ON TWO-WAY COMMUNICATION CHANNELS AND A PROBLEM BY ZARANKIEWICZ

RUDOLF AHLSWEDE

COLUMBUS and URBANA

SUMMARY

Shannon's two-way communication channels are studied in case of no feedback. Let $(n, N_1, N_2, \lambda_1, \lambda_2)$ denote a code for the two-way channel with word length n, code length N_1 and maximal error probability λ_1 in one direction, and code length N_2 and maximal error probability λ_2 in the other direction.

The following "strong converse type" estimate is proved (Theorem 1 in Section 3): Given $\varepsilon > 0$, λ_1, λ_2 strictly between 0 and 1, then every code $(n, N_1, N_2, \lambda_1, \lambda_2)$ satisfies

$$\left[\frac{1}{n}\log N_1 - \varepsilon, \quad \frac{1}{n}\log N_2 - \varepsilon\right] \in G_{\mathrm{I}}$$

for all sufficiently large n, where G_{I} denotes the inner capacity region.

This result implies that one can achieve a pair of rates $(R_1, R_2) \notin G_{\mathrm{I}}$ with codes of maximal error only if at least one of the error probabilities tends to one as the word length n tends to infinity.

Zarankiewicz [18] posed the problem to find the least $k = k_j(n)$ so that an $n \times n$ – matrix containing k ones and $n^2 - k$ zeros, no matter how distributed, contains a $j \times j$ submatrix (minor) consisting entirely of ones.

Theorem 2 (Section 4) gives the *lower* bound

$$k_i(n) \geqq (i!)^{2 \cdot i^{-2}} n^{2-(2/i)} \quad \text{for all } n \text{ and all } \; i \leqq n\,.$$

Using this result it is shown (Section 5) that in general one cannot reduce a code with average errors for two-way channels to a code with maximal errors without an essential loss in code length or error probability, whereas for one-way channels it is unessential whether one uses average or maximal errors.

1. INTRODUCTION

In the following we give a formal description of Shannon's two-way communication channel [7] which we abbreviate as t.w.c. – and we restate the main results about them. For proofs we refer to [7], familiarity with which we assume. We adopt the following notation:

Let $X = \{1, \ldots, a\}$, $Y = \{1, \ldots, b\}$, $\bar{X} = \{1, \ldots, \bar{a}\}$, $\bar{Y} = \{1, \ldots, \bar{b}\}$ be finite sets. $X(Y)$ and $\bar{Y}(\bar{X})$ are the input and output alphabets, respectively, at terminal I (II) of the t.w.c.

Write $X^t = X$, $Y^t = Y$, $\bar{X}^t = \bar{X}$, $\bar{Y}^t = \bar{Y}$ for $t = 1, 2, \ldots$

Define

$$X_n = \prod_{t=1}^{n} X^t, \quad Y_n = \prod_{t=1}^{n} Y^t, \quad \bar{X}_n = \prod_{t=1}^{n} \bar{X}^t, \quad \bar{Y}_n = \prod_{t=1}^{n} \bar{Y}^t \quad \text{for} \quad n = 1, 2, \ldots$$

Let $w(\bar{x}, \bar{y} \mid x, y)$ be a non-negative function defined for every element $(x, y, \bar{x}, \bar{y})$ of $X \times Y \times \bar{X} \times \bar{Y}$, and such that

$$\sum_{\bar{x} \in \bar{X}} \sum_{\bar{y} \in \bar{Y}} w(\bar{x}, \bar{y} \mid x, y) = 1 \tag{1.1}$$

for every $(x, y) \in X \times Y$.

The transition probabilities of a t.w.c. are defined by

$$P(\bar{x}_n, \bar{y}_n \mid x_n, y_n) = \prod_{t=1}^{n} w(\bar{x}^t, \bar{y}^t \mid x^t, y^t) \tag{1.2}$$

for every $x_n = (x^1, \ldots, x^n) \in X_n$, $y_n = (y^1, \ldots, y^n) \in Y_n$, $\bar{x}_n = (\bar{x}^1, \ldots, \bar{x}^n) \in \bar{X}_n$ and every $\bar{y}_n = (\bar{y}^1, \ldots, \bar{y}^n) \in \bar{Y}_n$, $n = 1, 2, \ldots$

A code (n, N_1, N_2) for the t.w.c. – neglecting feedback – is a system

$$\{(u_i, v_j, A_{ij}, B_{ij}) \mid i = 1, \ldots, N_1; j = 1, \ldots, N_2\} \tag{1.3}$$

where $u_i \in X_n$, $v_j \in Y_n$, $A_{ij} \subset \bar{X}_n$, $B_{ij} \subset \bar{Y}_n$ for $i = 1, \ldots, N_1$; $j = 1, \ldots, N_2$ and for fixed j, $j = 1, \ldots, N_2$,

$$A_{ij} \cap A_{i'j} = \emptyset \quad \text{for} \quad i \neq i' \tag{1.4}$$

and for fixed i, $i = 1, \ldots, N_1$,

$$B_{ij} \cap B_{ij'} = \emptyset \text{ for } j \neq j'.$$

For $A \subset \bar{X}_n$, $B \subset \bar{Y}_n$ define

$$P(A \mid x_n, y_n) = \sum_{\bar{x}_n \in A} \sum_{\bar{y}_n \in \bar{Y}_n} P(\bar{x}_n, \bar{y}_n \mid x_n, y_n) \tag{1.5}$$

for $(x_n, y_n) \in X_n \times Y_n$.

