Reprinted from INFORMATION AND CONTROL
All Rights Reserved by Academic Press, New York and London

Vol. 19, No. 2, September 1971
Printed in Belgium

Bounds on Algebraic Code Capacities for Noisy Channels. I

R. AHLSWEDE[1] AND J. GEMMA

Ohio State University, Columbus, Ohio, and
Battelle Memorial Institute, Columbus Laboratories

This paper continues the study of algebraic code capacities, which were introduced by Ahlswede (1971). He states an upper bound for the rates of codes which have the property that the code words form a linear space and the decoding procedure is arbitrary. It was asked (problem 5) whether this upper bound is actually the capacity if we deal with average errors. We answer this question in the affirmative for binary discrete memoryless channels. For non-binary discrete memoryless channels we obtain slightly weaker result: If we allow those codes which have as code words a coset of a group which is a linear space, then the upper bound is again the capacity. An example shows that the result is not true for maximal error.

In paragraph 3 we prove that the linear code capacity for compound channels with invariant transition probabilities equals the capacity for compound channels as given by Wolfowitz (1960).

I. BASIC DEFINITIONS AND AUXILIARY RESULTS

1. *Channels, Probabilistic Codes, and Errors*

Let $X = \{1,..., a\}$ denote the input alphabet and let $Y = \{1,..., a\}$ denote the output alphabet. Let $X_n = \prod_1^n X$ denote the set of sequences $x_n = (x^1,..., x^n)$ where $x^t \in X$, $t = 1,..., n$ and let $Y_n = \prod_{i=1}^n Y$ denote the corresponding set of sequences $y_n = (y^1,..., y^n)$, $y^t \in Y$, $t = 1,..., n$. We call x_n an input word of length n and y_n an output word of length n. We define a channel probability function (c.p.f.) to be an $a \times a$ stochastic matrix $w(\cdot \mid \cdot)$.

(1.1.1) A discrete memoryless channel (d.m.c.) is a system $\mathscr{P} = \{P_n(\cdot \mid \cdot) \mid n = 1, 2,...\}$, where

$$P_n(y_n \mid x_n) = \prod_{t=1}^{n} w(y^t \mid x^t)$$

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

for all $x_n \in X_n$, $y_n \in Y_n$, $n = 1, 2, \ldots$. We also refer to $\mathscr{P}$ as the d.m.c. given by $w(\cdot \mid \cdot)$.

Let S be an arbitrary set, and let $\{w(\cdot \mid \cdot \mid s) \mid s \in S\}$ be a collection of c.p.f.'s.

(1.1.2) We call $\mathscr{P}(S) = \{P_n(\cdot \mid \cdot \mid s) \mid s \in S, \ n = 1, 2, \ldots\}$ a compound channel if we are interested in the simultaneous behavior of these channels: each n-sequence x_n is transmitted according to $P_n(\cdot \mid \cdot \mid s)$ for some $s \in S$ and the channel may vary arbitrarily from one n sequence to the next. Given a probability distribution (p.d.) q on S,

(1.1.3) We define an averaged discrete channel $\mathscr{P}^* = \{P_n{}^*(\cdot \mid \cdot) \mid n = 1, 2, \ldots\}$ by

$$P_n{}^*(y_n \mid x_n) = \sum_{s \in S} q_s P_n(y_n \mid x_n \mid s)$$

for all $x_n \in X_n$, $y_n \in Y_n$, $n = 1, 2, \ldots$.

(1.1.4) A code (n, N) is a system $\{(u_1, A_1), \ldots, (u_N, A_N)\}$, where $u_i \in X_n$, $A_i \subset Y_n$, $i = 1, \ldots, N$, and $A_i \cap A_j = \varnothing$, $i \neq j$.

(1.1.5) A code (n, N) is called a code (n, N, λ) (with maximal error)

(i) for the d.m.c. $\mathscr{P}$ if

$$P_n(A_i \mid u_i) \geqslant 1 - \lambda, \qquad i = 1, \ldots, N,$$

(ii) for the compound channel $\mathscr{P}(S)$ if

$$P_n(A_i \mid u_i \mid s) \geqslant 1 - \lambda \qquad \text{for all} \quad s \in S, \quad i = 1, \ldots, N,$$

(iii) for the averaged channel $\mathscr{P}^*$ if

$$P_n{}^*(A_i \mid u_i) \geqslant 1 - \lambda, \quad i = 1, \ldots, N.$$

(1.1.6) A code (n, N) is called a code $(n, N, \bar{\lambda})$ (with average error)

(i) for the d.m.c. $\mathscr{P}$ if

$$\frac{1}{N} \sum_{i=1}^{N} P_n(A_i \mid u_i) \geqslant 1 - \bar{\lambda}.$$

(ii) for the compound channel $\mathscr{P}(S)$ if

$$\frac{1}{N} \sum_{i=1}^{N} P(A_i \mid u_i \mid s) \geqslant 1 - \bar{\lambda} \qquad \text{for all} \quad s \in S.$$

(iii) for the average channel $\mathscr{P}^*$ if

$$\frac{1}{N} \sum_{i=1}^{N} P^*(A_i \mid u_i) \geqslant 1 - \bar{\lambda}.$$

(1.1.7) Let $N(n, \lambda)$ denote the maximal length of an (n, N, λ) code, and let $N(n, \bar{\lambda})$ denote the maximal length of an $(n, N, \bar{\lambda})$ code.

(1.1.8) We say the (n, N) code $\{(u_i, A_i) \mid i = 1,..., N\}$ is a maximum likelihood code with respect to $\mathscr{P}$ if

$$\{y_n \mid y_n \in Y_n \text{ and } P(y_n \mid u_i) > \max_{j \neq i} P(y_n \mid u_j)\} \subset A_i$$

$$\subset \{y_n \mid y_n \in Y_n \text{ and } P(y_n \mid u_i) \geqslant \max_{j \neq i} P(y_n \mid u_j)\} \qquad \text{for} \quad i = 1,..., N.$$

(1.1.9) The (n, N) code (1.1.8) is called a *strict* maximum likelihood code (s.m.l.c.) if

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

We define the entropy of a probability vector $\pi = (\pi_1, ..., \pi_a)$ to be

$$(1.1.10) \qquad H(\pi) = - \sum_{i=1}^{a} \pi_i \log \pi_i .$$

(All logarithms in this paper are to the base 2.) Denote the rate for the probability π on X and c.p.f. $w(\cdot \mid \cdot \mid s)$ by

(1.1.11) $R(\pi, s) = H(\pi'(s)) - \sum_i \pi_i H(w(\cdot \mid i \mid s))$ where $\pi'(s)$ is the probability vector on Y given by

$$\pi'(s)_j = \sum_i \pi_i w(j \mid i \mid s) \qquad \text{for} \quad j = 1,..., a.$$

2. *Shannon's Channel Capacity*

(1.2.1) A number $C > 0$ is called (Shannon's or weak) capacity of a channel if

(i) for any $\delta > 0$ and λ $(0 < \lambda < 1)$ there exists a code $(n, 2^{n(C-\delta)}, \lambda)$ all sufficiently large n, and if

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

Part (i) is called the coding theorem and part (ii) is called the weak converse of the coding theorem.

(1.2.2) C is called the strong capacity if (i) holds and (ii) is replaced by:

(ii′) 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.

