Optimal Coding Strategies for Certain Permuting Channels

RUDOLF AHLSWEDE AND AMIRAM H. KASPI, ASSOCIATE MEMBER, IEEE

Abstract—We give optimal coding strategies for nonprobabilistic permuting (especially, "trapdoor") channels and also for a permuting relay channel. Our results open the door to a coding theory for nonprobabilistic (or deterministic) channels with memory.

I. Permuting Channel Models

WITHIN the broad class of finite-state channels [1] Blackwell's trapdoor channel (see [2], [3], and [6]) has attracted special interest. It has been generalized in [3] to permuting channels. Whereas in [1] the capacity is described as a "product space limit" for quite general finite-state channels, no "single-letter characterization" for the capacity is known, even for the seemingly simple trapdoor channel. Many possible permuting mechanisms exist which describe the behavior of certain channels. The permuting channel was introduced by Benjamin [3] as a model for typewriting, when most errors of the secretary consist in exchanging letters in the text.

In this paper we consider *nonprobabilistic* permuting channels. Especially as a model for the typing of natural languages, this model provides the following advantages. Several secretaries may type different parts of the text of a book. Even if their individual error probabilities were known, we might not know who types which part of the book. The situation can be worse in so far as the error probabilities are often not known at all or are even undefined. Our model includes these possibilities.

The nonprobabilistic permuting channel, to which we also refer as the P channel, can be described as follows. Consider a box that contains β balls at time $t = 0$. We assume that the balls are marked with numbers from $\{1, 2, \cdots, \alpha\}$. Thus the content or "state" of the box can be described as a multiset $S_0 = (s_0(1), \cdots, s_0(\alpha))$, where $s_0(i)$ is the number of balls marked with i and $\sum_{i=1}^{\alpha} s_0(i) = \beta$.

At times $t = 1, 2, \cdots$ a new ball is thrown into the box by a person or a device I, and one of the $\beta + 1$ balls now in the box is pulled out by a person or a device $\mathcal{O}$ and given to the decoder D. We consider the following two communication situations.

Situation J: I wants to transmit information to D. $\mathcal{O}$ takes arbitrary actions or (equivalently) tries to make the communication as difficult as possible; that is, he acts like a jammer. I does not know what is pulled out of the box.

Situation R: $\mathcal{O}$ wants to transmit information to D. I is cooperative; that is, he helps to optimize the transmission of information.

Further situations and coding problems are discussed in Section IV. We refer to the P channel in situation J also as a permuting jammer channel and in situation R as a permuting relay channel. In short, we speak of the PJ channel and the PR channel.

Motivation for the study of PJ channels was given earlier. PR channels address the question of how much information an be transmitted by modifications of a text, which are of the nature of secretarial mistakes (see [3]). The amount of information achievable depends on the structure of the text. It is zero for the text $111 \cdots 11$. Answers to the question may be of interest in analyzing historical documents, which were either accidentally or purposely altered during repeated copying.

One can conceive of much more general situations where the text is replaced by any configuration (for instance, pictures), and instead of the permutation rules any suitable set of rules describes the possible modifications. These investigations, of which our concrete results in Theorems 1 and 2 are just a beginning, are in the spirit of the abstract coding theory of [4].

II. On PJ Channels with Initial State S_0 Known to the Communicators

We are given $\mathscr{X} = \{1, 2, \cdots, \alpha\}$ and $S_0 = (s_0(1), \cdots, s_0(\alpha))$ with $\sum_{i=1}^{\alpha} s_0(i) = \beta$. In the notions to follow, their dependence on α, β, and S_0 will be omitted where this is unambiguous:

$x_t \in \mathscr{X}$ is the input and $y_t \in \mathscr{X}$ is the output at time t;

$$S_t = (s_t(1), \cdots, s_t(\alpha)) \tag{2.1}$$

is the state of the box after the tth output; (2.2)

$S_0 | x^n \rightsquigarrow y^n | S_n$ indicates that y^n can be produced as an output sequence by a suitable strategy of the jammer when $x^n \in \mathscr{X}^n$ is sent. (2.3)

Manuscript received December 19, 1984; revised August 24, 1986. This paper was presented at a workshop on information theory in Hakone, Japan, November 1984.

R. Ahlswede is with the Department of Mathematics, University of Bielefeld, P.O. Box 8640, D-4800 Bielefeld 1, West Germany.

A. H. Kaspi is with the Department of Electrical Engineering, Technion, Israel Institute of Technology, P.O. Box 2250, Haifa, Israel 32000.