A code (n, N_1, N_2) is an $(n, N_1, N_2, \lambda_1, \lambda_2)$ code if

$$P(A_{ij} \mid u_i, v_j) \geqq 1 - \lambda_1 \quad \text{and} \quad P(B_{ij} \mid u_i, v_j) \geqq 1 - \lambda_2 \tag{1.6}$$

for $i = 1, \ldots, N_1$; $j = 1, \ldots, N_2$.

A code (n, N_1, N_2) is an $(n, N_1, N_2, \bar{\lambda}_1, \bar{\lambda}_2)$ code if

$$(1.7)\qquad \frac{1}{N_1 \cdot N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(A_{ij} \mid u_i, v_j) \geqq 1 - \bar{\lambda}_1$$

and

$$\frac{1}{N_1 \cdot N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(B_{ij} \mid u_i, u_j) \geqq 1 - \bar{\lambda}_2 .$$

We say an $(n, N_1, N_2, \lambda_1, \lambda_2)$ code is a code with *maximal* errors λ_1, λ_2 and an $(n, N_1, N_2, \bar{\lambda}_1, \bar{\lambda}_2)$ code is a code with *average* errors $\bar{\lambda}_1, \bar{\lambda}_2$.

Obviously, every $(n, N_1, N_2, \lambda_1, \lambda_2)$ code is an $(n, N_1, N_2, \bar{\lambda}_1, \bar{\lambda}_2)$ code for $\lambda_1 = \bar{\lambda}_1$, $\lambda_2 = \bar{\lambda}_2$; the converse is not true.

We give now some further definitions.

Let p be a probability distribution on X and q be a probability distribution on Y. Define

$$(1.8)\qquad R_{12}(p, q) = \sum_{y \in Y} q(y) \sum_{x \in X} \sum_{\bar{x} \in X} p(x)\, w(\bar{x} \mid x, y) \log \frac{w(\bar{x} \mid x, y)}{\sum_{x \in X} p(x)\, w(\bar{x} \mid x, y)} ,$$

$$(1.9)\qquad R_{21}(p, q) = \sum_{x \in X} p(x) \sum_{y \in Y} \sum_{\bar{y} \in Y} q(y)\, w(\bar{y} \mid x, y) \log \frac{w(\bar{y} \mid x, y)}{\sum_{y \in Y} q(y)\, w(\bar{y} \mid x, y)}$$

and

$$(1.10)\qquad G_{\mathrm{I}} = \text{convex closed hull of the set}$$

$$\{(R_{12}(p, q), R_{21}(p, q)) \mid p, q \text{ prob. distr. on } X, Y\} ,$$

where the closure is taken with respect to the natural topology in the euclidean plane E^2.

It was proved in [7], page 625 that G_{I} contains with every point (R_1, R_2) also the projections $(R_1, 0)$, $(0, R_2)$.

Denote by $G_{\mathrm{I}}(\varepsilon)$ the points in E^2 which have a distance less than ε from G_{I}.

We are now ready to state Shannon's main result.

THEOREM S (Theorem 3 and equation (34) of [7]).

a) (Coding theorem) *For any point* (R_1, R_2) *in* G_{I} *and any* $\varepsilon > 0$ *there exists a code*

$$(n, N_1, N_2, \bar{\lambda}_1, \bar{\lambda}_2) = (n, e^{(R_1 - \varepsilon)n}, e^{(R_2 - \varepsilon)n}, e^{-A(\varepsilon)n}, e^{-A(\varepsilon)n})$$

for all sufficiently large n, and some positive $A(\varepsilon)$.

b) (Weak converse to the coding theorem) *Given* $\varepsilon > 0$, *then there exist* $\bar{\lambda}_1(\varepsilon)$, $\bar{\lambda}_2(\varepsilon)$ *strictly between* 0 *and* 1, *such that every code* $(n, N_1, N_2, \bar{\lambda}_1(\varepsilon), \bar{\lambda}_2(\varepsilon))$ *satisfies* $[(1/n) \log N_1, (1/n) \log N_2] \in G_{\mathrm{I}}(\varepsilon)$ *for all sufficiently large n.*

REMARK. Libkind [6] proved the weak converse also in case of feedback.

Shannon's proof of the coding theorem uses his random coding method [8] and in so far no other proof has been given, whereas for discrete memoryless channels (d.m.c) and several other channels different methods for a proof of the coding theorem exist (see for instance [10]).

Theorem S establishes the coding theorem for average errors, that is, code concept (1.7) is used, Shannon's random coding method works only for average errors. It seems to the author that a drawback of code concept (1.7) is that a small error probability is guaranteed only if both senders use their code words with equal probabilities. For a d.m.c. it is unimportant whether we work with average or with maximal errors (cf. [10], Lemma 4.2.1). However, for compound channels it already makes a difference for rates above capacity. The strong converse of the coding theorem holds in this case for maximal but not for average errors (cf. [2], [4]). This shows that even though Shannon used in his coding theory average errors only – which may be appropriate for all practical communication problems – there is certainly from a purely mathematical point of view a theory of coding for average errors and a theory of coding for maximal errors.

In section 3 we prove the following "strong converse type" estimate for t.w.c.: given $\varepsilon > 0$, λ_1, λ_2 strictly between 0 and 1, then every code $(n, N_1, N_2, \lambda_1, \lambda_2)$ satisfies $[(1/n) \log N_1 \; (1/n) \log N_2] \in G_I(\varepsilon)$ for all sufficiently large n. This implies that we can achieve a pair of rates $(R_1, R_2) \notin G_I$ with codes of maximal error only if at least one of the error probabilities tends to one as the word length n tends to infinity.

One would like to have a result like this also for average errors.

In case of a d.m.c. we can reduce a code with average error to a code with maximal error and still maintain the rate simply by application of Lemma 4.2.1 in [10].