Note that (i), (ii′) imply (i), (ii). (ii′) is called the strong converse of the coding theorem. Analogous definitions can be given for $(n, N, \bar{\lambda})$ codes. (i), (ii) are equivalent to

$$(1.2.3) \qquad \inf_{\lambda>0} \underline{\lim}_{n} \frac{1}{n} \log N(n, \lambda) = \inf_{\lambda>0} \overline{\lim}_{n} \frac{1}{n} \log N(n, \lambda) = C.$$

(i), (ii′) are equivalent to

$$(1.2.4) \quad \underline{\lim}_{n} \frac{1}{n} \log N(n, \lambda) = \overline{\lim}_{n} \frac{1}{n} \log N(n, \lambda) = C \qquad \text{for all } \lambda,\ 0 < \lambda < 1.$$

3. *Algebraic Codes*

We assume that X (resp. Y) is a Galois field with $a = p^s$ elements (which we denote by GF(a)) where p is a prime and s is an integer, and we identify X_n (resp. Y_n) with the vector space of dimension n over GF(a). That is, for $x_n = (x^1,..., x^n) \in X_n$, $\bar{x}_n = (\bar{x}^1,..., \bar{x}^n) \in X_n$, and $\lambda \in \text{GF}(a)$, we have

$$x_n + \bar{x}_n = (x_1 + \bar{x}_1 ,..., x^n + \bar{x}^n)$$

and

$$\lambda x_n = (\lambda x^1,..., \lambda x^n),$$

where the sums $x^i + \bar{x}^i$ and the products λx^i are defined in the sense of GF(a).

(1.3.1) A code (n, N) is a pseudo-group code if $\{u_1 ,..., u_N\}$ is a subgroup of X_n and the A_i's are arbitrary. Let φ denote the canonical isomorphism between X_n and Y_n : for $x_n \in X_n$, $\varphi x_n = y_n$, where $y^t = x^t$, $t = 1,..., n$.

(1.3.2) A pseudo-group code is called a group code if there exists a set of representatives $\{l_1 ,..., l_L\}$ of the cosets of $\{\varphi u_1 ,..., \varphi u_N\}$ for which $A_i = \{l_1 + \varphi u_i ,..., l_L + \varphi u_i\}$, $i = 1,..., N$.

(1.3.3) A group code is called as linear code if $\{u_1 ,..., u_N\}$ is a subspace of X_n. Note that if $a = p$, group codes and linear codes coincide.

(1.3.4) An (n, N) code is a pseudo-linear code if it is a pseudo-group code and $\{u_1 ,..., u_N\}$ is a subspace of X_n.

(1.3.5) $\{(u_1 , A_1),..., (u_N , A_N)\}$ is a pseudo-shifted group code (n, N) if there exists a pseudo-group code with code words $\{\bar{u}_1 ,..., \bar{u}_N\}$ and an $x_n \in X_n$

such that $u_i = \bar{u}_i + x_n$, $i = 1, \ldots, N$, and the decoding sets A_i, $i = 1, \ldots, N$, are arbitary.

(1.3.6) $\{(u_1, A_1), \ldots, (u_N, A_N)\}$ is a shifted group code (n, N) if there exists a group $\{\bar{u}_1, \ldots, \bar{u}_N\}$ and an $x_n \in X_n$ such that $u_i = \bar{u}_i + x_n$, $i = 1, \ldots, N$, and if there exists a system of representatives $\{l_1, \ldots, l_L\}$ of the cosets of $\{\varphi\bar{u}_1, \ldots, \varphi\bar{u}_N\}$ such that $A_i = \{l_1 + \varphi\bar{u}_i, \ldots, l_L + \varphi u_i\}$, $i = 1, \ldots, N$.

(1.3.7) An (n, N) code is a shifted linear code if it is a shifted group code obtained from a group code for which $\{u_1, \ldots, u_N\}$ is a subspace of X_n.

(1.3.8) We say that an (n, N) code $\{(u_1, A_1), \ldots, (u_N, A_N)\}$ is a pseudo-shifted linear code if there exists a pseudo-linear code with code words $\{\bar{u}_1, \ldots, \bar{u}_N\}$ and an $x_n \in X_n$ such that $u_1 = \bar{u}_1 + x_n, \ldots, u_N = \bar{u}_N + x_n$.

4. *Algebraic Code Capacities*

We introduce the concept of algebraic code capacities. We say that

$$C_u^+ = \inf_{\lambda > 0} \overline{\lim_n} \frac{1}{n} \log N_u(n, \lambda) \tag{1.4.1}$$

and

$$C_u^- = \inf_{\lambda > 0} \underline{\lim_n} \frac{1}{n} \log N_u(n, \lambda) \tag{1.4.2}$$

are the upper and lower capacities respectively of a particular algebraic code concept, where $N_u(u, \lambda)$ denotes the maximal length of (n, N, λ) codes of this type. We make this more precise in the following table:

(1.4.3)

Algebraic code concept	Maximal length of (n, N, λ) codes	Upper capacity (1.4.1)	Lower capacity (1.4.2)
Group code	$N_g(n, \lambda)$	C_g^+	C_g^-
Pseudo group code	$N_p(n, \lambda)$	C_p^+	C_p^-
Linear code	$N_l(n, \lambda)$	C_l^+	C_l^-
Pseudo linear code	$N_{l_p}(n, \lambda)$	$C_{l_p}^+$	$C_{l_p}^-$
Shifted group code	$N_{sg}(n, \lambda)$	C_{sg}^+	C_{sg}^-
Pseudo shifted group code	$N_{sp}(n, \lambda)$	C_{sp}^+	C_{sp}^-
Shifted linear code	$N_{sl}(n, \lambda)$	C_{sl}^+	C_{sl}^-
Pseudo shifted linear code	$N_{sl_p}(n, \lambda)$	$C_{sl_p}^+$	$C_{sl_p}^-$

If $C_g^+ = C_g^-$, we talk about the group code capacity (for maximal error) C_g. Analogously, we define C_p, C_l, C_{lp}, C_{sp}, C_{sl}, and C_{sl_p}. If we talk about average error, we talk about the quantities $N_u(n, \bar{\lambda})$, $\bar{C}_u^+$, $\bar{C}_u^-$ corresponding to the quantities in (1.4.3). If $\bar{C}_g^+ = \bar{C}_g^-$, we talk about the group code capacity (for average error) $\bar{C}_g$. Analogously, we define $\bar{C}_p$, $\bar{C}_l$, $\bar{C}_{l_p}$, $\bar{C}_{sg}$, $\bar{C}_{sp}$, $\bar{C}_{sl}$, and $\bar{C}_{sl_p}$.

