The Annals of Mathematical Statistics
1971, Vol. 42, No. 1, 224–240

GROUP CODES DO NOT ACHIEVE SHANNON'S CHANNEL CAPACITY FOR GENERAL DISCRETE CHANNELS[1]

BY R. AHLSWEDE

The Ohio State University

0. Summary. Elias [9], [10] proved that group codes achieve Shannon's channel capacity for binary symmetric channels. This result was generalized by Dobrushin [7] (and independently by Drygas [8]) to discrete memoryless channels satisfying a certain symmetry condition and having a Galois field as alphabet. We prove that group codes do not achieve the channel capacity for general discrete memoryless channels. It therefore makes sense to introduce a group code capacity and to talk about a group coding theorem and its weak and strong converses. The group coding theorem and its weak converse can be established for several reasonable channels such as the discrete memoryless channel, compound channels, and averaged channels. An example of a channel is given for which Shannon's capacity is positive and the group code capacity is zero. Using group codes, one can therefore expect high rates only for channels with a simple probabilistic structure.

1. Preliminaries.

(A) *Channels, probabilistic codes and errors.* Let $X = \{1, \cdots, a\}$ be the "input alphabet" and $Y = \{1, \cdots, a\}$ be the "output alphabet" of the channels we shall study 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 S be any countable set, and let $\mathscr{C} = \{w(\cdot\,|\,\cdot\,|\,s)\,|\,s \in S\}$ be a set if $(a \times a)$-stochastic matrices $w(\cdot\,|\,\cdot\,|\,s)$. For every $s \in S$ we define a discrete memoryless channel (d.m.c.) $P(\cdot\,|\,\cdot\,|\,s)$ by

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

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

Consider now the class of channels

$$\mathscr{C}_n = \{P(\cdot\,|\,\cdot\,|\,s)\,|\,s \in S\}. \tag{1.2}$$

If we are interested in the simultaneous behavior of all these channels we call this indexed set of channels a compound channel [18]. (Sender and receiver communicate without knowing which individual channel actually governs the transmission of any one n-sequence.) Given any probability distribution q on S, then we can define an averaged discrete channel $\bar{P}(\cdot\,|\,\cdot)$ by

$$\bar{P}(y_n\,|\,x_n) = \sum_{s \in S} q_s P(y_n\,|\,x_n\,|\,s) \quad \text{for every} \quad x_n \in X_n, \quad y_n \in Y_n, \tag{1.3}$$

Received December 12, 1969; revised August 12, 1970.

[1] Research of the author supported by the National Science Foundation under Grant Contract No. GP-9464 to The Ohio State University.

([1], [2]). Throughout this paper we shall be concerned only with channels defined under (1.1), (1.2), (1.3).

(1.4) A code (n, N) is a system $\{(u_i, A_i) \mid i = 1, \cdots, N\}$, where $u_i \in X_n$, $A_i \subset Y_n$, $A_i \cap A_j = \varnothing$ for $i \neq j$.

(1.5) A code (n, N) is a λ-code (n, N, λ)

(a) for the d.m.c. $P(\cdot \mid \cdot \mid s)$, if $P(A_i \mid u_i \mid s) \geqq 1 - \lambda$ for $i = 1, \cdots, N$.
(b) for the compound channel $\{P(\cdot \mid \cdot \mid s) \mid s \in S\}$, if

$$P(A_i \mid u_i \mid s) \geqq 1 - \lambda \quad \text{for} \quad i = 1, \cdots, N \quad \text{and for all} \quad s \in S.$$

(c) for the averaged channel $\bar{P}(\cdot \mid \cdot)$, if

$$\bar{P}(A_i \mid u_i) \geqq 1 - \lambda \quad \text{for} \quad i = 1, \cdots, N.$$

(1.6) A code (n, N) is a $\bar{\lambda}$-code $(n, N, \bar{\lambda})$

(a) for the d.m.c. $P(\cdot \mid \cdot \mid s)$, if $1/N \sum_{i=1}^{N} P(A_i \mid u_i \mid s) \geqq 1 - \bar{\lambda}$.
(b) for the compound channel $\{P(\cdot \mid \cdot \mid s) \mid s \in S\}$ if,

$$\inf_{s \in S} 1/N \sum_{i=1}^{N} P(A_i \mid u_i \mid s) \geqq 1 - \bar{\lambda}.$$

(c) for the averaged channel $\bar{P}(\cdot \mid \cdot)$, if $1/N \sum_{i=1}^{N} \bar{P}(A_i \mid u_i) \geqq 1 - \lambda$.

In case (1.5) we talk about maximal errors and in case (1.6) we talk about average errors. In probabilistic coding theory for a single channel it is unimportant whether we work with average or with maximal errors (cf. [17], Lemma 3.1.1). However, for two channels treated simultaneously it makes a difference, as was shown in [1], Example 1, and also in [4]. For group codes the difference becomes even more essential.

(B) *Shannon's channel capacity.*

(1.7) A number $C > 0$ is called (Shannon's) capacity of a channel, if

(a) for any $\delta > 0$ and $\lambda (0 < \lambda < 1)$ there exists a λ-code $(n, 2^{n(C-\delta)}, \lambda)$ for all sufficiently large n, and if (b) for any $\delta > 0$ there exists a $\lambda = \lambda(\delta)$ such that for all sufficiently large n there does not exist a λ-code $(n, 2^{n(C+\delta)}, \lambda)$.

Part (a) is called coding theorem and part (b) is called the weak converse of the coding theorem.

(1.8) C is called (strong or Wolfowitz's) capacity if (a) holds and (b) is replaced by

(b′) for any $\delta > 0$ and $\lambda (0 < \lambda < 1)$ there does not exist a code $(n, 2^{n(C+\delta)}, \lambda)$ for all sufficiently large n.

(a), (b′) imply (a), (b).

(b′) is called the strong converse of the coding theorem. Analogous definitions can be given for $(n, N, \bar{\lambda})$ codes.

(1.9) Let $N(n, \lambda)$ be the maximal length of a (n, N, λ) code and let $N(n, \bar{\lambda})$ be the maximal length of a $(n, N, \bar{\lambda})$ code for the channel in question. (a), (b) are equivalent to

(1.10) $\quad \inf_{\lambda>0} \liminf_{n\to\infty} n^{-1} \log N(n, \lambda) = \inf_{\lambda>0} \limsup_{n\to\infty} n^{-1} \log N(n, \lambda) = C.$

(a), (b′) are equivalent to

(1.11) $\quad \liminf_{n\to\infty} n^{-1} \log N(n, \lambda) = \limsup_{n\to\infty} n^{-1} \log N(n, \lambda) = C \quad$ for all $\quad \lambda,$

$$0 < \lambda < 1.$$

In case (b′) does not hold, one can ask whether

(1.12) $\lim_{n\to\infty} n^{-1} \log N(n, \lambda) = C(\lambda)$ exists for certain λ's. $C(\lambda)$ was introduced in [17]. In [4] it is shown that for compound channels $\lim_{n\to\infty} n^{-1} \log N(n, \bar{\lambda}) = C(\bar{\lambda})$ exists except for at most finitely many $\bar{\lambda}$'s and that $C(\bar{\lambda}_1) \neq C(\bar{\lambda}_2)$ for certain values of $\bar{\lambda}_1$, $\bar{\lambda}_2$, whereas always $C(\lambda) = C$ [16].

(C) *Algebraic codes.* The main criteria for the "goodness" of codes (n, N) for channels are

(1) low error probability
(2) large code length
(3) short encoding and decoding procedures.