IEEE Log Number 8611651.

0018-9448/87/0500-0310$01.00 ©1987 IEEE

We also write $S_0|x^n \rightsquigarrow y^n$ if we are not interested in S_n. Next we have

$$\mathscr{Y}(S_0|x^n) = \mathscr{Y}(x^n) = \{y^n \in \mathscr{X}^n \colon S_0|x^n \rightsquigarrow y^n\}. \quad (2.4)$$

An (n, M) code for the PJ channel is a set

$$\mathscr{U} \subset \mathscr{X}^n, |\mathscr{U}| = M, \text{ with the property}$$

$$\mathscr{Y}(u) \cap \mathscr{Y}(u') = \varnothing, \qquad \text{for all } u, u' \in \mathscr{U} \text{ with } u \neq u'. \quad (2.5)$$

Finally,

$$M(n) = \text{maximal } M \text{ for which an } (n, M) \text{ code exists} \quad (2.6)$$

$$C_J(\alpha, \beta, S_0) = \lim_{n \to \infty} \frac{1}{n} \log M(n)$$

is the capacity of the PJ channel with parameters α, β and initial state S_0. (2.7)

Our first result is the following.

Proposition 1: For all $\alpha \geq 2$, $\beta \geq 1$, and S_0:

$$C_J(\alpha, \beta, S_0) \geq (\beta + 1)^{-1} \log \alpha.$$

Proof: For a sequence $v^l = (v_1, \cdots, v_l) \in \mathscr{X}^l$ we define

$$|v^l|_i = |\{t \colon v_t = i, 1 \leq t \leq l\}|, \qquad i \in \mathscr{X}. \quad (2.8)$$

For $n = (\beta + 1)^m$ we choose the set $\mathscr{U}$ of words of length n which are obtained by an m-fold concatenation of words of length $\beta + 1$ where each word uses one letter only, that is,

$$\mathscr{U} = \{11 \cdots 1, 22 \cdots 2, \cdots, \alpha\alpha \cdots \alpha\}^m, \quad (2.9)$$

and verify that $\mathscr{U}$ is an (n, α^m) code. Otherwise, we have for some distinct $a^n, a'^n \in \mathscr{U}$ and $x^n \in \mathscr{X}^n$,

$$S_0|a^n \rightsquigarrow x^n \text{ and } S_0|a'^n \rightsquigarrow x^n. \quad (2.10)$$

Now let t be the smallest integer such that

$$a_{(t-1)(\beta+1)+1} = r \neq a'_{(t-1)(\beta+1)+1}. \quad (2.11)$$

Then also for $l = t \cdot (\beta + 1)$,

$$S_0|a^l \rightsquigarrow x^l|S_l \text{ and } S_0|a'^l \rightsquigarrow x^l|S'_l. \quad (2.12)$$

We derive a contradiction as follows. From (2.12) for all $i \in \mathscr{X}$

$$|a^l|_i + s_0(i) = |x^l|_i + s_l(i) \quad (2.13)$$

$$|a'^l|_i + s_0(i) = |x^l|_i + s'_l(i). \quad (2.14)$$

Furthermore, by definition of l and r

$$|a^l|_r = |a'^l|_r + (\beta + 1). \quad (2.15)$$

Equations (2.13)–(2.15) for $i = r$ imply that $(\beta + 1) = s_l(r) - s'_l(r)$, which contradicts $0 \leq s_l(r)$, $s'_l(r) \leq \beta$. Therefore, $\mathscr{U}$ is an (n, α^m) code and

$$N((\beta + 1)m) \geq \alpha^m, \qquad \text{for all } m = 1, 2, \cdots.$$

The result follows by definition (2.7).

Remark: Inspection of the preceding arguments shows that we have actually established somewhat more, namely, the following result. If for an arbitrary $\mathscr{U} \subset \mathscr{X}^n$ the property

$$(*)\; ||u|_v - |u'|_v| \geq \beta + 1, \qquad \text{for all distinct } u, u' \in \mathscr{U}$$

holds, then $\mathscr{U}$ is a code. Without any reference to coding theory one can study in the cases $\alpha \geq 2$ and $\beta \geq 1$ the combinatorial problem of the maximum cardinality of a $\mathscr{U} \subset \mathscr{X}^n$ satisfying $(*)$. Next we prove that the bound in Proposition 1 is optimal for $\alpha = 2$.