5. *Auxiliary Results*

Before proceeding to the main results, we first state some known theorems and introduce some concepts to which we will refer later. We state now a fundamental result in coding theory, to which we will refer many times throughout this paper. We precede the statement of the theorem by some definitions and notation. Let $U \times V$ be a finite or countable probability space with elements (u, v) and a probability distribution $Q(u, v)$. Let $P(v \mid u)$ be the conditional probability on V given u, and let $Q'(u)$, $Q''(v)$ be the marginals of Q on U and V, respectively. Let $u_1^*, \ldots, u_N^*$ be pairwise independent random variables taking values in U according to $P(u_i^* = u) = Q'(u)$. For each set of values of $u_1^*, \ldots, u_N^*$ we define N disjoint subsets $A_1^*, \ldots, A_N^*$ of V by $A_i^* = \{v \mid P(v \mid u_i^*) > \max_{j \neq i} P(v \mid u_j^*)\}$ and N random variables $\mathscr{E}_1, \ldots, \mathscr{E}_N$ by

$$\mathscr{E}_i = P(A_i^{*c} \mid u_i^*) = \sum_{v \in A_i^{*c}} P(v \mid u_i^*).$$

(1.5.1) Let

$$I(u, v) = \log \frac{Q(u, v)}{Q'(u)Q''(v)}.$$

Then we have the following theorem due to Shannon (1957):

THEOREM 1.5.1 (Random Coding Theorem). *Let $\alpha > 1$ be arbitrary. We have*

$$\frac{1}{N} \sum_{i=1}^{N} E\mathscr{E}_i \leqslant \frac{1}{\alpha} + Q\{(u, v) \mid I(u, v) \leqslant \log \alpha N\}.$$

Another result we will make use of is due to Fano (in Wolfowitz (1964)). Using the same notation as for the random coding theorem, let $\{(u_1, A_1), \ldots, (u_N, A_N)\}$ be an $(N, \bar{\lambda})$ code. Without loss of generality, we may assume that $A_1 \cup \cdots \cup A_N = V$. Let Q_0' be the distribution on U defined by $Q_0'(u_i) = 1/N$, $i = 1, \ldots, N$. Let $P(\cdot \mid \cdot)$ denote a channel and let $Q(u, v) = Q_0'(u)P(v \mid u)$.

(1.5.2) Let $R(Q') = E(I(u, v))$. Then we have

THEOREM 1.5.2. *For any channel, a code* $(N, \bar{\lambda})$ *satisfies*

$$(1 - \bar{\lambda}) \log N \leqslant R(Q_0') + 1.$$

We now introduce the concept of a systematic code and state a lemma relating systematic codes and linear codes. We first introduce some necessary preliminary notation. Let σ be a permutation of $\{1,..., n\}$, and let σi denote the image of i under σ. Then σ induces a mapping σ^* of X_n onto X_n and a mapping σ^{**} of Y_n onto Y_n given by

$$(1.5.3) \qquad \sigma^* x_n = \sigma^*(x^1,..., x^n) = (x^{\sigma 1},..., x^{\sigma n})$$

and

$$\sigma^{**} y_n = \sigma^{**}(y^1,..., y^n) = (y^{\sigma 1},..., y^{\sigma n})$$

for $x_n \in X_n$, $y_n \in Y_n$. It follows from (1.1.1) that

$$(1.5.4) \quad P_n(y_n \mid x_n) = P_n(\sigma^{**} y_n \mid \sigma^* x_n) \qquad \text{for} \quad x_n \in X_n, \quad y_n \in Y_n.$$

(1.5.5) An (n, N) linear code is called a systematic code if there exists a matrix $P = (p_{ij})$, $i = 1,..., k$, $j = k + 1,..., n$, with coefficients in $\mathrm{GF}(p^s)$ such that $\{u_1,..., u_N\} = \{u \mid u = (a^1,..., a^k, b^{k+1},..., b^n)$, where $a^i \in \mathrm{GF}(p^s)$, $i = 1,..., k$, and

$$b^j = \sum_{i=1}^{k} a^i p_{ij} \text{ for } j = k + 1,..., n\}.$$

The first k components are called the information digits and the last $(n - k)$ components are called the check digits. We have the following lemma (Ahlswede, 1971 and Peterson, 1961):

LEMMA 1.5.1. *If* $\{(u_1, A_1),..., (u_N, A_N)\}$ *is a linear code, then there exists a permutation* σ *such that* $\{(\sigma^* u_1, \sigma^{**} A_1),..., (\sigma^* u_N, \sigma^{**} A_N)\}$ *is a systematic code and* $P(A_i \mid u_i) = P(\sigma^{**} A_i \mid \sigma^* u_i)$.

II. ALGEBRAIC CODE CAPACITIES FOR SEVERAL CHANNELS

The results in this chapter extend theorems of Elias (1955) and Dobrushin (1963), and partially resolve a problem raised by Ahlswede (1971, unsolved Problem 5). We define a channel with invariant transition probabilities (c.i.t.p.) as a d.m.c. given by a matrix $w(\cdot \mid \cdot)$ which satisfies $w(j \mid i) =$

$w(j+k \mid i+k)$ for all $i, j, k \in \mathrm{GF}(a)$, where the sums $j+k$, $i+k$ are defined in the sense of $\mathrm{GF}(a)$. We state these results in our terminology.

THEOREM 2.1 (Elias). *Let* $X = Y = \mathrm{GF}(2)$. *Let* $\mathscr{P}$ *be a binary symmetric* d.m.c. *Then* $C = C_g = \bar{C}_g$.

THEOREM 2.2 (Dobrushin). *Let* $X = Y = \mathrm{GF}(a)$, *where* $a = p^k$. *For a* c.i.t.p.

(i) $C = C_l = \bar{C}_l$ *and therefore also*

(ii) $C = C_{sl} = \bar{C}_{sl}$.

The definition of $\bar{C}_{l_p}$ given in (I.4) depends on the way in which we define the field structures in X and Y. Let $\bar{C}_{l_p}^*$ be the value of $\bar{C}_{l_p}$ corresponding to an optimal choice of field structures. Let π^* be the uniform distribution on X. Ahlswede (1971) asked whether $\bar{C}_{l_p}^* = R(\pi^*, w)$.

1. *The Pseudo Linear Code Capacity for the Binary Discrete Memoryless Channel*

The proof of the theorem which follows makes use of an idea of Elias (1955).

THEOREM 2.1.1. *Let* $\mathscr{P}$ *be a* d.m.c. *with* $X = Y = \mathrm{GF}(2)$. *Then*

$$N_p(n, \bar{\lambda}) > 2^{nR(\pi^*, w) - K_{\bar{\lambda}}\sqrt{n}}$$

where $K_{\bar{\lambda}}$ *is a constant depending on* $\bar{\lambda}$ *but not on* $\mathscr{P}$ *or* n *and* $\pi^* = (\frac{1}{2}, \frac{1}{2})$.

Proof. It is sufficient to prove the result for large n. Suppose we have a pseudo-linear code with 2^k elements. Let G denote the set of code words. Then we can find a set of generators, $u_1, \dots, u_k$, such that $u \in G$ implies

$$u = \sum_{i=1}^{k} \alpha_i u_i, \qquad \alpha_i \in \mathrm{GF}(2), \qquad i = 1, \dots, k.$$

The idea of the proof is as follows: select generators at random, form a pseudo-linear code, and apply the random coding theorem to prove the result.

Fix n and k, $k < n$. Independently select sequences

$$u_1 = (u_1^1, u_1^2, \dots, u_1^n), \dots, u_k = (u_k^1, \dots, u_k^n),$$

where $u_i^t \in X$, $i = 1, \dots, k$, $t = 1, \dots, n$, choosing the components of each sequence independently with probability $\frac{1}{2}$ that either element in X will be chosen.

Form a code with 2^k words, u_j, $j = 0, \ldots, 2^k - 1$, by taking all possible linear combinations of $u_1, \ldots, u_k$. Note that the u_j are not necessarily distinct, since $u_1, \ldots, u_k$ may not have been linearly independent.