It is in general impossible to do best in all three respects and therefore one has to compromise between the different aspects. Various approaches have been given. Probabilistic coding theory usually is concerned with coding procedures which give optimal code length for fixed error probability or optimal error probability for fixed code length. The main goal of algebraic coding theory has been the construction of codes which satisfy (3) and which have certain "error correcting abilities." A central role in the theory is played by group codes (cf. [6], [13]), which were introduced by Hamming [12] and Slepian [14]. We repeat now very briefly some of the basic definitions.

(1.13) We assume that X and Y are Galois fields with $a = p^s$ elements and we identify X_n (respectively Y_n) with the vector space of dimension n over the Galois field X (respectively Y), i.e., we assume that for

$$x_n = (x^1, \cdots, x^n) \in X_n, \qquad \bar{x}_n = (\bar{x}^1, \cdots, \bar{x}^n) \in X_n, \qquad \lambda \in X;$$

$$(x^1, \cdots, x^n) + (\bar{x}^1, \cdots, \bar{x}^n) = (x^1 + \bar{x}^1, \cdots, x^n + \bar{x}^n), \quad \lambda(x^1, \cdots, x^n) = (\lambda x^1, \cdots, \lambda x^n),$$

where the sum $x^i + \bar{x}^i$ and the product λx^i are understood in the sense of the Galois field X.

(1.14) A code (n, N) is said to be a *pseudo group code* if

(a) the set $\{u_1, \cdots, u_N\}$ forms a subgroup of the additive group X_n.
(b) the A_i's are arbitrary.

Denote by φ the canonical isomorphism between X_n and Y_n: for $x_n \in X_n$, $\varphi x_n = y_n$, where $y^t = x^t$ for $t = 1, \cdots, n$.

(1.15) A pseudo group code is a *group code* if there exists a system of representatives (called leaders) $\{l_1, \cdots, l_L\}$ of the cosets of $\{\varphi u_1, \cdots, \varphi u_N\}$ such that $A_i = \{l_1 + \varphi u_i, \cdots, l_L + \varphi u_i\}$ for all $i = 1, \cdots, N$.

(1.16) A group code is called a *linear* (or parity-check) code if $\{u_1, \cdots, u_N\}$ is a subspace of X_n.

(In case $a = p$ every group code is also a linear code.) The additional algebraic structure required for group codes and especially for linear codes allows decoding methods which need fewer computational steps (cf. [6]).

The question remains: Given a channel (for instance one of those described in Section 1(A), how well can we do with group codes and with linear codes with respect to (1) and (2)? As a measure for (1) and (2) we introduce the following quantities:

(1.17) $$N_g(n, \lambda) = \text{maximal length of } (n, N, \lambda)\text{-group codes}$$
$$N_g(n, \bar{\lambda}) = \text{maximal length of } (n, N, \bar{\lambda})\text{-group codes.}$$

The corresponding quantities for linear codes and for pseudo group codes shall be denoted by

$$N_l(n, \lambda), N_l(n, \bar{\lambda}), N_p(n, \lambda), N_p(n, \bar{\lambda}).$$

(1.18)
$$C_g^+ = \inf_{\lambda>0} \limsup_{n\to\infty} n^{-1} \log N_g(n, \lambda).$$
$$C_g^- = \inf_{\lambda>0} \liminf_{n\to\infty} n^{-1} \log N_g(n, \lambda).$$
$$C_l^+ = \inf_{\lambda>0} \limsup_{n\to\infty} n^{-1} \log N_l(n, \lambda).$$
$$C_l^- = \inf_{\lambda>0} \liminf_{n\to\infty} n^{-1} \log N_l(n, \lambda).$$
$$C_p^+ = \inf_{\lambda>0} \limsup_{n\to\infty} n^{-1} \log N_p(n, \lambda).$$
$$C_p^- = \inf_{\lambda>0} \liminf_{n\to\infty} n^{-1} \log N_p(n, \lambda).$$

(1.19) Replacing λ by $\bar{\lambda}$ in (1.18) we get quantities

$$\bar{C}_g^+, \bar{C}_g^-, \bar{C}_l^+, \bar{C}_l^-, \bar{C}_p^+, \bar{C}_p^-.$$

(1.20) In case $C_g^+ = C_g^-$ we talk about the group code capacity (for maximal errors) C_g and in case $\bar{C}_g^+ = \bar{C}_g^-$ about the group code capacity (for average errors) $\bar{C}_g$. Analogously we define C_l, $\bar{C}_l$, C_p, $\bar{C}_p$, and we introduce the names *linear code capacity* and *pseudo group code capacity.*

2. Auxiliary results. Dobrushin defines in [7] a channel with invariant transition probabilities (c.i.t.p.) as a d.m.c. given by a matrix w, which satisfies

(2.1) $$w(j \mid i) = w(j+k \mid i+k) \quad \text{for all} \quad i, j, k \in \{1, \cdots, a\}.$$

We state his main result in our terminology as

LEMMA 1. *Let $a = p^k$ and $X = GF(a)$. For a* c.i.t.p. *the equalities $C = C_g = \bar{C}_g$ hold.*

Lemma 1 generalizes a theorem of Elias [9], saying that $C = C_g = \bar{C}_g$ for binary symmetric channels.

LEMMA 2. *Let $a = p^k$ and $X = GF(a)$. For a* c.i.t.p. *the equalities $C = C_l = \bar{C}_l$ hold.*

This result is due to Drygas [8]. It implies Lemma 1.

(2.2). We call the stochastic matrix $(\tilde{w}(i|j))$ $i = 1, \cdots, a$; $j = 1, \cdots, a$ a-ary symmetric, if

$$\tilde{w}(j|i) = 1-\varepsilon \quad \text{for} \quad i = j,$$

$$= \frac{\varepsilon}{a-1} \quad \text{for} \quad i \neq j.$$

(2.3). The d.m.c. $P(\cdot|\cdot)$ determined by $w(\cdot|\cdot)$ is called an a-ary symmetric channel.

(2.4) We say the (n, N) code $\{(u_i, A_i) | i = 1, \cdots, N\}$ is a *strict* maximum likelihood code (s.m.l.c.) with respect to $P(\cdot|\cdot)$, if

$$A_i = \{y_n | y_n \in Y_n \text{ and } P(y_n|u_i) > P(y_n|u_j) \text{ for } j \neq i\} \quad \text{for} \quad i = 1, \cdots, N.$$

(2.5) $w(\cdot|\cdot|s^*)$ is dominated by the a-ary symmetric matrix $\tilde{w}(\cdot|\cdot)$ if $w(i|i|s^*) \geqq 1-\varepsilon$ for $i = 1, \cdots, a$.

Accordingly we say $P(\cdot|\cdot|s^*)$ is dominated by $\tilde{P}(\cdot|\cdot)$ if $w(\cdot|\cdot|s^*)$ is dominated by $\tilde{w}(\cdot|\cdot)$.

After these preparations we can state

LEMMA 3. *Let $P(\cdot|\cdot|s^*)$ be dominated by the a-ary symmetric channel $\tilde{P}(\cdot|\cdot)$ and let $\{(u_i, A_i) | i = 1, \cdots, N\}$ be a* s.m.l.c. *with respect to $\tilde{P}(\cdot|\cdot)$, then*

$$P(A_i|u_i|s^*) \geqq \tilde{P}(A_i|u_i) \quad \text{for} \quad i = 1, \cdots, N. \tag{2.6}$$

This was proved in [3], Lemma 1.