Theorem 1: For $\alpha = 2$, $\beta \geq 1$, and all S_0,

$$C_J(2, \beta, S_0) = (\beta + 1)^{-1}.$$

First we establish an auxiliary result in greater generality than needed to prove

$$C_J(2, \beta, S_0) \leq (\beta + 1)^{-1}. \quad (2.16)$$

Lemma 1: For $\alpha \geq 2$, $\beta \geq 1$, and arbitrary S_0, if

$$S_0|a_1 \cdots a_n \rightsquigarrow c = c_1 \cdots c_n \text{ and}$$

$$S_0|a_1 \cdots a_n \rightsquigarrow c' = c'_1 \cdots c'_n,$$

then an $x = x_1 \cdots x_n$ exists with the properties

$$S_0|c \rightsquigarrow x \text{ and } S_0|c' \rightsquigarrow x.$$

Proof (induction in n): For $n = 1$, $S_0|i \rightsquigarrow k$ and $S_0|i \rightsquigarrow l$ imply that for any j with $s_0(j) \geq 1$, $S_0|k \rightsquigarrow j$ and $S_0|l \rightsquigarrow j$. For $n - 1 \to n$, the assumption for n implies

$$S_0|a^{n-1} \rightsquigarrow c^{n-1}|S_{n-1} \qquad S_0|a^{n-1} \rightsquigarrow c'^{n-1}|S'_{n-1} \quad (2.17)$$

and by induction hypothesis

$$\exists x^{n-1} \colon S_0|c^{n-1} \rightsquigarrow x^{n-1}|\Sigma_{n-1}$$

and

$$S_0|c'^{n-1} \rightsquigarrow x^{n-1}|\Sigma'_{n-1}. \quad (2.18)$$

The assumption for n gives

$$s_0(i) + |a^n|_i = |c^n|_i + s_n(i), \qquad i \in \mathscr{X} \quad (2.19)$$

$$s_0(i) + |a^n|_i = |c'^n|_i + s'_n(i), \qquad i \in \mathscr{X} \quad (2.20)$$

and from (2.18) in *analogous notation*

$$s_0(i) + |c^{n-1}|_i = |x^{n-1}|_i + \sigma_{n-1}(i), \qquad i \in \mathscr{X} \quad (2.21)$$

$$s_0(i) + |c'^{n-1}|_i = |x^{n-1}|_i + \sigma'_{n-1}(i), \qquad i \in \mathscr{X}. \quad (2.22)$$

Now (2.19) and (2.20) yield

$$s_n(i) + |c^{n-1}|_i + |c_n|_i = s'_n(i) + |c'^{n-1}|_i + |c'_n|_i.$$

Using (2.21) and (2.22), we can eliminate c^{n-1} and c'^{n-1}. Thus

$$s_n(i) + \left[\left(|x^{n-1}|_i - s_0(i)\right) + \sigma_{n-1}(i)\right] + |c_n|_i$$
$$= s'_n(i) + \left[\left(|x^{n-1}|_i - s_0(i)\right) + \sigma'_{n-1}(i)\right] + |c'_n|_i$$

which yields for all $i \in \mathscr{X}$

$$s_n(i) + \sigma_{n-1}(i) + |c_n|_i = s'_n(i) + \sigma'_{n-1}(i) + |c'_n|_i. \quad (2.23)$$

We now show that (2.23) implies

$$(\Sigma_{n-1} \cup \{c_n\}) \cap (\Sigma'_{n-1} \cup \{c'_n\}) \neq \varnothing. \quad (2.24)$$

This will complete the proof because we can choose for x_n any element in the intersection.

More abstractly, we can view the situation as follows. We are given two matrices:

$$(A_{ij})_{\substack{1 \le i \le \alpha \\ 1 \le j \le 3}} \qquad (B_{ij})_{\substack{1 \le i \le \alpha \\ 1 \le j \le 3}}$$

with nonnegative integers as entries and the properties

$$\text{a)} \quad \sum_j A_{ij} = \sum_j B_{ij}, \qquad \text{for all } 1 \le i \le \alpha$$

$$\text{b)} \quad \sum_i A_{i1} = \sum_i A_{i2} = \sum_i B_{i1} = \sum_i B_{i2} = \beta$$

$$\text{c)} \quad \sum_i A_{i3} = \sum_i B_{i3} = 1.$$

We have used (2.23) for a) and the properties for the quantities in (2.23) for b) and c). From a)–c) we now derive

$$\text{d)} \quad \exists \text{ row } i: A_{i2} + A_{i3} \ge 1 \qquad B_{i2} + B_{i3} \ge 1.$$

Assume that d) does not hold. Then two distinct indices i_1 and i_2 exist with $A_{i_1 3} = 1$ and $B_{i_2 3} = 1$. Define now $I = \{i\colon\ B_{i2} > 0\} \cup \{i_2\}$. Necessarily, $A_{i2} = 0$ for $i \in I$ and $i_1 \notin I$.

