}$, the same value at most l_s times.

Let s^* be the maximal s for which (3.4) holds and define I^* by

(3.5) $I^* = \min\{I, s^* + 1\}$, where I is as defined in Lemma 5.

The result (3.4), the definitions of s^* and I^* and Lemma 5 yield

(3.6) $$l_{I^*} \leqq \max\{1 + Q^*, [Q^{-1} m_0(\varepsilon)], [Q^{-1} l^*(b, \bar{H})]\} = l_0(\varepsilon, b, \bar{H}),$$

say, and

(3.7) $$\sum_{s=1}^{I^*} l_s \leqq lQ^* + IQ^*.$$

Now suppose that a finite set of messages $\{1, \cdots, N\}$, one of which will be presented to the sender for transmission, is given. We encode message $i \in \{1, \cdots, N\}$ as

(3.8) $$u_i = (u_{l_1}(i),\ \Phi_{l_2} u_{l_1}(i),\ \Phi_{l_2}\Phi_{l_1} u_{l_1}(i), \cdots, \Phi_{l_{I^*}} \cdots \Phi_{l_1} u_{l_1}(i)).$$

What the receiver receives is a matter of chance. Suppose he has received

$$y = (y_{l_1}, y_{l_2}, \cdots, y_{l_{I^*}}).$$

He now associates with this sequence the sequence of lists $(T(y_{l_1}), T(y_{l_2}), \cdots, T(y_{l_{I^*}}))$. Now define sets $V_k (k = 0, 1, \cdots, I^* - 1)$ recursively as follows

(3.9) $$V_0 = T(y_{l_{I^*}}),$$
$$V_k = \Phi^{-1}_{l_{I^*-k}} V_{k-1} \cap T(y_{l_{I^*-k}}).$$

Having received y, the receiver decides on the list $T(y) = V_{I^*-1}$. $T(y)$ is a subset of $\{u_l(i) \mid i = 1, \cdots, N_l\}$ and

$$|T(y)| \leqq |T(y_{l_{I^*}})| \prod_{s=1}^{I^*} l_s. \tag{3.10}$$

Hence, by Lemma 5,

$$|T(y)| \leqq \exp\{l_{I^*} \log a + (1 - (\log Q)^{-1}) \log^2 l\} = L, \tag{3.11}$$

say. Set $n' = \sum_{s=1}^{I^*} l_s$. It follows from $I^* \leqq I$ and from Lemma 5 that

$$n' \leqq lQ^* + 1 - (\log Q)^{-1} \log l. \tag{3.12}$$

With u_i as defined in (3.8) and with $A_i = \{y \mid y \in Y_{n'}, u_i \in T(y)\}$ we obtain an (n', N, L) list code $\{(u_i, A_i) \mid i = 1, \cdots, N\}$.

B. *Refinement of the coding scheme and calculation of the error probability.* Suppose message i is encoded as u_i. Using the scheme $\{(u_i, A_i) \mid i = 1, \cdots, N\}$ a decoding error is made, if for some s, $1 \leqq s \leqq I^*$,

$$\Phi_{l_s} \cdots \Phi_{l_1} u_{l_1}(i) \notin T(y_{l_s}).$$

It follows from (3.1) that the probability for a decoding error λ satisfies

$$\lambda \leqq \sum_{s=1}^{I^*} \lambda_{l_s} \leqq \sum_{s=1}^{I^*} e^{-El_s}. \tag{3.13}$$

Since $l_1 > l_2 > l_3 > \cdots$ the numbers $\lambda_{l_1}, \lambda_{l_2}, \cdots$ are increasing and the bound at the right of (3.13) may be very large. We therefore modify our scheme as in [4] or [5]. This modification is of course unnecessary for the zero error case, because there $\lambda_{l_s} = 0$ for $s = 1, \cdots, I^*$. We now encode i into u_i^*, which is obtained from u_i as follows:

(a) u_i^* has the same first $d(l) = c \log l$, c suitable, components as u_i;

(b) for $s > d(l)$ we repeat every component of u_i $[l^{\frac{1}{4}}]$ times. It was shown in [4], Equations (2.14) to (2.25), that one can choose c such that

$$\lambda \leqq l^{-E^*(\varepsilon) l^{\frac{1}{4}}}, \quad E^*(\varepsilon) \text{ suitable}, \tag{3.14}$$

$$l_s \leqq l^{\frac{1}{4}} \text{ for } s \geqq d \tag{3.15}$$

and

$$n \leqq n' + f^*(\varepsilon) l^{\frac{1}{4}} \log l, \tag{3.16}$$

where n denotes the total number of letters needed and $f^*(\varepsilon)$ is chosen suitably.

It follows from (3.12) and (3.16) that

$$n \leqq lQ^* + \{f^*(\varepsilon) l^{\frac{1}{4}} \log l + 1 - (\log Q)^{-1} \log l\}. \tag{3.17}$$

Since the list size of the modified scheme is the same as the list size of the original scheme, the coding theorems now follow from (3.11), (3.14) and (3.17).