(2.7) As usual, the Hamming distance between n-sequences is defined as $h(x_n, y_n)$ = number of components in which x_n and y_n are different. We shall write $h(x_n, y_n) = h(x_n)$, if y_n is the zero vector, and call $h(x_n)$ the weight of x_n.

Let $\{(u_i, A_i) | i = 1, \cdots, N\}$ be a group code for the a-ary symmetric channel with maximal error λ, where $0 < \lambda < \frac{1}{2}$. Furthermore, we assume that there exist coset leaders of $\{\varphi u_1, \cdots, \varphi u_N\}$ with minimal weight, which we denote by $l_1, \cdots, l_L$, such that $A_i = \{\varphi u_i + l_j | j = 1, \cdots, L\}$ for $i = 1, \cdots, N$.

It follows from the definition of the a-ary symmetric channel that the transition probability $\tilde{P}(y_n|x_n)$ depends only on $h(x_n, y_n)$, and that $P(\varphi u_i + l_j | u_i)$ has the same value for all $i = 1, \cdots, N$. Consequently,

$$\tilde{P}(A_i|u_i) = \tilde{P}(A_j|u_j) \geqq 1-\lambda \quad \text{for} \quad i, j = 1, \cdots, N. \tag{2.8}$$

However $\{(u_i, A_i) | i = 1, \cdots, N\}$ is not necessarily a s.m.l.c. with respect to $\tilde{P}(\cdot|\cdot)$.

(2.9) Define $B_i = \{y_n | \tilde{P}(y_n|u_i) > \tilde{P}(y_n|u_j) \text{ for all } j \neq i\}$.

It follows from the definitions of A_i and B_i that

$$B_i \subset A_i \tag{2.10}$$

and

$$\{(u_i, B_i) \mid i = 1, \cdots, N\} \quad \text{is a s.m.l.c. with respect to} \quad \tilde{P}(\cdot \mid \cdot). \tag{2.11}$$

LEMMA. 4. $\tilde{P}(B_i \mid u_i) \geqq 1-2\lambda$ *for* $i = 1, \cdots, N$.

PROOF. We describe our decoding scheme by the matrix

$$\mathscr{L} = \begin{pmatrix} l_1 + \varphi u_1, \cdots, l_1 + \varphi u_N \\ \vdots \\ l_L + \varphi u_1, \cdots, l_L + \varphi u_N \end{pmatrix}.$$

The elements of the ith column of $\mathscr{L}$ constitute A_i.

(a) Choose u_1 to be the zero vector. $\varphi u_1 + l_j$ is not in B_1 if there exists a $k \neq 1$ such that $\tilde{P}(\varphi u_1 + l_j \mid u_k) = \tilde{P}(\varphi u_1 + l_j \mid u_1)$. Consequently $h(l_j) = h(u_1 + l_j - u_k) = h(u_1 + l_j + u_k{}^*)$, where $u_k{}^* = -u_k$. Therefore, and because u_1 is the zero vector we have $\tilde{P}(\varphi u_k{}^* + l_j \mid u_1) = \tilde{P}(\varphi u_1 + l_j \mid u_1)$. That means if $\varphi u_1 + l_j$ is not in B_1 then there exists an element $\varphi u_k{}^* + l_j$ which

(1) lies also in the jth row;
(2) is not in A_1; and
(3) satisfies $\tilde{P}(\varphi u_k{}^* + l_j \mid u_1) = \tilde{P}(\varphi u_1 + l_j \mid u_1)$.

It follows from (1), (2), (3) and $\tilde{P}(A_1 \mid u_1) \geqq 1-\lambda$ that

$$\tilde{P}(B_1 \mid u_1) \geqq 1-2\lambda. \tag{2.12}$$

(b) We omit now in our decoding scheme all rows with an index j for which $\varphi u_1 + l_j \notin B_1$. We denote the set of remaining elements in the kth column by $\bar{B}_k$ and show now that

$$\bar{B}_k \subset B_k \quad \text{for} \quad k = 2, \cdots, N. \tag{2.13}$$

Assume (2.13) does not hold for $k = r$. Then we have an element

$$\varphi u_r + l_i \in \bar{B}_r \quad \text{and} \quad \varphi u_r + l_i \notin B_r.$$

Consequently there exists an element $u_s \neq u_r$ such that

$$\tilde{P}(\varphi u_r + l_i \mid u_s) = \tilde{P}(\varphi u_r + l_i \mid u_r) \quad \text{and therefore} \quad h(u_r - u_s + l_i) = h(l_i).$$

Defining $u_t = u_r - u_s$, we get $h(\varphi u_t + l_i, u_1) = h(\varphi u_1 + l_i, u_1)$. $t \neq 1$ would imply that the ith row is excluded, in contradiction to $\varphi u_r + l_i \in \bar{B}_r$. Consequently, $u_t = u_1$, and therefore $u_r = u_s$ in contradiction to our assumption $u_s \neq u_r$. That proves (2.13). (2.12), (2.13), and $\tilde{P}(\bar{B}_i \mid u_i) = \tilde{P}(B_1 \mid u_1)$ imply that $\tilde{P}(B_i \mid u_i) > 1-2\lambda$ for $i = 1, \cdots, N$.

(2.14) The entropy of a probability vector $\pi = (\pi_1, \cdots, \pi_c)$ is defined to be $H(\pi) = -\sum_{i=1}^{c} \pi_i \log \pi_i$.

(2.15) Denote the "rate" for the probability vector π on X and matrix $w(\cdot \mid \cdot \mid s)$ by $R(\pi, w(\cdot \mid \cdot \mid s)) = H(\pi'(s)) - \sum_{i=1}^{a} \pi_i H(w(\cdot \mid \cdot \mid s))$, where $\pi'(s) = \pi \cdot w(\cdot \mid \cdot \mid s)$.

Now we can state

LEMMA 5. *The capacity of the a-ary symmetric channel is*

$$C = \max R(\pi, w(\cdot \mid \cdot)) = \log a + (1-\varepsilon)\log(1-\varepsilon) + \varepsilon \log \frac{\varepsilon}{a-1}.$$

PROOF. $C = \max_\pi R(\pi, w(\cdot \mid \cdot))$ is a well-known formula for the channel capacity of a d.m.c. The second equality follows from straightforward computation.

LEMMA 6.

(a) *For the compound channel defined under* (1.2) *we have*

$$C = \max_\pi \inf_{s \in S} R(\pi, w(\cdot \mid \cdot \mid s)). \tag{2.16}$$

(b) *Let us assume that the probability distribution q on S satisfies $q_s > 0$ for all $s \in S$. Then we have for the averaged channel defined under* (1.3)

$$C = \max_\pi \inf_{s \in S} R(\pi, w(\cdot \mid \cdot \mid s)). \tag{2.17}$$

(a) was proved in [16], (b) was proved in [2].

Let $\{u_i = (u_i^1, \cdots, u_i^n) \mid i = 1, \cdots, N\}$ be any system of code words in X_n. This system induces a probability distribution π^t on $X^t (t = 1, 2, \cdots, n)$ given by

$$\pi_i^t = \frac{\left|\{u_j^t \mid u_j^t = i, j \in \{1, \cdots, N\}\}\right|}{N} \qquad (i = 1, \cdots, a). \tag{2.18}$$

LEMMA 7. *Given a (n, N, λ)-code $\{(u_i, A_i) \mid i = 1, \cdots, N\}$ for the* d.m.c. *determined by $w(\cdot \mid \cdot)$. For any $b (0 < b < 1)$ the estimate*

$$\log N \leqq \sum_{t=1}^n \sum_{i=1}^a \sum_{j=1}^a \pi_i^t w(j \mid i) \log \frac{w(j \mid i)}{\sum_{h=1}^a \pi_h^t w(j \mid h)} + k(\lambda, w, b) n^{\frac{1}{2}} - \log(1-\lambda) b$$

holds, where π^t is defined as under (2.18) *and $k(\lambda, w, b)$ is a known function.*

This result is due to Augustin ([5], Satz 8.2).