(2.1.1) $u_j = \alpha_{j1}u_1 + \alpha_{j2}u_2 + \cdots + \alpha_{jk}u_k$, where $\alpha_{jt} \in X$,

$t = 1, \ldots, k$. Let u_0 be the zero vector corresponding to $\alpha_{0t} = 0$, $t = 1, \ldots, k$. Then we have

(2.1.2) $u_j^t = \alpha_{j1}u_1^t + \cdots + \alpha_{jk}u_k^t$, $t = 0, \ldots, 2^k - 1$. Since the totality of components of $u_1, \ldots, u_k$ are independent, the u_j^t, $t = 1, \ldots, n$ are independent. Hence the components of each word are chosen independently. Moreover,

(2.1.3) $P\{u_j^t = 0\} = P\{u_j^t = 1\} = \frac{1}{2}$, $t = 1, \ldots, n, j \neq 0$. If $j \neq 0$, there is at least one coefficient $\alpha_{jr} \neq 0$. If $\sum_{l=1}^{k} \alpha_{jl}u_l^t = 0$, obtain a new sequence (u_l^{*t}), where

$$u_l^{*t} = u_l^t \qquad l \neq r,$$

$$u_r^{*t} = u_r^t + 1.$$

For this sequence,

$$\sum_{l=1}^{k} \alpha_{jl}u_l^{*t} = 1.$$

Hence, there are at least as many sequences $(u_1^t, \ldots, u_k^t)$ producing $u_j^t = 1$ as there are $u_j^t = 0$. By symmetry, we obtain that there are as many sequences producing $u_j^t = 1$ as there are producing $u_j^t = 0$. Thus, the components of the words u_j, $j \neq 0$, are independent and equidistributed.

We now show that the words u_j, u_m, $j, m \neq 0, j \neq m$ are independent. We show

(2.1.4) $P(u_j^t = x, u_m^t = x') = \frac{1}{4}$ for $x, x' \in X$, $t = 1, \ldots, n$. It then follows that

(2.1.5) $P\{u_j^t = x, u_m^t = x'\} = P\{u_j^t = x\}P\{u_m^t = x'\}$. Then u_j^t, u_m^t are independent for each t which implies that u_j, u_m are independent. To prove (2.1.4) we note that since $j \neq m$, we have $\alpha_{jr} \neq \alpha_{mr}$ for some r, say $\alpha_{jr} = 1$ and $\alpha_{mr} = 0$. Then given $(u_1^t, \ldots, u_k^t)$ such that

(2.1.6) $$\sum_{l=1}^{k} \alpha_{jl}u_l^t = 0, \qquad \sum_{l=1}^{k} \alpha_{ml}u_l^t = 0,$$

define $u_l^{*t} = u_l^t \; l \neq r$, and $u_r^{*t} = u_r^t + 1$. Then

$$\sum_{l=1}^{k} \alpha_{jl} u_l^{*t} = 1, \qquad \sum_{l=1}^{k} \alpha_{ml} u_l^{*t} = 0. \tag{2.1.7}$$

Hence, we can now see that there is a one-to-one correspondence between sequences which produce $u_j^t = u_m^t = 0$ and ones that produce $u_j^t = 1$, $u_m^t = 0$. Hence,

(2.1.8) $P\{u_j^t = 0, u_m^t = 0\} = P\{u_j^t = 1, u_m^t = 0\}$. But

(2.1.9) $\frac{1}{2} = P\{u_m^t = 0\} = P\{u_j^t = 0, u_m^t = 0\} + P\{u_j^t = 1, u_m^t = 0\}$ which implies

(2.1.10) $P\{u_j^t = 0, u_m^t = 0\} = P\{u_j^t = 1, u_m^t = 0\} = \frac{1}{4}$ and the other relations in (2.1.4) easily follow.

(2.1.11) Let $s = 2^k - 1$. Since $u_1, \ldots, u_s$ are pairwise independent, we may apply Theorem 1.5.1. Let K_0 be an ensemble of codes of the above type. Then for arbitrary $\beta > 1$, we have that

$$\frac{1}{s} \sum_{i=1}^{s} E\mathscr{E}_i \leqslant \frac{1}{\beta} + Q_n\{(u, v) \mid I_n(u, v) \leqslant \log \beta s\}, \tag{2.1.12}$$

where $Q_n\{(u, v)\} = Q_n'(u) P_n(v \mid u)$ and Q_n' is the source distribution on X_n given by

$$Q_n'(u) = \prod_{t=1}^{n} Q'(u^t) = \frac{1}{2^n} \qquad \text{for all} \quad u \in X_n \,. \tag{2.1.13}$$

If (2.1.12) is less or equal $\bar{\lambda}' < \frac{1}{2}$, then there is a code $(n, s, \bar{\lambda}')$.

(2.1.14) Let d, d' be such that

$$2^{-d\sqrt{n}} + Q_n\{I_n \leqslant nR(\pi^*, w) - d' \sqrt{n}\} \leqslant \bar{\lambda}',$$

where $I_n = I_n(u, v) = \sum_{t=1}^{n} I^t(u^t, v^t)$. This is possible by Chebyshev's inequality and the fact that $E_{Q'}(I^t) = R(\pi^*, w)$ and hence $E_{Q_n'}(I_n) = nR(\pi^*, w)$.

(2.1.15) Choose $\beta = 2^{d\sqrt{n}}$ and k such that

$$s = 2^k - 1 \leqslant 2^{nR(\pi^*, w) - (d + d')\sqrt{n}} \leqslant 2^{k+1} - 1.$$

Note that $\beta s \leqslant 2^{nR(\pi^*, w) - d'\sqrt{n}}$ implies

$$Q_n\{I_n \leqslant \log \beta s\} \leqslant Q_n\{I_n \leqslant nR(\pi^*, w) - d' \sqrt{n}\}. \tag{2.1.16}$$

Since $2^{nR(\pi^*,w)-(d+d')\sqrt{n}} < 2^{k+1}$ there is a constant $K_{\bar{\lambda}'}$ such that

$$2^{nR(\pi^*,w)-K_{\bar{\lambda}'}\sqrt{n}} < 2^k. \tag{2.1.17}$$

Hence we have a code $(n, s, \bar{\lambda}')$, $\{(u_1{}^0, A_1{}^0),..., (u_s{}^0, A_s{}^0)\}$. Add the code word u_0 to this code with decoding set $\varnothing$. Then the average error for this new code is given by

$$\frac{1}{s+1} + \frac{1}{s+1} \sum_{i=1}^{s} P(A_i^{0c} \mid u_i{}^0) \leqslant \frac{1}{s+1} + \frac{s}{s+1} \bar{\lambda}'.$$

Since s tends toward ∞ as n tends toward ∞, there exists n^* such that $n \geqslant n^*$ implies

$$\frac{1}{s} + \frac{s}{s+1} \bar{\lambda}' < 2\bar{\lambda}' = \bar{\lambda}.$$

We now have a code $(n, 2^k, \bar{\lambda})$, where

$$2^k \geq 2^{nR(\pi^*,w)-K_{\bar{\lambda}}\sqrt{n}}.$$

Now replacing the preceding decoding scheme by a maximum likelihood decoding scheme for u_0, $u_1{}^0,..., u_s{}^0$ we can only improve on the error probability.

Now if u_0, $u_1{}^0,..., u_s{}^0$ are not distinct, the set of generators was not linearly independent. If r is the maximal number of linearly independent code words in the set of generators, then 2^r is the number of distinct code words. Replace the dependent code words by $k - r$ words so that we achieve a set of k linearly independent generators, and hence 2^k distinct code words. Decode maximum likelihood and obtain a $(n, 2^k, \bar{\lambda})$ code and the theorem is proved.