In Section 5 we prove that the analogous result is not true for t.w.c. Our proof uses an estimate concerning a problem of Zarankiewicz [18], which we derive in Section 4.

It seems not unlikely that for maximal errors the region of achievable rates G_I^* is in general smaller than G_I.

2. AUXILIARY RESULTS

LEMMA 1. *Let Z_s, $s = 1, \ldots, d$, be non-negative chance variables, defined on the same probability space, such that*

$$EZ_s \leqq \alpha\,, \quad s = 1, \ldots, d\,.$$

For any $\beta > 0$ the probability of

$$B^* = \{Z_s \leqq d(\alpha + \beta) \text{ for } s = 1, \ldots, d\}$$

satisfies

$$P(B^*) \geqq \frac{\beta}{\alpha + \beta}.$$

This is a trivial refinement of the lemma in [7], for a proof see [3], page 467.

In [1] we proved (a coding theorem and) a strong converse of the coding theorem for *non-stationary* d.m.c., thus generalizing the results of [9] from the stationary to the nonstationary case. Then Augustin [5] found a simpler proof. His main result is stated as Lemma 2 below. (Compare [11] for a related result.)

Before we can state the Lemma, we need some preparatory definitions.

Let $\tilde{X} = \{1, \ldots, \tilde{a}\}$, $\tilde{Y} = \{1, \ldots, \tilde{b}\}$ and define

$$\tilde{X}_n = \prod_1^n \tilde{X}, \quad \tilde{Y}_n = \prod_1^n \tilde{Y} \quad \text{for} \quad n = 1, 2, \ldots$$

Let $(F^t(\cdot|\cdot))_{t=1,2,\ldots}$ be a sequence of stochastic matrices, i.e.

(2.1) $$F^t(\tilde{y} \mid \tilde{x}) \geqq 0 \quad \text{for every} \quad \tilde{x} \in \tilde{X}, \; \tilde{y} \in \tilde{Y}$$

and

$$\sum_{\tilde{y} \in \tilde{Y}} F^t(\tilde{y} \mid \tilde{x}) = 1 \quad \text{for every} \quad \tilde{x} \in \tilde{X} \quad \text{and} \quad t = 1, 2, \ldots$$

The transition probabilities of a nonstationary d.m.c. are defined by

(2.2) $$F(\tilde{y}_n \mid \tilde{x}_n) = \prod_{t=1}^n F^t(\tilde{y}^t \mid \tilde{x}^t) \quad \text{for every} \quad \tilde{x}_n = (\tilde{x}^1, \ldots, \tilde{x}^n) \in \tilde{X}_n$$

$$\text{and every } \tilde{y}_n = (\tilde{y}^1, \ldots, \tilde{y}^n) \in \tilde{Y}_n, \; n = 1, 2, \ldots$$

An (n, N, λ) code for the nonstationary d.m.c. F is a system

$$\{(\tilde{u}_i, A_i) \mid i = 1, \ldots, N\},$$

where $\tilde{u}_i \in \tilde{X}_n$, $A_i \subset \tilde{Y}_n$, $i = 1, \ldots, N$, $A_i \cap A_j = \emptyset$ for $i \neq j$ and which satisfies

(2.3) $$F(A_i \mid \tilde{u}_i) \geqq 1 - \lambda, \quad i = 1, \ldots, N.$$

Let $\{(\tilde{u}_i = (\tilde{u}_i^1, \ldots, \tilde{u}_i^n), A_i) \mid i = 1, \ldots, N\}$ be an (n, N, λ) code for F.

Define

(2.4) $$\pi^t(\tilde{x}) = \frac{|\{i \mid \tilde{u}_i^t = \tilde{x}, i \in \{1, \ldots, N\}\}|}{N} \quad \text{for} \quad \tilde{x} \in \tilde{X}, \quad t = 1, 2, \ldots, n,$$

and

(2.5) $$\pi'^t(\tilde{y}) = \sum_{\tilde{x} \in \tilde{X}} \pi^t(\tilde{x}) F^t(\tilde{y} \mid \tilde{x}) \quad \text{for} \quad \tilde{y} \in \tilde{Y}, \quad t = 1, 2, \ldots, n,$$

and

$$R_n = \sum_{t=1}^{n} \frac{1}{N} \sum_{i=1}^{N} \sum_{\tilde{y}^t \in \tilde{Y}^t} F^t(\tilde{y}^t \mid \tilde{u}_i^t) \log \frac{F^t(\tilde{y}^t \mid \tilde{u}_i^t)}{\pi''^t(\tilde{y}^t)} . \tag{2.6}$$

LEMMA 2 (Theorem 3 of [5]). *Let* $\{(\tilde{u}_i, A_i) \mid i = 1, \ldots, N\}$ *be an* (n, N, λ) *code and let* $\pi^t(\cdot)$, $\pi''^t(\cdot)$, $t = 1, \ldots, n$, *be defined as in* (2.4), (2.5). *The following estimates hold for any* λ, d, $0 < \lambda, d < 1$; $n = 1, 2, \ldots$:

a) $$\log(\lambda\, dN) \leqq R_n + \frac{1}{\lambda(1-d)} k(\tilde{a}) \sqrt{n} ,$$

where k *depends only on* $\tilde{a}$ *and not on* $(F^t)_{t=1,2,\ldots}$.

b) $$R_n = \sum_{t=1}^{n} \sum_{\tilde{x}^t} \sum_{\tilde{y}^t} \pi^t(\tilde{x}^t)\, F^t(\tilde{y}^t \mid \tilde{x}^t) \,.\, \log \frac{F^t(\tilde{y}^t \mid \tilde{x}^t)}{\pi''^t(\tilde{y}^t)} .$$

The strong converse of the coding theorem for nonstationary d.m.c. is an immediate consequence of Lemma 2.