(Unfortunately, this estimate does not hold for $(n, N, \bar{\lambda})$-codes. This can be seen from the following argument: Assume the estimate extends to $(n, N, \bar{\lambda})$-codes, then one could derive the strong converse for compound channels for average errors in contradiction to [1] example 1, and [4] Theorem 1).

3. The existence of the group code capacity for discrete memoryless channels, compound channels and averaged channels. We recall the definitions given in paragraph 1, especially those in Section 1 (C).

THEOREM 1. *Let $a = p^s$, where p is prime and s a positive integer, and let $X = Y = GF(a)$.*

Then we have for a d.m.c.

$$C_g^- = C_g^+ = C_g.$$

(The group code capacity exists for maximal errors).

PROOF. We shall use Lemmas 1, 3, 4, 5. Clearly C_g^- and C_g^+ exist, because $\log N_g(n, \lambda) \leqq n \log a$. For $\delta > 0$ and $\varepsilon (0 < \varepsilon < \frac{1}{2})$ there exists a $k = k(\delta, \varepsilon)$ such that $1/k \log N_g(k, \varepsilon) \geqq C_g^+ - \delta/4$. Let $\{(u_i, A_i) \mid i = 1, \cdots, N_g(k, \varepsilon)\}$ be a $(k, N_g(k, \varepsilon), \varepsilon)$ group code for the d.m.c. $P(\cdot \mid \cdot)$.

(3.1) Let $\qquad w'(j \mid i) = P(A_j \mid u_i) \quad$ for $\quad i, j = 1, \cdots, N_g(k, \varepsilon)$.

(3.2) Using the mappings ψ and χ, where $\psi u_i = i$ and $\chi A_i = i$, we let $\overline{X} = \{\psi u_i \mid i = 1, \cdots, N_g(k, \varepsilon)\}$ be the input alphabet and $\overline{Y} = \{\chi A_i \mid i = 1, \cdots, N_g(k, \varepsilon)\}$ be the output alphabet of the d.m.c. $P'(\cdot \mid \cdot)$ determined by the stochastic matrix $w'(\cdot \mid \cdot)$. The set $\{u_i \mid i = 1, \cdots, N_g(k, \varepsilon)\}$ is a subgroup of X_k and therefore isomorphic to a finite direct sum of cyclic groups of order p. The mappings ψ and χ induce this group structure in $\overline{X}$ and $\overline{Y}$.

(3.3) We denote the $N_g(k, \varepsilon)$-ary symmetric channel with alphabets $\overline{X}$ and $\overline{Y}$ by P^*.

Let now $\{(v_j, D_j) \mid j = 1, \cdots, N^*\}$ be a $(t, N^*, \lambda/2)$ group code for P^*. In accordance with Lemma 4 we can modify the sets D_j such that $\{(v_j, D_j) \mid j = 1, \cdots, N^*\}$ is an s.m.l.c. with respect to P^* with maximal error λ. P' is dominated by P^*. As a consequence of Lemma 3 we get that

(3.4) $\{(v_j, D_j) \mid j = 1, \cdots, N^*\}$ is a (t, N^*, λ) code for P' and *a fortiori* that

(3.5) $\{(v_j, D_j) \mid j = 1, \cdots, N^*\}$ is a (t, N^*, λ) group code for P'.

Let $N^*(t, \lambda)$ be the maximal length of a (t, N^*, λ) group code for P^*. It follows from Lemma 1 and Lemma 5 that

(3.6) $\quad N^*(t, \lambda/2) > \exp\{(1-\varepsilon)\log N_g(k, \varepsilon) + (1-\varepsilon)\log(1-\varepsilon) + \varepsilon \log \varepsilon - \eta\} t$

for t sufficiently large. (3.5) and (3.6) imply that

(3.7) there exists a (t, N^*, λ) group code for P' with

$$N^* > \exp\{(1-\varepsilon)\log N_g(k, \varepsilon) + (1-\varepsilon)\log(1-\varepsilon) + \varepsilon \log \varepsilon - \eta\} t$$

for t sufficiently large.

We have now to embed $\{v_j \mid j = 1, \cdots, N^*\}$ into X_{tk} and $\{D_j \mid j = 1, \cdots, N^*\}$ into Y_{tk} in order to get a group code for the original channel $P(\cdot \mid \cdot)$. The v_j are sequences of length t whose components are elements of $\{1, \cdots, N_g(k, \varepsilon)\}$. Define $\psi_t^{-1} v_j = \psi_t^{-1}(v_j^1, \cdots, v_j^t)$ as $(\psi^{-1} v_j^1, \cdots, \psi^{-1} v_j^t)$ and call the image $v_j{}^{**}$. $v_j{}^{**}$ has now elements of $\{u_i \mid i = 1, \cdots, N_g(k, \varepsilon)\}$ as components.

Define now $v_j{}^*$ as sequence of length tk, whose components coincide with the components of $\psi^{-1} v_j^1$ in the first k places, with the components of $\psi^{-1} v_j^2$ in the places $k+1, \cdots, 2k$ and so on. $v_j{}^*$ is an element of X_{tk}. For an element $z \in D_j$, $z = (z^1, \cdots, z^t)$, where $z^\tau \in \{1, \cdots, N_g(k, \varepsilon)\}$ for $\tau = 1, \cdots, t$, we define $z^{**} = \chi_t^{-1} z$ as $(\chi^{-1} z^1, \cdots, \chi^{-1} z^t)$, where $\chi^{-1} z^\tau \in \{A_i \mid i = 1, \cdots, N_g(k, \varepsilon)\}$. z^{**} is a sequence of A_i's of length t. Take now the Cartesian product of the components of z^{**} and

embed this set in a canonical way into Y_{tk}. Call this set z^* and define the decoding sets A_j^* as $\bigcup_{z \in D_j} z^*$. Obviously

$$P(A_j^* \mid v_j^*) \geqq 1-\lambda. \tag{3.8}$$

We have to show now that there exist coset leaders $l_1^*, \cdots, l_L^*$ of $\{v_j^* \mid j = 1, \cdots, N^*\}$ such that $A_j^* = \{\varphi v_j^* + l_i^* \mid i = 1, \cdots, L^*\}$. Take the coset leaders of the group code $\{(v_j, D_j) \mid j = 1, \cdots, N^*\}$. Their images in Y_{tk} are now sets which are Cartesian products of A_i's. Choose all their elements to be the coset leaders $\{l_1^*, \cdots, l_{L^*}^*\}$ of $\{\varphi v_j^* \mid j = 1, \cdots, N^*\}$. Then we have $A_j^* = \{\varphi v_j^* + l_i^* \mid i = 1, \cdots, L^*\}$. (3.7), (3.8) imply now

$$N_g(tk, \lambda) > \exp\{(1-\varepsilon)\log N_g(k, \varepsilon) + (1-\varepsilon)\log(1-\varepsilon) + \varepsilon \log \varepsilon - \eta\}t \tag{3.9}$$

for t sufficiently large, and

$$\begin{aligned} &\frac{1}{tk}\log N_g(tk, \lambda) \\ &\geqq \frac{1}{k}\{(1-\varepsilon)\log N_g(k, \varepsilon) + (1-\varepsilon)\log(1-\varepsilon) + \varepsilon\log\varepsilon - \eta\} \qquad (3.10) \\ &\geqq (1-\varepsilon)\left(C_g^+ - \frac{\delta}{4}\right) + \frac{1}{k}\{(1-\varepsilon)\log(1-\varepsilon) + \varepsilon\log\varepsilon - \eta\} \\ &\geqq C_g^+ - \delta/2 \quad \text{for} \quad \varepsilon, \eta \quad \text{sufficiently small and} \quad t \quad \text{sufficiently large.} \end{aligned}$$