For completeness, we include the weak converse of Theorem 2.1.1. This result was proved by Ahlswede (1971). The proof given here is different. Let $\{(u_1, A_1),..., (u_N, A_N)\}$ be a linear or pseudo linear code. Let

$$\pi_i{}^t = \frac{|\{u_j{}^t \mid u_j{}^t = i, j = 1,..., N\}|}{N} \quad \text{for} \quad i = 1,..., a.$$

We first state a result of Ahlswede (1971):

Lemma 2.1.1. *Let $X = Y = \mathrm{GF}(a)$, where $a = p^s$. Then either*

$$\pi_0{}^t = 1 \quad \text{or} \quad \pi_i{}^t = \frac{1}{a}, \quad i = 1,..., a.$$

Theorem 2.1.2. *Let $X = Y = \mathrm{GF}(a)$, where $a = p^s$. Let $\mathscr{P}$ be a* d.m.c. *given by w. Then $(1 - \bar{\lambda})N_{l_p}(n, \bar{\lambda}) \leqslant nR(\pi^*, w) + 1$, where*

$$\pi^* = \left(\frac{1}{p^s}, \dots, \frac{1}{p^s}\right).$$

Proof. According to Lemma 1.5.1, we can restrict ourselves to systematic codes. Let $\{u_1, \dots, u_N\}$ be the code words of any systematic code as described under (1.5.5). We have from Theorem 1.5.2 that $(1 - \bar{\lambda}) \log N \leqslant R(Q'_{0n}) + 1$, where $Q'_{0n}(u_i) = 1/N$, $i = 1, \dots, N$. Let Q'^t_0 be the marginal distribution of Q'_{0n} on $\{u_i{}^t \mid i = 1, \dots, N\}$. We have

(2.1.18) $R(Q'_{0n}) \leqslant \sum_{t=1}^{n} R(Q'^t_0)$. By Lemma 2.1.1, we have that $Q'^t_0(0) = 1$ or $Q'^t_0(i) = 1/p^s$, $i = 1, \dots, p^s$. Since $R(\pi^t, w) = 0$ if $\pi_0{}^t = 1$, we have from (2.1.18) $R(Q'_{0n}) \leqslant nR(\pi^*, w)$ where $\pi^* = (1/p^s, \dots, 1/p^s)$.

From Theorem 2.1.1 and Theorem 2.1.2 we have:

Theorem 2.1.3. *Let $\mathscr{P}$ be a* d.m.c. *given by w. Let $X = Y = \mathrm{GF}(2)$. Then*

$$\bar{C}_p = R(\pi^*, w), \qquad \textit{where} \quad \pi^* = (\tfrac{1}{2}, \tfrac{1}{2}).$$

Theorem 2.1.3 is a solution of unsolved Problem 5 of Ahlswede (1971) in the case $X = Y = \mathrm{GF}(2)$.

Theorem 2.1 is a corollary of Theorem 2.1.3. To see this, we first prove the following

Lemma 2.1.2. *Let $\mathscr{P}$ be a* c.i.t.p. *Let $G = \{u_1, \dots, u_N\}$ be a subgroup of X_n. Then there exists a maximum likelihood decoding scheme* (1.1.8) *for G which is also a group decoding scheme* (1.3.2), *and maximal error equals average error for this decoding scheme.*

Proof. Let u_1 be the zero code word. Let $V_1 = \{y_n \mid y_n \in Y_n$ and $P(y_n \mid u_1) = \max_{j \neq 1} P(y_n \mid u_j)\}$. Let V_1' be a set obtained from V_1 by choosing exactly one representative of each coset of G which has elements in V_1. Then we define

$$A_1 = \{y_n \mid y_n \in Y_n \text{ and } P(y_n \mid u_1) > \max_{j \neq 1} P(y_n \mid u_j)\} \cup V_1'.$$

(2.1.19) Let $A_i = \{v + u_i \mid v \in A_1\}$, $i = 2, \dots, N$. Then $\{(u_1, A_1), \dots, (u_N, A_N)\}$ is a maximum likelihood code since $A_i \cap A_j = \varnothing$, $i \neq j$, and

$$\begin{aligned} P(v + u_i \mid u_i) = P(v \mid u_1) &\geqslant \max_{j \neq 1} P(v \mid u_j) \\ &= \max_{j \neq i} P(v + u_i \mid u_j) \end{aligned}$$

for all $v \in A_1$, $i = 2,..., N$. But (2.1.19) is a group decoding scheme if we let $\{l_1,..., l_L\} = A_1$. From (2.1.19) it follows that $P(A_i \mid u_i) = P(A_j \mid u_j)$, $i, j = 1,..., N$. Hence maximal error and average error coincide.

Since a binary symmetric channel is a c.i.t.p., Theorem 2.1.3 and Lemma 2.1.2 imply Theorem 2.1.

2. *The Pseudo Shifted Linear Code Capacity for the Discrete Memoryless Channel*

The following theorem is proved with the help of the methods of Dobrushin (1963).

THEOREM 2.2.1. *Let $\mathscr{P}$ be a* d.m.c. *given by w, and let $X = Y =$* GF(a), *where $a = p^s$. Then*

$$N_{sl_p}(n, \bar{\lambda}) > 2^{nR(\pi^*, w) - K_{\bar{\lambda}}\sqrt{n}}$$

where $\pi^* = (1/p^s,..., 1/p^s)$ and $K_{\bar{\lambda}}$ is a constant depending on $\bar{\lambda}$ and a but not on w or n.

Proof. It suffices to prove the result for large n. Suppose we have a shifted linear code with a^k words. Then we may find a set of generators $\tilde{u}_1,..., \tilde{u}_k$ such that for any code word u, $u = \sum_{i=1}^{k} \alpha_i \tilde{u}_i + e$, where $\alpha_i \in \mathrm{GF}(a)$, $i = 1,..., k$ and $e \in X_n$.

To choose $\tilde{u}_1,..., \tilde{u}_k$, independently select k sequences of length n, each component in X, choosing the components of each sequence independently with probability $1/a$ that any element in X will be chosen.

Form a code with a^k words by taking all possible linear combinations of $\tilde{u}_1,..., \tilde{u}_k$, and adding a word $e = (e^1,..., e^n) \in X_n$, chosen in such a way that each component $e^t \in X^t$ is equally likely and the choice is independent of $\tilde{u}_1{}^t,..., \tilde{u}_k{}^t$, and e^t is independent of e^s, $t \neq s$.