3. A STRONG CONVERSE TYPE ESTIMATE FOR T.W.C. WITHOUT FEEDBACK FOR MAXIMAL ERRORS

In this section we shall prove the

THEOREM 1 (Strong converse for t.w.c. without feedback). *Given* $\varepsilon > 0$, λ_1, λ_2 *strictly between* 0 *and* 1, *then every code* $(n, N_1, N_2, \lambda_1, \lambda_2)$ *satisfies*

$$\left(\frac{1}{n} \log N_1, \frac{1}{n} \log N_2\right) \in G_I(\varepsilon)$$

for all sufficiently large n.

Proof. Let $\{(u_i, v_j, A_{ij}, B_{ij}) \mid i = 1, \ldots, N_1; j = 1, \ldots, N_2\}$ be an $(n, N_1, N_2, \lambda_1, \lambda_2)$ code for the t.w.c.

Write $u_i = (u_i^1, \ldots, u_i^n)$ for $i = 1, \ldots, N_1$ and $v_j = (v_j^1, \ldots, v_j^n)$ for $j = 1, \ldots, N_2$ and define

$$p^t(x) = \frac{|\{i \mid u_i^t = x, i \in \{1, \ldots, N_1\}\}|}{N_1} \tag{3.1}$$

for $x \in X^t$, $t = 1, 2, \ldots, n$, and

$$q^t(y) = \frac{|\{j \mid v_j^t = y, j \in \{1, \ldots, N_2\}\}|}{N_2} \tag{3.2}$$

for $y \in Y^t$, $t = 1, 2, \ldots, n$. $p^t(\cdot)$ is a probability distribution on X^t and $q^t(\cdot)$ is a probability distribution on Y^t.

For every $v_j = (v_j^1, \ldots, v_j^n)$, $j = 1, \ldots, N_2$, we define probability distributions $\bar{p}^t(\cdot \mid v_j^t)$ on $\bar{X}^t$, $t = 1, \ldots, n$, by

$$\bar{p}^t(\bar{x}^t \mid v_j^t) = \sum_{x \in X^t} p^t(x)\, w(\,{}^t \mid x, v_j^t) \tag{3.3}$$

for all $\bar{x}^t \in \bar{X}^t$.

For every $u_i = (u_i^1, \ldots, u_i^n)$, $i = 1, \ldots, N_1$, we define probability distributions $\bar{q}^t(\cdot \mid u_i^t)$ on $\bar{Y}^t$, $t = 1, 2, \ldots, n$, by

$$\bar{q}^t(\bar{y}^t \mid u_i^t) = \sum_{y \in Y^t} q^t(y)\, w(\bar{y} \mid u_i^t, y) \tag{3.4}$$

for all $\bar{y}^t \in \bar{Y}^t$.

For fixed v_j, $P(\cdot \mid \cdot, v_j)$ given by

$$P(\bar{x}_n \mid x_n, v_j) = \prod_{t=1}^{n} w(\bar{x}^t \mid x^t, v_j^t) \tag{3.5}$$

for all $x_n = (x^1, \ldots, x^n) \in X_n$, $\bar{x}_n = (\bar{x}^1, \ldots, \bar{x}^n) \in \bar{X}_n$, defines the transition probabilities of a nonstationary d.m.c. for words of lengths n. Similarly, for fixed u_i, $P(\cdot \mid u_i, \cdot)$ given by

$$P(\bar{y}_n \mid u_i, y_n) = \prod_{t=1}^{n} w(\bar{y}^t \mid u_i^t, y^t) \tag{3.6}$$

for all $y_n = (y^1, \ldots, y^n) \in Y_n$, $\bar{y}_n = (\bar{y}^1, \ldots, \bar{y}^n) \in \bar{Y}_n$, defines the transition probabilities of a nonstationary d.m.c. for words of length n.

Define

$$R_{12}(v_j^t) = \sum_{x^t} p^t(x^t)\, w(\bar{x}^t \mid x^t, v_j^t) \,.\, \log \frac{w(\bar{x}^t \mid x^t, v_j^t)}{\bar{p}^t(\bar{x}^t)} \tag{3.7}$$

for $j = 1, \ldots, N_2$, $t = 1, 2, \ldots, n$ and

$$R_{21}(u_i^t) = \sum_{y^t} q^t(y^t)\, w(\bar{y}^t \mid u_i^t, y^t) \,.\, \log \frac{w(\bar{y}^t \mid u_i^t, y^t)}{\bar{q}^t(\bar{y}^t)} \tag{3.8}$$

for $i = 1, \ldots, N_1$, $t = 1, 2, \ldots, n$.

We denote $\sum_{t=1}^{n} R_{12}(v_j^t)$ by $R_{12}(v_j)$ and $\sum_{t=1}^{n} R_{21}(u_i^t)$ by $R_{21}(u_i)$.

$\{(u_i, A_{ij}) \mid i = 1, \ldots, N_1\}$ is a code with maximal error λ_1 for all $P(\cdot \mid \cdot, v_j)$, $j = 1, \ldots, N_2$; and $\{(v_j, B_{ij}) \mid j = 1, \ldots, N_2\}$ is a code with maximal error λ_2 for all $P(\cdot \mid u_i, \cdot)$, $i = 1, \ldots, N$.