Every nonnegative integer can be written as $n = tk + l$, where $0 \leqq l < k$. Using $N_g(tk+l, \lambda) \geqq N_g(tk, \lambda)$ and $\lim_{t\to\infty}(tk+l)/tk = 1$ we get $1/n \log N_g(n, \lambda) \geqq C_g^+ - \delta$ for all sufficiently large n and therefore $C_g^+ = C_g^- = C_g$.

REMARKS.

(1) The existence of C_p can be proved in the same way.

(2) The continuity of C_g, C_p as functions of w can be proved by using the fact that the channel capacity of the a-ary symmetric channel is continuous in ε.

THEOREM 2. *Let $X = Y = GF(a)$, $a = p^s$. Then we have for the compound channel* $\{P(\cdot \mid \cdot \mid s) \mid s \in S\}$

$$C_g^+ = C_g^- = C_g.$$

PROOF. The proof is essentially the same as the proof for Theorem 1. For $\delta > 0$ and $\varepsilon (0 < \varepsilon < 1)$ there exists a $k = k(\delta, \varepsilon)$ such that $1/k \log N_g(k, \varepsilon) \geqq C_g^+ - \delta/4$. Let $\{(u_i, A_i) \mid i = 1, \cdots, N_g(k, \varepsilon)\}$ be a $(k, N_g(k, \varepsilon), \varepsilon)$ group code for the compound channel $\{P(\cdot \mid \cdot \mid s) \mid s \in S\}$. We use $\{\psi u_i \mid i = 1, \cdots, N(k, \varepsilon)\}$ as input alphabet and $\{\chi A_i \mid i = 1, \cdots, N_g(k, \varepsilon)\}$ as output alphabet of the compound channel $\{P'(\cdot \mid \cdot \mid s) \mid s \in S\}$ determined by the class of stochastic matrices $\{w'(\cdot \mid \cdot \mid s) \mid s \in S\}$, where $w'(j \mid i \mid s) = P(A_j \mid u_i \mid s)$ for $i, j = 1, \cdots, N_g(k, \varepsilon)$. Define the d-ary symmetric channel P^* as above. A $(t, N, \lambda/2)$ group code for P^* corresponds to a (tk, N, λ)

group code for $\{P(\cdot\,|\,\cdot\,|\,s)\,|\,s\in S\}$. The remaining steps are the same as in the proof for Theorem 1.

For convenience we choose S to be the set of natural numbers $\mathbb{N}$.

(3.11) We denote the group capacity of $\{P(\cdot\,|\,\cdot\,|\,s)\,|\,s=1,\cdots,l\}$ by $C_g(l)$.

THEOREM 3. $X = Y = GF(a)$, $a = p^s$.

$P(\cdot\,|\,\cdot) = \sum_{s\in\mathbb{N}} q_s P(\cdot\,|\,\cdot\,|\,s)$ $\qquad (q_s > 0$ *for all* $s\in\mathbb{N})$.

We have $C_g^+ = C_g^- = C_g = \inf_l C_g(l)$.

PROOF.

(a) First we prove $C_g^- \geqq \inf_l C_g(l)$. Given $\gamma > 0$, $\lambda > 0$ choose $l^*(\gamma, \lambda)$ such that $|C_g(l^*) - \inf_l C_g(l)| \leqq \gamma$ and $\sum_{\kappa=1}^{l^*} q_\kappa \geqq 1-\lambda/2$.

A code with error $\lambda/2$ for the compound channel $\{P(\cdot\,|\,\cdot\,|\,s)\,|\,s=1,\cdots,l^*\}$ is then a code for $P(\cdot\,|\,\cdot)$ with an error less than $1-(1-\lambda/2)(1-\lambda/2) \leqq \lambda$. The maximal length of a (n, N, λ) group code for $\{P(\cdot\,|\,\cdot\,|\,s)\,|\,s=1,\cdots,l^*\}$ may be denoted by $N^{l^*}(n, \lambda)$. Then we have

$$N_g(n,\lambda) \geqq N^{l^*}(n, \lambda/2), \tag{3.12}$$

$$\liminf_{n\to\infty} n^{-1}\log N_g(n,\lambda) \geqq \lim_{n\to\infty} n^{-1}\log N^{l^*}(n,\lambda/2) \geqq C_g(l^*), \tag{3.13}$$

and consequently

$$C_g^- = \inf_{\lambda>0} \liminf_{n\to\infty} n^{-1}\log N_g(n,\lambda) \geqq \inf_l C_g(l).$$

(b) We prove now

$$C_g^+ \leqq \inf_l C_g(l).$$

Fix any $l\in N$. Define $\eta_l = \inf_{\kappa=1,\cdots,l} q_\kappa > 0$. Choose $\lambda_\nu = \nu^{-1}\eta_l$ for $\nu = 2, 3, \cdots$. Suppose $\{(u_i, A_i)\,|\,i=1,\cdots,N\}$ is a λ_ν-code for $P(\cdot\,|\,\cdot)$. Then

$$q_\kappa P(A_i\,|\,u_i\,|\,\kappa) \geqq 1 - \sum_{s\neq\kappa} q_s - \lambda_\nu \quad \text{for all} \quad \kappa\in\mathbb{N}, i = 1,\cdots,N,$$

and therefore

$$P(A_i\,|\,u_i\,|\,\kappa) \geqq 1-\lambda_\nu/q_\kappa \geqq 1-1/\nu \qquad \text{for} \quad \kappa = 1,\cdots,l; i = 1,\cdots,N. \tag{3.14}$$

Letting ν tend to infinity we get

$$C_g'^+ \leqq C_g(l). \tag{3.15}$$

The choice of l was arbitrary, therefore $C_g^+ \leqq \inf_l C_g(l)$.

4. Examples of channels for which the group code capacity is smaller than Shannon's channel capacity. The definition of the group code capacity C_g given in paragraph 1(C) depends on the way in which we define the field structures in X and Y. The field structure serves only as a tool to deal with purely probabilistically described channels. We introduce of course field structures which optimize C_g.

(4.1) Let C_g^* be the group code capacity corresponding to an optimal choice of the field structures.

The following example demonstrates the need for the introduction of C_g^*.

EXAMPLE 1. Choose $X = Y = \{1, 2, 3\}$; $GF(3) = \{0', 1', 2'\}$. Consider the d.m.c. given by

$$w = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Let 1 serve as $0'$, 2 as $1'$ and 3 as $2'$ in the input space X and in the output space Y. If u is a code word then also $-u$ is a code word. $P(y_n \mid u) = P(y_n \mid -u)$ for $y_n \in Y_n$ implies that there exists no (n, N, λ) group code for $N > 1$, $\lambda < \frac{1}{2}$. C_g is zero.