(2.2.1) $u_j = \alpha_{j1}\tilde{u}_1 + \cdots + \alpha_{jk}\tilde{u}_k + e, j = 1,..., a^k$, where $\alpha_{jl} \in X$, $l = 1,..., k$. Let $\bar{u}_j{}^t$ denote the quantity

$$\bar{u}_j{}^t = \alpha_{j1}\tilde{u}_1{}^t + \cdots + \alpha_{jk}\tilde{u}_k{}^t$$

and let $\bar{u}_i$ denote the quantity

$$\bar{u}_i = (\bar{u}_i{}^1,..., \bar{u}_i{}^n).$$

Consider the expression $P(u_i{}^t = j \mid \bar{u}_i{}^t = l)$. Since $u_i{}^t = \bar{u}_i{}^t + e^t$, we have

$$P(u_i{}^t = j \mid \bar{u}_i{}^t = l) = P(e^t = j - l) = \frac{1}{a}.$$

But then we have

$$P(u_i^t = j) = \sum_{l=1}^{a} P(u_i^t = j \mid \bar{u}_i^t = l)\, P(\bar{u}_i^t = l) = \frac{1}{a}\,.$$

Similarly,

$$P(u_i = j) = \frac{1}{a^n}\,.$$

Next we show that u_i and u_j are independent for $i \neq j$. We have

$$P\{\tilde{u}_1 = x_1 ,..., \tilde{u}_k = x_k \mid u_i = x\} = \frac{P\{\tilde{u}_1 = x_1 ,..., \tilde{u}_k = x_k \,, u_i = x\}}{P\{u_i = x\}}$$

$$= a^n P\{\tilde{u}_1 = x_1 ,..., \tilde{u}_k = x_n \,, e = x - \alpha_{i1}x_1 - \cdots - \alpha_{ik}x_k\}$$

$$= a^n \frac{1}{(a^n)^{k+1}} = \frac{1}{(a^n)^k}\,.$$

Hence the variables $\tilde{u}_1 ,..., \tilde{u}_k$ remain independent and identically distributed under the assumption $u_i = x$. Note next that the mapping $b \rightarrow \alpha b$, α, $b \in \mathrm{GF}(a)$, $\alpha \neq 0$, is a one-to-one mapping of GF(a) onto GF(a). Thus if a random variable ξ is uniformly distributed on GF(a), so is $\alpha\xi$. It follows now from the definition (2.2.1) of u_i , u_j that

$$u_j = u_i + \tilde{\alpha}_1\tilde{u}_1 + \cdots + \tilde{\alpha}_k\tilde{u}_k\,.$$

Assume $\tilde{\alpha}_k \neq 0$. Then

$$P\{u_j = \bar{x} \mid u_i = x, \tilde{u}_1 = x_1 ,..., \tilde{u}_{k-1} = x_{k-1}\}$$
$$= P\{\tilde{\alpha}_k\tilde{u}_k = \bar{x} - x - \tilde{\alpha}_1 x_1 - \cdots - \tilde{\alpha}_{k-1}x_{k-1} \mid u_i = x\} = \frac{1}{a^n}$$

and hence, by the formula for total probability,

$$P\{u_j = \bar{x} \mid u_i = x\} = \frac{1}{a^n}\,.$$

Hence u_i and u_j are independent.

Now since $u_1 ,..., u_{a^k}$ are pairwise independent, we may apply the random coding theorem, Theorem 1.5.1, to the ensemble of codes of the above type and obtain that the expected value of the average errors of the codes in the ensemble is less or equal

$$\frac{1}{\beta} + Q_n\{(u, v) \mid I(u, v) \leqslant \log \beta a^k\}. \tag{2.2.2}$$

If (2.2.2) is less or equal $\bar{\lambda}$, there is a code (n, a^k, $\bar{\lambda}$).

Let d, d' be such that

$$2^{-d\sqrt{n}} + Q_n\{I_n \leqslant nR(\pi^*, w) - d'\sqrt{n}\} \leqslant \bar{\lambda}. \tag{2.2.3}$$

This is possible, as in (2.1.14), by Chebyshev's inequality and $E_{Q_n'}(I_n) = nR(\pi^*, w)$. Similar to (2.1.15), choose $\beta = 2^{d\sqrt{n}}$ and k such that

$$a^k \leqslant 2^{nR(\pi^*, w)-(d+d')\sqrt{n}} \leqslant a^{k+1}. \tag{2.2.4}$$

Then $\beta a^k \leqslant 2^{nR(\pi^*, w)-d'\sqrt{n}}$ implies

$$Q_n\{I_n \leqslant \log \beta a^k\} \leqslant Q_n\{I_n \leqslant nR(\pi^*, w) - d'\sqrt{n}\}. \tag{2.2.5}$$

From (2.2.4), since $a^{k+1} \geqslant 2^{nR(\pi^*, w)-(d+d')\sqrt{n}}$, there is a constant $K_{\bar{\lambda}}$ such that

$$2^{nR(\pi^*, w)-K_{\bar{\lambda}}\sqrt{n}} < a^k. \tag{2.2.6}$$

From (2.2.5) and (2.2.6), it follows that

$$N_{l_p}(n, \bar{\lambda}) > 2^{nR(\pi^*, w)-K_{\bar{\lambda}}\sqrt{n}}.$$

We now prove the weak converse of Theorem 2.2.1.

THEOREM 2.2.2. *Let $\mathscr{P}$ be a* d.m.c. *given by w*, with $X = Y = \mathrm{GF}(a)$ *where $a = p^s$. Then*

$$(1 - \bar{\lambda})N_{sl_p}(n, \bar{\lambda}) \leqslant nR(\pi^*, w) + 1,$$

where $\pi^* = (1/p^s, \ldots, 1/p^s)$.

Proof. Let G' denote the set of code words in an $(n, N, \bar{\lambda})$ pseudo shifted linear code. From Lemma 1.5.1 we may assume that G' was obtained from a systematic code, whose code words we denote by G, by the addition of some $x_n \in X_n$, that is, $G' = \{u + x_n \mid u \in G\}$. Since for G, either $\pi_0{}^t = 1$ or $\pi_i{}^t = 1/p^s$, $i = 1, \ldots, p^s$, it easily follows that either there is a j such that $\pi_j{}^t = 1$ or $\pi_i{}^t = 1/p^s$, $i = 1, \ldots, p^s$, for G'. The theorem now follows from Theorem 1.5.2 and (2.1.18) in Theorem 2.1.2.

From Theorem 2.2.1 and Theorem 2.2.2 we have

THEOREM 2.2.3. *Let $\mathscr{P}$ be a* d.m.c. *given by w with $X = Y = \mathrm{GF}(a)$, where $a = p^s$. Then*

$$\bar{C}_{sl_p} = R(\pi^*, w)$$

where $\pi^* = (1/p^s, \ldots, 1/p^s)$.

Part (ii), Theorem 2.2, follows from Theorem 2.2.3 and Lemma 2.1.2.

3. *The Linear Code Capacity for Compound Channels with Invariant Transition Probabilities*

We now extend Theorem 2.2 to the case of a compound channel with finitely many channels. In order to prove this result, we first prove two lemmas.

(2.3.1) A symmetric channel is an $a \times a$ stochastic matrix whose rows are permutations of each other and whose columns are permutations of each other. The following lemma is due to Dobrushin (1963).

LEMMA 2.3.1. *A* c.i.t.p. *is a symmetric channel.*

Proof. The permutation $j_l \rightarrow j_{l'}$ defined by $j_{l'} - j_l = i_{k'} - i_k$ has the property that

$$w(j_{l'} \mid i_{k'}) = w(j_l \mid i_k).$$

To see this, note that

$$\begin{aligned} w(j_l \mid i_k) &= w(j_l - j_{l'} \mid i_k - j_{l'}) \\ &= w(i_k - i_{k'} \mid i_k - j_{l'}) \\ &= w(j_{l'} \mid i_{k'}). \end{aligned}$$

Similarly, the permutation $i_l \rightarrow i_{l'}$ defined by $i_{l'} - i_l = j_{k'} - j_k$ has the property that

$$w(j_k \mid i_l) = w(j_{k'} \mid i_{l'}).$$

Hence a c.i.t.p. is symmetric.