Now

$$\sum_{i \in I;\, l=1,2,3} B_{il} \ge \sum_{i \in I} B_{i2} + \sum_{i \in I} B_{i3} = \beta + 1.$$

However,

$$\sum_{i \in I;\, l=1,2,3} A_{il} = \sum_{i \in I} A_{il} \le \beta,$$

contradicting a). With the identifications

$$A_{i1} = s_n(i) \qquad A_{i2} = \sigma_{n-1}(i) \qquad A_{i3} = |c_n|_i$$
$$B_{i1} = s'_n(i) \qquad B_{i2} = \sigma'_{n-1}(i) \qquad B_{i3} = |c'_n|_i$$

we see that d) implies (2.24). This completes the proof of Lemma 1.

It is convenient to give an interpretation of Lemma 1 and the maximal code length $M(n)$ in terms of a graph G_n which has vertex set $\mathcal{X}^n$ and the following adjacency structure: $u, u' \in \mathcal{X}^n$ are adjacent iff a $v \in \mathcal{X}^n$ exists with $S_0|u \rightsquigarrow v$ and $S_0|u' \rightsquigarrow v$. Lemma 1 says that the sets $\mathcal{Y}(u)$, $u \in \mathcal{X}^n$, are *cliques*, i.e., complete subgraphs of G_n. Clearly, $M(n)$ is the independence number of G_n. Furthermore, clique (n), $\triangleq$ minimal number of cliques needed to cover G_n, satisfies

$$\text{clique}\,(n) \ge M(n). \tag{2.25}$$

Notice that the *desired inequality* (2.16) is an immediate consequence of (2.25) and the following result.

Lemma 2: For $\alpha = 2$, $\beta \ge 1$, and S_0 arbitrary,

$$\text{clique}\,((\beta+1)t) \le 2^t.$$

Proof: By the foregoing explanations it suffices to find a $\mathcal{U}_n \subset \mathcal{X}^n$, $n = (\beta+1)t$, such that

$$|\mathcal{U}_n| = 2^t \text{ and } \bigcup_{u \in \mathcal{U}_n} \mathcal{Y}(u) = \mathcal{X}^n. \tag{2.26}$$

With $\mathcal{V} = \{11 \cdots 1, 22 \cdots 2\} \subset \mathcal{X}^{\beta+1}$, define for $t = 1, 2, \cdots$

$$\mathcal{U}_n \triangleq \mathcal{V}^t.$$

Now for any $S_0 = (s_0(1), s_0(2))$, $\Sigma_{i=1}^2 s_0(i) = \beta$, and any word $x^{\beta+1} \in \mathcal{X}^{\beta+1}$,

$$|x^{\beta+1}|_1 + |x^{\beta+1}|_2 = \beta + 1 \qquad \text{and } s_0(1) + s_0(2) = \beta$$

imply that either

$$|x^{\beta+1}|_1 \le s_0(1) \qquad \text{or } |x^{\beta+1}|_2 \le s_0(2).$$

In the first case, clearly, $S_0|22 \cdots 2 \rightsquigarrow x^{\beta+1}$, and in the second case $S_0|11 \cdots 1 \rightsquigarrow x^{\beta+1}$. This settles the first step of the induction beginning with $t = 1$. Since $\mathcal{U}_n = \mathcal{V}^{t-1} * \mathcal{V}$, and in the foregoing argument S_0 *was arbitrary*, the induction step proceeds in the same way. Q.E.D.

III. Optimal Coding Strategies for PR Channels

Recall definition (2.4). Now we have to find an $x^n \in \mathcal{X}^n$ such that $|\mathcal{Y}(S_0|x^n)|$ is maximal! Our *guiding idea* is to use a greedy approach, which means here the following stepwise optimization.

The x^n maximizing $|\mathcal{Y}(S_0|x^n)|$ is guessed at by extending an x^{n-1} maximizing $|\mathcal{Y}(S_0|x^{n-1})|$. This suggests that *periodic* sequences x^n might be optimal (see Lemma 5, to follow). Lemmas 3 and 4 show that this is indeed the case.