However, if 2 serves as $0'$, 1 serves as $1'$ and 3 serves as $2'$ in X, and 1 serves as $0'$, 2 as $1'$ and 3 as $2'$ in Y, then $\{u_1, u_2, u_3\} = \{(1', 2'), (2', 1'), (0', 0')\}$ forms a group and by choosing the coset leaders $(1', 1')$, $(2', 2')$ we get a $(n, N, \lambda) = (2, 3, 0)$ group code. By taking direct sums of the group $\{u_1, u_2, u_3\}$ we can achieve $C_g^* \geqq \frac{1}{2} \log 3 > 0$.

(4.2) Let τ be a permutation of $\{1, \cdots, n\}$, where τi is the image of i under τ.

(4.3) τ induces a mapping τ^* of X_n onto X_n and a mapping τ^{**} of Y_n onto Y_n:

$$\tau^* x_n = \tau^*(x^1, \cdots, x^n) = (x^{\tau 1}, \cdots, x^{\tau n})$$

$$\tau^{**} y_n = \tau^{**}(y^1, \cdots, y^n) = (y^{\tau 1}, \cdots, y^{\tau n}) \qquad (x_n \in X_n, y_n \in Y_n).$$

It follows from definitions (1.1) and (1.3) that

$$P(y_n \mid x_n \mid s) = P(\tau^{**} y_n \mid \tau^* x_n \mid s) \quad \text{for} \quad x_n \in X_n, y_n \in Y_n, s \in S \tag{4.4}$$

and

$$\bar{P}(y_n \mid x_n) = \bar{P}(\tau^{**} y_n \mid \tau^* x_n) \qquad \text{for} \quad x_n \in X_n, y_n \in Y_n.$$

Thus the invariance property (4.4) holds for d.m.c., compound channels and averaged channels. (It also holds for channels with arbitrarily varying channel probability functions and therefore holds for some of the most reasonable channels.)

(4.5) A linear (n, N) code $\{(u_i, A_i) \mid i = 1, \cdots, N\}$ is called a *systematic code*, if there exists a matrix $P = (p_{ij})$ $i = 1, \cdots, k$; $j = k+1, \cdots, n$ with coefficients in $GF(p^s)$ such that

$$\{u_1, \cdots, u_N\} = \{u \mid u = (a^1, \cdots, a^k, b^{k+1}, \cdots, b^n), \tag{4.6}$$

where $a^i \in GF(p^s) = X$ for $i = 1, \cdots, k$ and $b^j = \sum_{i=1}^{k} a^i p_{ij}$ for $j = k+1, \cdots, n\}$. The first k components are called information digits, and the last $n-k$ components are called check digits.

(4.7) Given a linear code $\{(u_i, A_i) \mid i = 1, \cdots, N\}$, then there exists a permutation τ, such that $\{(\tau^* u_i, \tau^{**} A_i) \mid i = 1, \cdots, N\}$ is a systematic code and $P(A_i \mid u_i \mid s) = P(\tau^{**} A_i \mid \tau^* u_i \mid s)(s \in S;\ i = 1, \cdots, N)$, $\bar{P}(A_i \mid u_i) = \bar{P}(\tau^{**} \mid \tau^* u_i)(i = 1, \cdots, N)$.

(Using property (4.4) the proof of Theorem 3.3 in [13] carries over verbatim and gives (4.7).)

Because of (4.7) we can limit ourselves to the study of systematic codes. We give now an example of a d.m.c. with $C_g^* < C$.

EXAMPLE 2. Define w by $w(i \mid i) = 1$ for $i = 1, \cdots, p-1$, $w(j \mid p) = 1/p$ for $j = 1, \cdots, p$. Obviously $C > \log(p-1)$. Suppose $C_g^* = C$. Fix field structures in X, Y for which $C_g = C_g^*$. Then for arbitrarily small λ there exists a $n_0(\lambda)$ such that for $n \geqq n_0(\lambda)$ there exists a systematic code with maximal error less than λ and length greater than $\exp(n \log(p-1))$. This code has at least $k \geqq n \log(p-1)/\log p$ information digits. The code word $(p, \cdots, p, b^{k+1}, \cdots, b^n)$ needs a decoding set with at least $\exp(H(1/p, \cdots, 1/p)k - K(\lambda)k^{\frac{1}{2}})$ elements. Therefore we need $\exp(H(1/p, \cdots, 1/p)k - K(\lambda)k^{\frac{1}{2}})$ cosets. But $|A_1 \cup \cdots \cup A_n| \geqq \exp(n \log(p-1) + H(1/p, \cdots, 1/p)k - K(\lambda)n^{\frac{1}{2}}) \geqq \exp(n \log(p-1) + n \log(p-1) - K(\lambda)n^{\frac{1}{2}}) > \exp(n \log p)$, for n large and $p \geqq 3$, contradicts $|A_1 \cup \cdots \cup A_n| = |X_n| = \exp(n \log p)$.

For the symmetric channels considered by Elias, Dobrushin and Drygas

(4.8) $C_g = C_g^* = \bar{C}_g = \bar{C}_g^* = C$ always holds.

It seems very likely that $\bar{C}_g$ (respectively $\bar{C}_g^*$) can be greater than C_g (respectively C_g^*) for general d.m.c. We give now an example of a d.m.c. for which $\bar{C}_g^* < C$.

EXAMPLE 3. Define w by

$$
\begin{aligned}
w(i \mid i) &= 1 \qquad \text{for} \quad i = 1, \cdots, [p/2]; \\
w(j \mid i) &= 2/p \qquad \text{for} \quad i = [p/2]+1, \cdots, p, \\
&\qquad\qquad\quad\ \ j = [p/2]+1, \cdots, p-1; \\
w(j \mid i) &= 3/p \qquad \text{for} \quad i = [p/2]+1, \cdots, p, \\
&\qquad\qquad\quad\ \ j = p.
\end{aligned}
$$

Suppose $\bar{C}_g^* = C$. Fix field structures in X, Y for which $\bar{C}_g = C$. Obviously $C > \log p/2$.

For any λ there exists a $n_0(\lambda)$ such that for $n \geqq n_0(\lambda)$ there exists a systematic code with average error less than λ and length greater than $\exp(n \log p/2)$. This code has at least $k \geqq n \log \frac{1}{2} p/\log p$ information digits. Divide the set $\{1, \cdots, p\}$ into the sets $R = \{1, \cdots, [p/2]\}$ and $Q = \{[p/2]+1, \cdots, p\}$. It follows from Chebyshev's inequality that the proportion of code words, which have more than

(4.9) $((p - [p/2])/p - \varepsilon)k$ information digits in Q, tends to 1 as n tends to infinity. Let λ be less than $\frac{1}{2}$, then we can find a subcode of $\{(u_i, A_i) \mid i = 1, \cdots, N\}$ which has a length $N^* \geqq N/2$ and a *maximal* error less than 2λ. Denote this $(n, N^*, \lambda/2)$-code by $\{(u_i^*, A_i^*) \mid i = 1, \cdots, N^*\}$. By choosing n large enough we can guarantee that more than half of the u_i^*'s are of type (4.9).