Application of Lemma 2 yields for $d = \frac{1}{2}$

$$\log(\lambda_1 \,.\, \tfrac{1}{2} N_1) \leqq R_{12}(v_j) + \frac{2}{\lambda_1} k(a) \sqrt{n} \quad \text{for} \quad j = 1, \ldots, N_2 \tag{3.9}$$

and

$$\log(\lambda_2 \cdot \tfrac{1}{2}N_2) \leqq R_{21}(u_i) + \frac{2}{\lambda_2}\, k(b)\sqrt{n} \quad \text{for} \quad i = 1, \dots, N_1\,. \tag{3.10}$$

We write the system of inequalities (3.9) more explicitly as

$$\begin{aligned} \log N_1 &\leqq R_{12}(v_1^1) + \dots + R_{12}(v_1^n) + \frac{2}{\lambda_1}\, k(a)\sqrt{n} - \log\frac{\lambda_1}{2}\,, \\ \log N_1 &\leqq R_{12}(v_2^1) + \dots + R_{12}(v_2^n) + \frac{2}{\lambda_1}\, k(a)\sqrt{n} - \log\frac{\lambda_1}{2}\,, \\ &\dots \\ \log N_1 &\leqq R_{12}(v_{N_2}^1) + \dots + R_{12}(v_{N_2}^n) + \frac{2}{\lambda_1}\, k(a)\sqrt{n} - \log\frac{\lambda_1}{2}\,. \end{aligned} \tag{3.11}$$

Summing the right sides of the inequalities and dividing by N_2 yields

$$\log N_1 \leqq \frac{1}{N_2}\sum_{j=1}^{N_2}\sum_{t=1}^{n} R_{12}(v_j^t) + \frac{2}{\lambda_1}\, k(a)\sqrt{n} - \log\frac{\lambda_1}{2}\,. \tag{3.12}$$

(3.12) and (3.1) imply

$$\log N_1 \leqq \sum_{t=1}^{n}\sum_{x^t\in X^t} p^t(x^t)\, R_{12}(x^t) + \frac{2}{\lambda_1}\, k(a)\sqrt{n} - \log\frac{\lambda_1}{2}\,. \tag{3.13}$$

Analogously one can show that

$$\log N_2 \leqq \sum_{t=1}^{n}\sum_{y^t\in Y^t} q^t(y^t)\, R_{21}(y^t) + \frac{2}{\lambda_2}\, k(b)\sqrt{n} - \log\frac{\lambda_2}{2}\,. \tag{3.14}$$

Recalling definitions (1.8), (1.9) we see that

$$\sum_{x^t\in X^t} p^t(x^t)\, R_{12}(x^t) = R_{12}(p^t, q^t)$$

and

$$\sum_{y^t\in Y^t} q^t(y^t)\, R_{21}(y^t) = R_{21}(p^t, q^t)\,.$$

We obtain therefore from (3.13), (3.14) that

$$\frac{1}{n}\log N_1 \leqq \frac{1}{n}\sum_{t=1}^{n} R_{12}(p^t, q^t) + \frac{2}{\lambda_1}\, k(a)\, n^{-1/2} - \frac{1}{n}\log\frac{\lambda_1}{2} \tag{3.15}$$

and

$$\frac{1}{n}\log N_2 \leqq \frac{1}{n}\sum_{t=1}^{n} R_{21}(p^t, q^t) + \frac{2}{\lambda_2} k(b)\, n^{-1/2} - \frac{1}{n}\log\frac{\lambda_2}{2}\,. \tag{3.16}$$

The theorem follows now from (3.15) and (3.16).

REMARK 1. Lemma 2 does not hold for average errors. Since, if the contrary would be true, one could use it to prove a strong converse of the coding theorem for *compound* channels with average errors by arguments used in [1], page 37. But this would be a contradiction to Theorem 1 of [4].

REMARK 2. One can prove b) in Theorem 5 by using Theorem 1, Lemma 1 and an additional approximation argument. But this proof is relatively complicated as compared to the proof given by Shanonn and we therefore omit the lengthy details of this argument.

4. A RANDOM VERSION OF A PROBLEM BY ZARANKIEWICZ

Zarankiewicz [18] posed the problem to find the least $k = k_j(n)$ so that an $n \times n$ – matrix, containing k ones and $n^2 - k$ zeros, no matter how distributed, contains a $j \times j$ submatrix (minor) consisting entirely of ones. This problem naturally generalizes to that of finding the least $k = k_{i,j}(m, n)$ so that an $m \times n$ matrix containing k ones and $mn - k$ zeros, no matter how distributed, contains an $i \times j$ submatrix consisting entirely of ones. Several asymptotic and non-asymptotic results have been obtained under various conditions on m, n, i, j. (See references [12], ..., [20], and especially [13] for a more systematic account.)

We limit ourself here to the case $m = n$, $i = j$ – even so our results can be generalized –, and we are interested only in asymptotic results.

Hartman, Mycielski and Ryll - Nardzewski [14] obtained bounds for $k_2(n)$, which were improved by Kövari, Sos and Turan [16], who showed that

$$\lim_{n\to\infty} n^{-3/2}\, k_2(n) = 1\,. \tag{4.1}$$

Brown (see [13]) proved the first inequality in

$$2^{-1} \leqq \overline{\lim_{n\to\infty}}\, n^{-5/3}\, k_3(n) \leqq 2^{-2/3} \tag{4.2}$$

thus partially confirming a conjecture of Kövari et al. [16], who gave the second inequality, and of Erdös (see [13]). The existence of $\lim_{n\to\infty} n^{-5/3}\, k_3(n)$ is still unproved.

For $i \geqq 4$ only *upper* bounds on $k_i(n)$ are known (cf. [13], page 130).

Recently Guy and Znám [13] proved by a simple application of the pigeon-hole principle the

THEOREM. *If an $m \times n$ matrix contains more than nu ones, and it can be shown that*

$$n\binom{u}{i} \geqq (j-1)\binom{m}{i},$$

then there is an $i \times j$ submatrix consisting entirely of ones.