(2.3.2) Let $\mathscr{P}(S)$ be a compound channel given by $\{w(\cdot \mid \cdot \mid s) \mid s \in S\}$, $\mid S \mid < \infty$, $X = Y = \mathrm{GF}(a)$, where each $w(\cdot \mid \cdot \mid s)$ is a c.i.t.p. Let $\mathscr{P}^*$ denote an averaged channel over $\mathscr{P}(S)$ with distribution $\{q_s \mid s \in S\}$. Let $E_n(K)$ denote the expected value of the average errors of codes in a system K of (n, N) pseudo-linear codes, where the probability distribution on K is determined by $\mathscr{P}^*$ and by a source distribution Q_n' on X_n. Let $Q_n = Q_n' \times P_n^*$.

(2.3.3) Let K^* be the system of (n, N) pseudo shifted linear codes obtained from K, and let the distribution on K^* be such that for each of the a^n pseudo shifted linear codes corresponding to the pseudo-linear code G, say G_i^*, $i = 1, \ldots, a^n$ we have

$$Q_n^*(G_i^*) = \frac{1}{a^n} Q_n(G).$$

LEMMA 2.3.2. *Under the conditions* (2.3.2) *and* (2.3.3) *we have*

$$E_n(K) = E_n(K^*).$$

Proof. Let $G^* = \{(u_1^*, A_1^*), \ldots, (u_N^*, A_N^*)\} \in K^*$. Let G be a code in K such that the set of code words G_u^* of G^* is obtained from the set of code words G_u of G by adding an element $x_n \in X_n$, that is,

$$G_u^* = \{u + x_n \mid u \in G_u\}.$$

Then for each $P_n(\cdot \mid \cdot \mid s)$ we have

$$P_n(v + x_n \mid u + x_n \mid s) = P_n(v \mid u \mid s)$$

so that

$$P^*(v + x_n \mid u + x_n) = P^*(v \mid u) \qquad \text{for all} \quad v \in Y_n, \quad u \in G_u. \tag{2.3.4}$$

Then from (2.3.4)

$$\begin{aligned}
\bar{\lambda}(G^*) &= \frac{1}{N} \sum_{i=1}^{N} P^*(A_i^{*c} \mid u_i^*) \\
&= \frac{1}{N} \sum_{i=1}^{N} P^*(A_i + x_n \mid u_i + x_n) \\
&= \frac{1}{N} \sum_{i=1}^{N} P^*(A_i{}^c \mid u_i) \\
&= \ (G).
\end{aligned}$$

It now follows from (2.3.3) that $E_n(K) = E_n(K^*)$.

THEOREM 2.3.1. *Let* $\mathscr{P}(S)$ *be a compound channel given by* $\{w(\cdot \mid \cdot \mid s) \mid s \in S\}$, $|S| < \infty$, where $w(\cdot \mid \cdot \mid s)$ *is a* c.i.t.p. *for all* $s \in S$. *Let* $X = Y =$ GF(a) *where* $a = p^s$. *Then*

$$C_l = \bar{C}_l = C = \max_{\pi} \inf_{s} R(\pi, s)$$

Proof. First we note that for each $w(\cdot \mid \cdot \mid s)$, $\pi^* = (1/a, \ldots, 1/a)$ is the maximizing input distribution since each channel is symmetric by Lemma 2.3.1. Hence $C = \min_{s \in S} C_s$. Also, since each channel is symmetric, the output distribution corresponding to π^* is again π^*. We proceed now precisely as in Theorem 2.2.1 to choose a^k code words which form an (n, a^k) shifted linear code, are equidistributed, and are pairwise independent.

We consider the ensemble K of (n, a^k) linear codes on the channel

$$P_n{}^*(\cdot \mid \cdot) = \sum_{s\in S} q_s P_n(\cdot \mid \cdot \mid s), \tag{2.3.5}$$

where $\{q_s \mid s \in S\}$ is a p.d. on S and $q_s > 0$, $s \in S$. Let K^* be the ensemble of shifted linear codes corresponding to K. By Lemma 2.3.2, we have $E_n(K) = E_n(K^*)$. Hence by Theorem 1.5.1 applied to K^*, we obtain

$$E_n(K) = E_n(K^*) \leqslant \frac{1}{\beta} + Q_n\{(u, v) \mid I_n(u, v) \leqslant \log \beta a^k\} \tag{2.3.6}$$

We have

$$\begin{aligned} I_n(u, v) &= \log \frac{Q_n(u, v)}{Q_n'(u)Q_n''(v)} \\ &= \log \frac{P_n{}^*(v \mid u)}{Q_n''(v)} \\ &= \log \frac{\sum_{s\in S} q_s P_n(v| \, u \, |s)}{1/a^n} \\ &= \log a^n + \log \sum_{s\in S} q_s P_n(v| \, u \, |s) \\ &\geqslant \log a^n + \sum_{s\in S} q_s \log P_n(v| \, u \, |s) \\ &= \log \prod_{s\in S} \left[\frac{P_n(v| \, u \, |s)}{Q_n''(v)}\right]^{q_s} \\ &= \sum_{s\in S} q_s \log \frac{P_n(v| \, u \, |s)}{Q_n''(v)} \\ &= \sum_{s\in S} q_s I_n(u, v; s). \end{aligned} \tag{2.3.7}$$

From (2.3.7) we have that $I_n(u, v) \leqslant \log \beta \alpha^k$ implies

$$\sum_{s\in S} q_s I_n(u, v; s) \leqslant \log \beta a^k$$

and thus

$$\begin{aligned} & Q_n\{(u, v) \mid I_n(u, v) \leqslant \log \beta a^k)\} \\ & \qquad \leqslant Q_n \left\{(u, v) \,\middle|\, \sum_{s\in S} q_s I_n(u, v; s) \leqslant \log \beta a^k\right\} \\ & \qquad \leqslant \sum_{s\in S} Q_n\{(u, v) \mid I_n(u, v; s) \leqslant \log \beta a^k\}. \end{aligned} \tag{2.3.8}$$

Let d, d' be such that

$$2^{-d\sqrt{n}} + \sum_{s\in S} Q_n(I_n(\cdot, \cdot; s) \leqslant nC - d'\sqrt{n}) \leqslant \lambda. \tag{2.3.9}$$

That we can obtain (2.3.9) follows from Chebyshev's inequality and from

$$E_{Q_n'}(I_n(\cdot, \cdot; s)) = nC_s, \qquad s \in S.$$

Choose $\beta = 2^{-d\sqrt{n}}$ and k such that

$$a^k \leqslant 2^{nC-(d+d')\sqrt{n}} \leqslant a^{k+1}. \tag{2.3.10}$$

Then $\beta a^k \leqslant 2^{nC-d'\sqrt{n}}$ implies

$$\begin{aligned} Q_n\{I_n(u, v; s) &\leqslant \log \beta a^k\} \\ &\leqslant Q_n\{I_n(u, v; s) \leqslant nC - d'\sqrt{n}\} \end{aligned} \tag{2.3.11}$$

for all $s \in S$. From (2.3.6), (2.3.8), (2.3.9), and (2.3.11), we obtain $E_n(K) \leqslant \lambda$, and from (2.3.10) we obtain that there exists a constant $K_{\bar{\lambda}}$ such that

$$a^k > 2^{nC-K_{\bar{\lambda}}\sqrt{n}}.$$

Hence there is a code $(n, 2^{nC-K_{\bar{\lambda}}\sqrt{n}}, \bar{\lambda})$. This proves $\bar{C}_{l_p} \geqslant C$. Since, clearly, $\bar{C}_{l_p} \leqslant C$ we have $\bar{C}_{l_p} = C$. By Lemma 2.1.2, we have $C_l = \bar{C}_l = \bar{C}_{l_p} = C$, and the theorem is proved. The following results were announced by Ahlswede (1971):

COROLLARY 2.3.1. *For the compound channel $\mathscr{P}(S)$ given by $\{w(\cdot \mid \cdot \mid s) \mid s \in S\}$, $\mid S \mid < \infty$, where each $w(\cdot \mid \cdot \mid s)$ is binary symmetric* d.m.c.,

$$C_g = \bar{C}_g = C = \max_{\pi} \inf_{s\in S} R(\pi, s).$$

Proof. This follows directly from Theorem 2.3.1.