C. *Reduction to list size smaller than n.* Above we obtained an (n,N,L,λ) list code $\{(u_i^*, A_i^*) \mid i = 1, \cdots, N\}$, where $N = [e^{Bl}]$ and L satisfies (3.11). Since L is smaller than $\exp\{K\log^2 l\}$ for some constant K, we obtain by applying Lemma 4 and (3.1) *twice* a list size smaller than $K_1(b)\, l(\log l)^{-1} \cdot \log^2 l$ for l sufficiently large. $K_1(b)$ is a suitable constant.

Now we repeat this double reduction once more and obtain a list size smaller than

$$K_1(b)\, l(\log l)^{-1} \cdot \frac{\log l}{\log \log l} \leqq l$$

for l sufficiently large. Our estimates are obtained by a nearly optimal exploitation of the condition $M \cdot L \leqq tl$ in Lemma 4. It is clear that the number of additional letters needed for our reduction is of small magnitude and that the increase in error probability can also be ignored.

Actually one could continue to reduce the list size from l to $\log l$ and so on. But thus one achieves the capacities only for larger and larger block lengths. It would be of interest to obtain results for a *constant* list size, but those results would have to be obtained by a different approach.

D. *The converses.* Since (b) is a special case of (a), it suffices to prove converses only in the cases (a) and (c).

It is well known (see, for instance, Lemma 4 of [7]) that

$$A = \max_{\pi} \min_{w \in \bar{\bar{W}}} R(\pi, w) = \min_{w \in \bar{\bar{W}}} \max_{\pi} R(\pi, w).$$

Let w' be such that $A = \max_\pi R(\pi, w')$, let $\mathcal{D}'$ be the d.m.c. corresponding to w', and denote its capacity by D'. The strong converse theorem for a d.m.c. $\mathcal{D}$ in the case of list codes (see [12]) says that:

(3.18) one can give a function $c(\lambda)$ explicitly such that there does not exist a list code (n, N, L, λ) with $NL^{-1} > \exp\{Dn + c(\lambda)\sqrt{n}\}$. Since $w' \in \bar{\bar{W}}$ and since $D' = A$, we obtain that $A_l \leqq A$.

The result stated in (3.18) was derived in [12] by using codes of fixed decomposition. The proof for the strong converse theorem for compound channels given in [11] also uses codes of fixed decomposition and can be carried over *verbally* to the case of list codes. One thus obtains that $C_l \leqq C$.

Remark. Our results are concerned with the *existence* of certain codes with certain properties. Our approach leads to no code *construction*, because Lemma 4 gives no construction of the mapping Φ. It would be desirable to know a Φ explicitly. However, since we obtained the existence of Φ by a simple counting argument, one could actually make a random choice according to the uniform

distribution over the set of all mappings from $\bar{N}$ into $\bar{L}$. (For our coding scheme we would have to make about log n such choices.) Thus, we are in a situation which might be compared with the situation in Shannon's random coding method.

References

[1] SHANNON, C. E. (1956) The zero error capacity of a noisy channel. *IRE Trans. Inf. Th.* IT-2, 8–19.

[2] BERGE, C. (1962) *The Theory of Graphs and its Applications.* English translation. Methuen, London.

[3] ELIAS, P. (1955) List decoding for noisy channels. *Technical Report* 335, Research Laboratory of Electronics, M. I. T., Cambridge, Mass.

[4] AHLSWEDE, R. (1972) A constructive proof of the coding theorem for discrete memoryless channels with noiseless feedback. To appear in the *Transactions of the Sixth Prague Conference on Inf. Th., Random Processes and Statistical Decision Functions.*

[5] AHLSWEDE, R. (1973) The capacity of channels with arbitrarily varying channel probability functions in the presence of feedback. *Zeit. Wahrscheinlichkeitsth.* 25, 239–252.

[6] AHLSWEDE, R. (1970) A note on the existence of the weak capacity for channels with arbitrarily varying channel probability functions and its relation to Shannon's zero error capacity. *Ann Math. Statist.* 41, 1027–1033.

[7] AHLSWEDE, R. AND WOLFOWITZ, J. (1970) The capacity of a channel with arbitrarily varying channel probability functions and binary output alphabet. *Zeit. Wahrscheinlichkeitsth.* 15, 186–194.

[8] SHANNON, C. E., GALLAGER, R. G. AND BERLEKAMP, E. (1967) Lower bounds to error probability for coding on discrete memoryless channels. *Inf. and Control* 10, 65–103.

[9] FORNEY, G. D. (1968) Exponential error bounds for erasure, list and decision feedback schemes. *IEEE Trans. Inf. Th.* IT-14, 206–220.

[10] AHLSWEDE, R. (1968) The weak capacity of averaged channels. *Zeit. Wahrscheinlichkeitsth.* 11, 61–73.

[11] WOLFOWITZ, J. (1960) Simultaneous channels. *Arch. Rational Mech. Anal.* 4, 371–386.

[12] NISHIMURA, S. (1969) The strong converse theorem in the decoding scheme of list size *L. Kodai Math. Sem. Rep.* 21, 418–425.

[13] ELIAS, P. (1958) Zero error capacity for list detection. *Quarterly Progress Report No.* 48, Research Laboratory of Electronics, M. I. T., Cambridge, Mass.