As an immediate consequence of this Theorem one obtains

$$\varlimsup_{n\to\infty} k_i(n)\, n^{-(2-1/i)} \leqq (i-1)^{1/i} \quad \text{for} \quad i \geqq 2\,. \tag{4.3}$$

This bound is sharp for $i = 2$, but not sharp for $i = 3$ as can be seen by comparison with (4.1) and (4.2).

By a rather simple reasoning we obtain the *lower* bound on $k_i(n)$:

$$k_i(n)\, n^{-(2-2/i)}\, (i!)^{-2/i^2} \geqq 1 \quad \text{for all } n \text{ and all} \quad i \leqq n\,. \tag{4.4}$$

(4.4) together with (4.3) gives a good estimate on $k_i(n)$ for larger values of i.

In the sequel we shall refer to the problem of Zarankiewicz as Problem Z and to the problem, which we introduce now, as Problem R.

Let $\overline{M}(n, k)$ be the set of all $n \times n$ – matrices with k ones and $n^2 - k$ zeros in its entries.

Clearly,

$$|\overline{M}(n, k)| = \binom{n^2}{k}. \tag{4.5}$$

Let $R(n, k)$ be a random matrix with values in $\overline{M}(n, k)$. We assume that $R(n, k)$ takes any value $M(n, k)$, $M(n, k) \in \overline{M}(n, k)$, with probability $\binom{n^2}{k}^{-1}$. Whether $R(n, k)$ contains an $i \times i$ – submatrix with all entries one is now a matter of chance.

Denote by $k_i(n, \varepsilon)$ the smallest integer k for which the probability $p(n, k, i, \varepsilon)$ that $R(n, k)$ contains an $i \times i$ – submatrix with all entries one is greater than or equal to $1 - \varepsilon$, where $0 < \varepsilon < 1$.

Problem R consists in finding estimates for $k_i(n, \varepsilon)$.

If we allow the value $\varepsilon = 0$ in the definition given above then we obtain that $k_i(n)$ equals $k_i(n, 0)$.

Since $k_i(n, \varepsilon)$ increases as ε decreases we get

$$k_i(n) \geqq k_i(n, \varepsilon) \tag{4.6}$$

for all ε, $0 < \varepsilon < 1$, and also

$$k_i(n) \geqq k_i(n, \varepsilon_n) \tag{4.7}$$

for every sequence $(\varepsilon_n)_{n=1,2,\ldots}$ converging to 0.

We shall make use only of relation (4.6).

Lower bounds on $k_i(n, \varepsilon)$ are a fortiori lower bounds on $k_i(n)$.

Denote by $T(n, i)$ the total number of $i \times i$ – submatrices of an $n \times n$ – matrix. We have

$$(4.8) \qquad T(n, i) = \binom{n}{i}^2 .$$

An $(i \times i, 1)$ – submatrix is an $i \times i$ – submatrix (of an $n \times n$ – matrix) with all its entries one. We denote the number of matrices in $\overline{M}(n, k)$ which contain a particular $(i \times i, 1)$ – submatrix by $N(n, k, i)$.

Obviously,

$$(4.9) \qquad N(n, k, i) = \binom{n^2 - i^2}{k - i^2} .$$

$N^*(n, k, i)$ shall count the matrices in $\overline{M}(n, k)$ which contain *at least* one $(i \times i, 1)$ – submatrix.

Let $\overline{L}$ be a system of l $(i \times i, 1)$ – submatrices and let $\overline{M}(n, k, \overline{L})$ be the subset of matrices of $\overline{M}(n, k)$ which contain every element of $\overline{L}$ as a submatrix.

We define now $F_l(n, k, i)$ by

$$(4.10) \qquad F_l(n, k, i) = \sum_{|\overline{L}| = l} |\overline{M}(n, k, \overline{L})| \quad \text{for} \quad l = 1, 2, \ldots$$

For $F_1(n, k, i)$ we obtain

$$(4.11) \qquad F_1(n, k, i) = T(n, i) \cdot N(n, k, i) .$$

It follows from the inclusion-exclusion-principle that we can express $N^*(n, k, i)$ in terms of the $F_l(n, k, i)$. We state this more explicitly as

LEMMA.

a) $$N^*(n, k, i) = \sum_{l=1}^{k} (-1)^{l+1} F_l(n, k, i) ,$$

b) $$\sum_{l=1}^{t} (-1)^{l+1} F_l(n, k, i) \leq N^*(n, k, i) \leq \sum_{l=1}^{t+1} (-1)^{l+1} F_l(n, k, i)$$

for any even integer t.

Since $p(n, k, i) = N^*(n, k, i) \binom{n^2}{k}^{-1}$, good estimates on $F_l(n, k, i)$ would lead to good estimates for $k_i(n, \varepsilon)$. However, it seems to be not easy to get those estimates. The reason for this is that elements of $\overline{L}$ may have entries in common for a large proportion of $\overline{L}$'s.