COROLLARY 2.3.2. *Let $\mathscr{P}(S)$ be a compound channel given by $\{w(\cdot \mid \cdot \mid s) \mid s \in S\}$, $\mid S \mid < \infty$, where each $w(\cdot \mid \cdot \mid s)$ is a binary symmetric* d.m.c. *Let $\{q_s \mid s \in S\}$ be a* p.d. *on S with $q_s > 0$ for all $s \in S$. Then, for the "binary symmetric averaged channel" $\mathscr{P}^*$ given by*

$$P_n^*(\cdot \mid \cdot) = \sum_{s\in S} q_s P_n(\cdot \mid \cdot \mid s), \qquad n = 1, 2, \ldots,$$

we have

$$C_g = \bar{C}_g = C.$$

Proof. It was shown by Ahlswede (1971) that if C_g denotes the capacity of $\mathscr{P}(S)$, then C_g is also the capacity of $\mathscr{P}^*$. The result now follows from Corollary 2.3.1.

4. *Examples*

Note that Theorems 2.1.3, 2.2.3, and 2.3.1 hold regardless of the field structure defined on X. It was shown (Ahlswede, 1971) that for some channels the capacity of the channel depends on the field structure defined on X and Y. Let $C_u{}^*$, where u is one of the previously defined subscripts for C, denote the u-capacity corresponding to the optimal choice of field structures.

Theorem 2.2.3 implies $\bar{C}_{sl_p} = R(\pi^*, w)$. Example 5 of Ahlswede (1971) shows that $C_{l_p}^* < R(\pi^*, w)$. In Example 1 to follow, we show that $C_{sl_p}^* < R(\pi^*, w)$. It is perhaps surprising that, in this case, maximal error and average error lead to different capacities.

EXAMPLE 1. $C_{sl_p}^* < R(\pi^*, w)$. Let

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

Fix any field structure on $X = (a_1, a_2, a_3)$. For $\pi^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$,

$$R(\pi^*, w) = \log_2 3 - \tfrac{2}{3}.$$

A systematic code with rate $\sim R(\pi^*, w)$ must have $k \sim (1 - \frac{2}{3}\log 3)n$ information digits.

Let G' denote an arbitrary (n, N) pseudo shifted linear code obtained from a systematic code, and let $G_u{}'$ denote the set of code words of G'.

(2.4.1) There is a subset of $G_u{}'$ of cardinality 2^k with only 0's and 1's in the information digits.

To see this, note that $G_u{}' = \{u + x_n \mid u \in G_u\}$ for the systematic code G and some $x_n \in X_n$. Let G_{uk}, G'_{uk} denote the sets of elements obtained by taking the first k components of elements in G_u, $G_u{}'$, respectively, and let x_k denote that element of X_k whose components agree with the first k components of x_n. Since $G_{uk} = X_k$, it follows that $G'_{uk} = X_k$.

Assume $a_3 \neq 0$, and without loss of generality, assume $a_3 = 2$.

(2.4.2) Only for sequences u_1, u_2 such that there exists at least one component in which one sequence takes the value 2 and the other sequence takes one of the other values are there sets A_1, A_2 such that both $P(A_1 \mid u_1) > 0$ $P(A_2 \mid u_2) > 0$. Any two elements of G_u' must have property (2.4.2).

The subset (2.4.1) has this property only if its check sequences have this property. But the maximal cardinality of such a set of check sequences is 2^{n-k}.

We have $k \sim (1 - \frac{2}{3} \log 3)n > n/2$ so that $2^k > 2^{n-k}$. Hence the subset (2.4.1) does not have property (2.4.2) and so neither does G_u'. Hence $C^*_{sl_p} < R(\pi^*, w)$.

For completeness, we include the following example, which was proved by Ahlswede (1971) for an averaged channel.

EXAMPLE 2. There exists a compound channel such that

$$= \bar{C}_p{}^* < \inf_s R(\pi^*, s) \leqslant C.$$

Let

$$w_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

$$w_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

$$w_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

By symmetry, $R(\pi^*, w_1) = \inf_s R(\pi^*, s)$ and

$$\inf_s R(\pi^*, s) = \log 3 - \tfrac{2}{3} > 0.$$

Let G be an $(n, N, \bar{\lambda})$ pseudo-group code, where $\bar{\lambda} < \frac{1}{3}$. Since the code words of G form a group, $u \in G_u$ implies $-u \in G_u$. Let F be any field structure on $\{a_1, a_2, a_3\}$. Note that any changes in field structure are simply permutations of the input alphabets of w_1, w_2, w_3. (The field structure on the output alphabet is irrelevant in this case.) Let w^* be the channel

$$w^* = \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Since except for the zero element, $u \neq -u$ and

$$P_{*n}(y \mid u) = P_{*n}(y \mid -u) \qquad \text{for all} \quad y \in Y_n\,, \quad \text{for any } A \text{ and } \tilde{A},$$
$$P_{*n}(A \mid u) + P_{*n}(\tilde{A} \mid -u) \leqslant 1.$$

Hence

$$\frac{1}{N} \sum_{i=1}^{N} P_{*n}(A_i \mid u_i)$$
$$= \frac{1}{N} \left[P_{*n}(A \mid 0) + \sum_{\substack{\text{each unordered} \\ \text{pair } (u,-u)}} [P_{*n}(A_n \mid u) + P_{*n}(\tilde{A}_u \mid -u)] \right]$$
$$\leqslant \frac{1}{N} + \frac{1}{N} \frac{N-1}{2} < \frac{2}{3} \qquad \text{for} \quad N > 3.$$

Hence $\bar{C}_p{}^* = 0$.

RECEIVED: June 22, 1970; REVISED: January 25, 1971

REFERENCES

AHLSWEDE, R. (1971), Group codes do not achieve Shannon's channel capacity for general discrete channels, *Ann. of Math. Stat.* **42**, No. 1, 224-240.

DOBRUSHIN, R. L. (1963), Asymptotic optimality of group and systematic codes for some channels, *Theory of Probability and Applications* **8**, 47–59.

ELIAS, P. (1955), Coding for noisy channels, *IRE Convention Record*, Part 4, pp. 37–46.

PETERSON, W. W. (1961), "Error Correcting Codes," MIT Press and John Wiley and Sons, New York.

SHANNON, C. E. (1957), Certain results in coding theory for noisy channels, *Information and Control* **1**, 6–25.

WOLFOWITZ, J. (1960), Simultaneous channels, *Arch. Rat. Mech. Anal.* **4**, 371–386.

WOLFOWITZ, J. (1964), "Coding Theorems of Information Theory," 2nd Ed., Springer-Verlag, Berlin/New York.

Printed in Belgium by the St. Catherine Press Ltd., Tempelhof 37, Bruges