A code word of type (4.9) requires more than $\exp(H(2/p, 2/p, \cdots, 3/p)\cdot(p-[p/2])/p)k - K(\lambda)k^{\frac{1}{2}}$ elements in the decoding set. Therefore

$$|A_1 \cup \cdots \cup A_N| \geqq \frac{1}{4}\exp\left(n \log \frac{p}{2}\right)\exp\left(H\left(\frac{2}{p}, \cdots, \frac{3}{p}\right)\frac{p-[p/2]}{p}\, k - K(\lambda)k^{\frac{1}{2}}\right).$$

Using $k \geqq (\log(p/2)/\log p)n$, $H(2/p, \cdots, 3/p) \sim \log p/2$ and $(p-[p/2])/p \sim \frac{1}{2}$ for p large, we get

$$|A_1 \cup, \cdots, \cup A_N| \geqq \frac{1}{4}\exp\left(n \log \frac{p}{2} + n\frac{1}{2}\frac{\log p/2 \cdot \log p/2}{\log p} - K(\lambda)n^{\frac{1}{2}}\right).$$

But

$$\log \frac{p}{2} + \frac{1}{2}\frac{\log p/2 \cdot \log p/2}{\log p} = \log \frac{p}{2} + \frac{1}{2}\log \frac{p}{2}\left(1 - \frac{\log 2}{\log p}\right) > \log p,$$

for p large enough, in contradiction to $|X_n| = \exp(n \log p)$.

We give now an example of a discrete channel with $C > \bar{C}_g^* = 0$.

EXAMPLE 4. Let $a = 3$,

$$w(\cdot\,|\,\cdot\,|\,1) = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

$$w(\cdot\,|\,\cdot\,|\,2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

$$w(\cdot\,|\,\cdot\,|\,3) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

and define the averaged channel

$$P(\cdot\,|\,\cdot) = \tfrac{1}{3}P(\cdot\,|\,\cdot\,|\,1) + \tfrac{1}{3}P(\cdot\,|\,\cdot\,|\,2) + \tfrac{1}{3}P(\cdot\,|\,\cdot\,|\,3).$$

All fields we can define in X are necessarily additive cyclic groups with 3 elements. If we choose for the alphabet $X = \{a_1, a_2, a_3\}$ the group structure $\{0, 1, 2\}$ or $\{0, 2, 1\}$, then the d.m.c. $P(\cdot\,|\,\cdot\,|\,3)$ has group capacity ${}_3\bar{C}_g = 0$. This can be seen as follows: let u be a code word which is not the zero vector, then $-u$ also belongs to the code. But $P(y_n|u|3) = P(y_n|-u|3)$ implies that for $\bar{\lambda} < \frac{1}{3}$ $N_g(n, \bar{\lambda}) = 1$ and therefore ${}_3\bar{C}_g = 0$. If we choose group structure $\{1, 0, 2\}$ or $\{2, 0, 1\}$ then ${}_2\bar{C}_g = 0$ and if

we choose group structure $\{1, 2, 0\}$ or $\{2, 1, 0\}$ then ${}_1\bar{C}_g = 0$. Applying Theorem 3 we get $\bar{C}_g{}^* = 0$, but according to Lemma 6

$$C = \max_\pi \inf_{s=1,2,3} R(\pi, w(\cdot\,|\,\cdot\,|\,s)).$$

Choose $\pi^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, then

$$R(\pi^*, w(\cdot\,|\,\cdot\,|\,1)) = R(\pi^*, w(\cdot\,|\,\cdot\,|\,2)) = R(\pi^*, w(\cdot\,|\,\cdot\,|\,3))$$

by symmetry. An easy computation shows that $R(\pi^*, w(\cdot\,|\,\cdot\,|\,s)) > 0$, and therefore that $C > 0$. In this context it may be of interest that the following theorem holds:

THEOREM 4. *Let* $a = 2$.

(a) *For the compound channel* $\{P(\cdot\,|\,\cdot\,|\,s)\,|\,s \in S\}$, *where every* $P(\cdot\,|\,\cdot\,|\,s)$ *is a binary symmetric* d.m.c., *we have* $\bar{C}_g{}^* = \bar{C}_g = C = \max_\pi \inf_{s\in S} R(\pi, w(\cdot\,|\,\cdot\,|\,s))$.

(b) *Let* q_s *be a* p.d. *on* S *with* $q_s > 0$ *for all* $s \in S$, *then for the "binary symmetric" averaged channel*

$$P(\cdot\,|\,\cdot) = \textstyle\sum_{s\in S} q_s P(\cdot\,|\,\cdot\,|\,s), \qquad \bar{C}_g = \bar{C}_g{}^* = C.$$

The proof will appear in a forthcoming paper by J. Gemma and the author.

5. An upper bound for the capacity of linear and pseudo linear codes. A pseudo group code $\{(u_i, A_i)\,|\,i = 1, \cdots, N\}$ for which $\{u_i\,|\,u = 1, \cdots, N\}$ is a vector space shall be called a *pseudo linear code*. For these codes we define capacities C_{l_p}, $C_{l_p}^*$, $\bar{C}_{l_p}$, $\bar{C}_{l_p}^*$ as usual.

THEOREM 5. *Let* $a = p^s$. *For a* d.m.c. *given by* w *we have* $C_l{}^* \leqq C_{l_p}^* \leqq R(\pi^*, w)$, *where* $\pi^* = (1/p^s, \cdots, 1/p^s)$.

PROOF. According to (4.7) we can restrict ourselves to systematic codes. Let $\{u_1, \cdots, u_N\}$ be any systematic code as described under (4.5). Write this code as a matrix so that the ith code word u_i equals the ith row vector.

$$U = \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix} = \begin{pmatrix} a_1{}^1, \cdots, a_1{}^k, b_1{}^{k+1}, \cdots, b_1{}^n \\ \vdots \\ a_N{}^1, \cdots, a_N{}^k, b_N{}^{k+1}, \cdots, b_N{}^n \end{pmatrix}.$$

In the first k-columns every element of $GF(p^s)$ appears equally often. Consider now column $j(k+1 \leqq j \leqq n)$. Write the matrix P of (4.5) as $(P_{k+1}, \cdots, P_n)$, where P_j has k components. The set of all elements of the jth column in U equals $\{\alpha\,|\,\alpha = (A, P_j)$, where A is any vector with k components$\}$. ((A, P_j) is the scalar product of A and P_j.) The number of times $\alpha \in GF(p^s)$ occurs in the jth column of U is equal to the number of solutions of

$$(A, P_j) = \alpha. \tag{5.1}$$

For $\alpha \neq 0$ (5.1) has a solution if and only if P_j is not the zero vector. In this case we find all solutions of (5.1) by finding one solution and adding to it all solutions of

$$(A, P_j) = 0. \tag{5.2}$$

Therefore, if P_j is not the zero vector then all elements of $GF(p^s)$ appear equally often in the jth column of U, and if P_j is the zero vector, then only 0 occurs in the jth column of U. Define now ${\pi_i}^t$ as in (2.18), then either ${\pi_0}^t = 1$ or ${\pi_i}^t = 1/p^s$ for all i.

$$R(\pi^t, w) = 0, \qquad \text{if} \quad {\pi_0}^t = 1.$$

Application of Lemma 7 to our code completes the proof.

$$R(\pi^*, w) \quad \text{is in general smaller than} \quad C = \max_\pi R(\pi, w).$$

One can easily construct examples where $R(\pi^*, w) < C$ *even for* $a = 2$. This case was not covered by Example 2.