It follows from the Lemma that

$$N^*(n, k, i) \leq F_1(n, k, i) ,$$

and from (4.8), (4.9), and (4.11) that

$$N^*(n, k, i) \leqq \binom{n^2 - i^2}{k - i^2}\binom{n}{i}^2 .$$

Since $p(n, k, i) = N^*(n, k, i)\binom{n^2}{k}^{-1}$ we obtain

$$p(n, k, i) \leqq \binom{n^2 - i^2}{k - i^2}\binom{n}{i}^2\binom{n^2}{k}^{-1} \tag{4.12}$$

and by the definition of $k_i(n, \varepsilon)$ and $k_i(n)$ also

$$\binom{n^2 - i^2}{k_i(n, \varepsilon) - i^2}\binom{n}{i}^2\binom{n^2}{k_i(n, \varepsilon)}^{-1} \geqq 1 - \varepsilon \tag{4.13}$$

and

$$\binom{n^2 - i^2}{k_i(n) - i^2}\binom{n}{i}^2\binom{n^2}{k_i(n)}^{-1} \geqq 1 . \tag{4.14}$$

One easily verifies that $n^{2i}/(i!)^2 \,.\, (k_i(n))^{i^2}/n^{2i^2}$ is greater than the left side in (4.14) and therefore also

$$\frac{n^{2i}}{(i!)^2}\frac{(k_i(n))^{i^2}}{n^{2i^2}} \geqq 1 . \tag{4.16}$$

We thus have proved

THEOREM 2.

$$k_i(n) \geqq (i!)^{2 . i^{-2}} \,.\, n^{2(1-1/i)} \tag{4.17}$$

for all n and all $i \leqq n$.

Theorem 2 implies

$$k_i(n) \geqq n^{2-2/i} \quad \text{for all } n \text{ and all } \quad i \leqq n . \tag{4.18}$$

5. ON THE RELATIONSHIP BETWEEN CODES WITH AVERAGE ERROR AND CODES WITH MAXIMAL ERROR FOR T.W.C.

For one-way channels it is unessential whether we use average or maximal errors. This is due to the simple fact that an $(n, N, \bar{\lambda})$ code $\{(u_i, A_i) \mid i = 1, \ldots, N\}$ contains a subcode $\{(u_{i_v}, A_{i_v}) \mid v = 1, \ldots, [N/2]\}$ which is an $(n, [N/2], \lambda)$ code for $\lambda \geqq 2\bar{\lambda}$ (cf. Lemma 4.2.1 in [10]).

In this section we shall prove that in general one *cannot* reduce a code with average error for t.w.c. to a code with maximal error without an essential loss in code length or error probability.

Let $\{(u_i, v_j, A_{ij}, B_{ij}) \mid i = 1, \ldots, N_1; j = 1, \ldots, N_2\}$ be a code with average errors $\bar{\lambda}_1, \bar{\lambda}_2$, that is

$$(5.1) \qquad \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(A_{ij} \mid u_i, v_j) = 1 - \bar{\lambda}_1$$

and

$$(5.2) \qquad \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} P(B_{ij} \mid u_i, v_j) = 1 - \bar{\lambda}_2 .$$

Let $\lambda_1 > \bar{\lambda}_1$, $\lambda_2 > \bar{\lambda}_3$ and define

$$\bar{N} = \{(i, j) \mid i = 1, \ldots, N_1 \; j = 1, \ldots, N_2\} .$$

Let

$$\bar{N}(\lambda_1, \lambda_2) = \{(i, j) \mid (i, j) \in \bar{N},\; P(A_{ij} \mid u_i, v_j) > 1 - \lambda_1 \text{ and } P(B_{ij} \mid u_i, v_j) > 1 - \lambda_2\} .$$

The cardinality of $\bar{N}(\lambda_1, \lambda_2)$ depends on the distribution of the values $P(A_{ij} \mid u_i, v_j)$ and $P(B_{ij} \mid u_i, v_j)$.

We denote the minimal cardinality which $\bar{N}(\lambda_1, \lambda_2)$ achieves for values of $P(A_{ij} \mid u_i, v_j)$, $P(B_{ij} \mid u_i, v_j)$ $(i = 1, \ldots, N_1;\; j = 1, \ldots, N_2)$ satisfying (5.1) and (5.2) by $k(\lambda_1, \lambda_2)$. The minimal cardinality of $\bar{N}(\lambda_1, \lambda_2)$ is achieved if the $P(A_{ij} \mid u_i, v_j)$, $(i = 1, \ldots, N_1; j = 1, \ldots, N_2)$ take only the values 1 and $1 - \lambda_1$ and the $P(B_{ij} \mid u_i, v_i)$ take only the values 1 and $1 - \lambda_2$.

Hence,

$$k(\lambda_1, \lambda_2) \leqq \min_{i=1,2} \frac{\lambda_i - \bar{\lambda}_i}{\lambda_i} N_1 N_2 .$$

The system $\{(u_i, v_j), A_{ij}, B_{ij}) \mid (i, j) \in \bar{N}(\lambda_1, \lambda_2)\}$ is not a code in the sense of (1.3), (1.4), because $\bar{N}(\lambda_1, \lambda_2)$ is not a cartesian product of subsets of $\{1, \ldots, N_1\}$ and $\{1, \ldots, N_2\}$.

Our problem reduces now to the question whether we can find a set $G \subset \bar{N}(\lambda_1, \lambda_2)$ satisfying

a) $$G = G_1 \times G_2 , \quad G_1 \subset \{1, \ldots, N_1\} , \quad G_2 \subset \{1, \ldots, N_2\} ,$$

and

b) $$|G_1| \approx N_1 , \quad |G_2| \approx N_2 ,$$

where "$\approx$" means that the numbers are close to each other in a sense which we make precise later.

Let $M = (M_{st})$, $s = 1, \ldots, N_1$, $t = 1, \ldots, N_2$, be an $N_1 \times N_2$ – matrix with zeros and ones as entries, where

$$M_{st} = 1 \quad \text{for} \quad (s, t) \in \bar{N}(\lambda_1, \lambda_2),$$
$$M_{st} = 0 \quad \text{for} \quad (s, t) \notin \bar{N}(\lambda_1, \lambda_2).$$

The problem described above is equivalent to the problem to find a $|G_1| \times |G_2|$ – submatrix of M with all entries one.

We are lead to the problem of Zarankiewicz and we shall make use of (4.15)

We limit ourself to $N_1 = N_2 = N = e^{Rm}$, where $R > 0$, $m = 1, 2, \ldots$, and to $|G_1| = |G_2|$. $|G_1| \approx N_1$ shall mean that for any $\eta > 0$ $|G_1| \geqq Ne^{-\eta m}$ for all sufficiently large m. (4.18) yields for $i = Ne^{-\eta m}$ and $n = N$:

(5.4) $$k_i(N) \geqq e^{2Rm}(e^{Rm})^{-2\exp[-(R-\eta)m]}.$$

The right side in (5.3) is maximized for $\lambda_2 = \lambda_1 = 1$, therefore,

$$k(\lambda_1, \lambda_2) \leqq \min_{i=1,2} (1 - \bar{\lambda}_i) \,.\, N^2 \,.$$

However,

$$\lim_{m \to \infty} (e^{Rm})^{-2\exp[-(R-\eta)m]} = 1$$

and (5.4) imply that $k(\lambda_1, \lambda_2) < k_i(N)$ for m large enough. It is impossible to find the desired subcode.

In our argument we assumed $\bar{\lambda}_1, \bar{\lambda}_2$ to be constant. We can obtain this result also if

$$\bar{\lambda}_1 = e^{-E_1 m}, \quad \bar{\lambda}_2 = e^{-E_2 m} \quad \text{and} \quad E_1, E_2 < R, \quad \text{W.l.o.g.}$$

we can assume $E_1 \geqq E_2$.

Choose η such that $R - \eta > E$. It suffices to show that

(5.5) $$(1 - e^{-E_1 m}) < (e^{Rm})^{-2/\exp[(R-\eta)m]}$$

for all sufficiently large m, and for this it is enough to show that

(5.6) $$-e^{-E_1 m} < -2 \,.\, e^{-(R-\eta)m} \,.\, R \,.\, m \,.$$

(5.6) holds, because $R - \eta > E_1$.

In order to decide what happens in case $E_1, E_2 > R$ one would have to make a careful evaluation of the formula given in the Theorem by Guy and Znám.

ACKNOWLEDGEMENT. The author wishes to thank Professor Jack Wolfowitz for proposing the problem to find a strong converse to Theorem S, and for many stimulating conversations.

REFERENCES

A. Coding theory

[1] AHLSWEDE, R.: Beiträge zur Shannonschen Informationstheorie im Falle nichtstationärer Kanäle. Z. Wahrscheinlichkeitstheorie verw. Geb. *10* (1968), 1–42.

[2] AHLSWEDE, R.: Certain results in coding theory for compound channels. Proc. Coll. on Inf. Theory, Debrecen, Hungary, 35–60, 1967.

[3] AHLSWEDE, R., WOLFOWITZ, J.: Correlated decoding for channels with arbitrarily varying channel probability functions, Information and Control *14* (1969), 457–473.

[4] AHLSWEDE, R., WOLFOWITZ, J.: The structure of capacity functions for compound channels. Springer Lectures Notes, vol. 89, Prob. and Inf. Theory, Proc. of the International Symposium at McMaster University Canada, April 1968, 12–54, 1969.

[5] AUGUSTIN, U.: Gedächtnisfreie Kanäle für diskrete Zeit. Z. Wahrscheinlichkeitstheorie verw. Geb. *6* (1966), 10–61.

[6] LIBKIND, L. M.: Two-way discrete memoryless communication channels. Problemy Peredachi Informatsii *3* (1967), 2, 37–46.

[7] SHANNON, C. E.: Two-way communication channels. Proc. Fourth Berkeley Symposium, vol. I, 611–644.

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

[9] WOLFOWITZ, J.: The coding of messages subject to chance errors. Illinois Journ. Math. (1957), 591–606.

[10] WOLFOWITZ, J.: Coding theorems of information theory. Springer, Berlin–Heidelberg–New York – first edition 1961, second edition 1964.

[11] WOLFOWITZ, J.: Notes on a General Strong converse. Information and Control *12* (1968), 1–4.

B. Combinatorics (On a problem of Zarankiewicz)

[12] ČULÍK, K.: Poznámka k problému K. Zarankiewicze. Práce Brněnské základny ČSAV *26* (1955), 341–348.

[13] GUY, R. K., ZNÁM, S.: A problem of Zarankiewicz. In: W. T. Tutte (ed.): Recent Progress in Combinatorics. Academic Press, 1969.

[14] HARTMAN, S., MYCIELSKI, J., RYLL-NARDZEWSKI, C.: Colloq. Math. *3* (1954), 84–85.

[15] HYLTEN-CAVALLIUS, C.: On a combinatorial problem. Colloq. Math. *6* (1958), 59–65.

[16] KÖVARI, T., SOS, V., TURAN, P.: On a problem of K. Zarankiewicz. Colloq. Math. *3* (1954), 50–57.

[17] REIMAN, I.: Über ein Problem von K. Zarankiewicz. Acta Math. Acad. Sci. Hungar. *9* (1958), 269–279.

[18] ZARANKIEWICZ, K.: Problem P 101. Colloq. Math. *2* (1951), 301.

[19] ZNÁM, S.: On a combinatorial problem of K. Zarankiewicz. Colloq. Math. *11* (1963), 81–84.

[20] ZNÁM, S.: Two improvements of a result concerning a problem of K. Zarankiewicz. Colloq. Math. *13* (1965), 255–258.

THE OHIO STATE UNIVERSITY
and
UNIVERSITY OF ILLINOIS