THEOREM 6. *Let* $a = p^s$. *For a* d.m.c. *given by* w *we have*

$$\bar{C}_l{}^* \leqq \bar{C}_{l_p}^* \leqq R(\pi^*, w), \qquad \textit{where} \quad \pi^* = (1/p^s, \cdots, 1/p^s).^2$$

PROOF. We deal now with average errors, and in this case Lemma 7 does not apply directly. We make use of Lemma 2 in [4] which states

(5.3) *Let* $\{(u_i, A_i) \,|\, i = 1, \cdots, N\}$ *be a code for* $P(\cdot\,|\,\cdot)$ *with average error* $\bar{\lambda}$; *then there exists a subcode of length* $N' = N\varepsilon\,|\,(\bar{\lambda}+\varepsilon)$ *with maximal error* $\bar{\lambda}+\varepsilon$, *if* $\bar{\lambda}+\varepsilon < 1$.

Denote the subcode by

$$\{(u_{i_v}, A_{i_v}) \,|\, v = 1, \cdots, N'\}. \tag{5.4}$$

(5.5) We have omitted only $(1-\varepsilon/(\bar{\lambda}+\varepsilon))\,N$ sequences from our original set $\{u_1, \cdots, u_N\}$.

Define the matrix

$$U' = \begin{pmatrix} u_{i_1} \\ \vdots \\ u_{i_{N'}} \end{pmatrix}.$$

Assigning probability vectors ${\pi'}^t$ to the subcode (5.4) according to (2.18), we get

$$\left|{\pi_i'}^t - {\pi_i}^t\right| = \left|\frac{N_i^t}{N} - \frac{{N_i'}^t}{N'}\right|,$$

where N_i^t = number of times i occurs in the tth column of U and ${N_i'}^t$ = number of times i occurs in the tth column of U'. It follows from (5.5) that $\left|N_i^t/N - {N_i'}^t/N'\right|$ is less than a function $g(\varepsilon, \bar{\lambda})$, where $\lim_{\bar{\lambda}\to 0} g(\varepsilon, \bar{\lambda}) = 0$.

Using the continuity of $R(\pi, w)$ in π and applying Lemma 7 to U' we get $\bar{C}_{l_p}^* \leqq R(\pi^*, w) + f(\bar{\lambda}, \varepsilon)$, where $f(\bar{\lambda}, \varepsilon)$ tends to zero as $\bar{\lambda}$ tends to 0. Therefore we have $\bar{C}_{l_p}^* \leqq R(\pi^*, w)$.

---

[2] R. G. Gallager has pointed out to me that he independently found this result (unpublished). His proof uses Fano's Lemma instead of Lemma 7. Compare also his remarks on page 208 in [11].

Using Theorem 6 we can easily construct examples of channels with any alphabet-length $a = p^s$ for which even pseudo-linear codes with average error do not achieve the channel capacity.

We show now that the bound given on C_l^* in Theorem 5 is not sharp.

EXAMPLE 5. Consider a d.m.c. given by

$$w = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Assign to the alphabet $X = (a_1, a_2, a_3)$ a field structure for which $a_3 \neq 0$, because otherwise $C_l = 0$. W.l.o.g. choose $a_3 = 2$. For $\pi^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ we get $R(\pi^*, w) = \log_2 3 - \frac{2}{3}$. A systematic code with rate $R(\pi^*, w)$ has to have $k \sim n(\log 3 - \frac{2}{3})/\log 3$ information digits.

(5.6) There are 2^k code words which have only 0 and 1 as components in the information digits.

We call two sequences "separable" if there exists at least one component in which one sequence takes the value 2 and the other not. Only for those sequences can we find disjoint sets A_1 and A_2 such that $P(A_1 \mid u_1) > 0$ and $P(A_2 \mid u_2) > 0$. We call a set of sequences "separable" if any two elements are separable. The set of code words has to be separable. The subset described under (5.6) is separable only if its check sequences are separable. But the maximal cardinality of a separable set of check sequences is 2^{n-k}.

$$k \sim \frac{\log 3 - \frac{2}{3}}{\log 3} n > \tfrac{1}{2}n,$$

therefore $2^k > 2^{n-k}$. The subset is not separable, therefore the set of code words is also not separable. Consequently $C_l^* < R(\pi^*, w)$.

It might be pointed out that Example 5 illustrates the fact that the capacity of a "subchannel" can be greater than the capacity of the channel. The capacity of

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

is less than $\log_2 3 - \frac{2}{3}$ whereas the capacity of $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is $\log_2 2 > \log_2 3 - \frac{2}{3}$. It is to be pointed out that this is for the capacity C_l^* as defined. This is eliminated if C_l^* is defined as the maximum over all mappings of $GF(q^t)$ into the set of channel symbols for $q^t \leqq p^s$. Theorem 5 could be easily modified to be consistent with this alternate definition.

6. Unsolved problems.

(1) Can one prove the existence of $\bar{C}_g$?
(2) Can $\bar{C}_g$ be larger than C_g for discrete memoryless channels?
(3) In which cases does a group coding strong converse hold?
(4) What is the relationship between the different capacities?
(5) Is $\bar{C}_{l_p}^* = R(\pi^*, w)$?[3]

This problem is entirely different from the problem treated in [11], page 208. The most important and difficult problem, however, is

(6) Can one find explicit formulas for any one of the capacities defined?

REFERENCES

[1] AHLSWEDE, R. (1967). Certain results in coding theory for compound channels. *Colloq. Information Theory, Debrecen, Hungary*. 35–60.
[2] AHLSWEDE, R. (1968). The weak capacity of averaged channels. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **11** 61–73.
[3] 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.* **3** 1027–1033.
[4] AHLSWEDE, R. and WOLFOWITZ, J. (1969). The structure of capacity functions for compound channels. *Lecture Notes, Probability and Information Theory* **89** Springer, Berlin.
[5] AUGUSTIN, U. (1966). Gedächtnisfreie Kanäle für diskrete Zeit. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **6** 10–61.
[6] BERLEKAMP, E. R. (1968). *Algebraic Coding Theory*. McGraw-Hill, New York.
[7] DOBRUSHIN, R. L. (1963). Asymptotic optimality of group and systematic codes for some channels. *Theor. Probability Appl.* **8** 47–59.
[8] DRYGAS, H. (1965). Verschlüsselungstheorie für symmetrische Kanäle. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **4** 121–143.
[9] ELIAS, P. (1955). Coding for noisy channels. IRE Convention Record **4** 37–46.
[10] ELIAS, P. (1956). Coding for two noisy channels. *Information Theory* (Colin Cherry, ed.). Academic Press, New York.
[11] GALLAGER, R. G. (1968). *Information Theory and Reliable Communication*. Wiley, New York.
[12] HAMMING, R. W. (1950). Error detecting and error correcting codes. *Bell System Tech. J.* **29** 147–160.
[13] PETERSON, W. W. (1961). *Error Correcting Codes*. MIT Press.
[14] SLEPIAN, D. (1956). A class of binary signalling alphabets. *Bell System Tech. J.* **35** 203–234.
[15] WOLFOWITZ, J. (1957). The coding of messages subject to chance errors. *Illinois J. Math.* **1** 591–606.
[16] WOLFOWITZ, J. (1960). Simultaneous channels. *Arch. Rat. Mech. Anal.* **4** 371–386.
[17] WOLFOWITZ, J. (1963). Channels without capacity. *Information and Control* **6** 49–54.
[18] WOLFOWITZ, J. (1961). *Coding Theorems of Information Theory* (1st ed.). Springer–Verlag, Berlin.

[3] This question has meanwhile been decided in the affirmative by I. Gemma and the author. The proof will be given in the forthcoming paper "Bounds on algebraic code capacities for noisy channels" (to appear in *Information and Control*).