We consider the case $\alpha \ge 2$, $\beta = 1$. States can here be described by a single function $s_n (n = 0, 1, 2, \cdots)$, taking values in $\mathcal{X}$. We study the situation where the initial state S_0 is known to all participants I, $\mathcal{O}$, and D. Therefore, we can assume without loss of generality that $s_0 = 1$ *always*. For minor technical reasons we denote the maximal codelength for block-length n by $N(n+1)$. Thus

$$N(n+1) = \max_{x^n \in \mathcal{X}^n} |\mathcal{Y}(x^n)|, \qquad n \in \mathbf{N}. \tag{3.1}$$

Furthermore, we make the conventions

$$N(1) = N(0) = 1 \qquad N(-n) = 0, \qquad \text{for } n \in \mathbf{N}. \tag{3.2}$$

A few more definitions are needed:

$$\mathcal{Y}_i(x^n) \triangleq \{y^n \in \mathcal{Y}(x^n)\colon s_n = i\}, \qquad i \in \mathcal{X}. \tag{3.3}$$

Clearly,

$$\mathcal{Y}(x^n) = \bigcup_{i \in \mathcal{X}} \mathcal{Y}_i(x^n) \tag{3.4}$$

$$\mathcal{P}_r(\mathcal{X}) \triangleq \{E \subset \mathcal{X}\colon |E| = r\} \tag{3.5}$$

$$N(n+1, r) \triangleq \max_{x^n \in \mathcal{X}^n} \max_{I \in \mathcal{P}_r(\mathcal{X})} \sum_{i \in I} |\mathcal{Y}_i(x^n)| \tag{3.6}$$

for $1 \le r \le \alpha$, and by convention $N(n+1, 0) \triangleq 0$.

Lemma 3: We have

$$N(n+1, r) \le N(n) + N(n-1) + \cdots + N(n+1-r)$$

for all $n \in \mathbf{N}$ and all $r = 1, 2, \cdots, \alpha - 1$.

Proof: First we consider the case $\alpha - 1 \ge r > n + 1$ and show that here

$$N(n+1, r) = N(n) + \cdots + N(n-n). \tag{3.7}$$

For this, just notice that at most $n + 1 < r$ of the $\mathcal{Y}_i(x^n)$ are nonempty and, therefore, that

$$N(n+1, r) = N(n+1).$$

Furthermore, obviously $N(n+1) \le 2^n$, and by choosing an x^n with distinct letters different from the initial state $s_0 = 1$, we see that actually $N(n+1) = 2^n$. For the same reasons,

$$N(k) = 2^{k-1} \qquad \text{for } 2 \le k \le n+1$$

and since $N(1) = N(0) = 1$, (3.7) holds because

$$2^n = \sum_{l=0}^{n-1} 2^l + 1.$$

From (3.7) and the fact that $N(-1) + \cdots + N(n+1-r) = 0$, we get for $\alpha - 1 \ge r > n+1$

$$N(n+1, r) = N(n) + \cdots + N(-1) + \cdots + N(n+1-r) \qquad (3.8)$$

so that the desired inequality holds in this case.

We settle the remaining case $1 \le r \le \min(\alpha - 1, n+1)$ by induction in n. For $n = 1$,

$$r = 1\colon\ N(2,1) = 1 = N(1)$$
$$r = 2\colon\ N(2,2) = 2 = N(1) + N(0).$$

For $n - 1 \to n$, consider (x^n, I) with $|I| = r$ and estimate $\sum_{i \in I} |\mathcal{Y}_i(x^n)|$ from above. We always have

$$\mathcal{Y}_i(x^n) = \begin{cases} \mathcal{Y}_i(x^{n-1}) * x_n, & x_n \ne i \\ \bigcup_{k \in \mathcal{X}} \mathcal{Y}_k(x^{n-1}) * k, & x_n = i \end{cases}. \qquad (3.9)$$

For $x_n \notin I$, by (3.9),

$$\sum_{i \in I} |\mathcal{Y}_i(x^n)| = \sum_{i \in I} |\mathcal{Y}_i(x^{n-1})| \le \sum_{i \in \mathcal{X}} |\mathcal{Y}_i(x^{n-1})|$$
$$= |\mathcal{Y}(x^{n-1})| \le N(n).$$

For $X_n \in I$, again by (3.9),

$$\sum_{i \in I} |\mathcal{Y}_i(x^n)| = \Big|\bigcup_{k \in \mathcal{X}} \mathcal{Y}_k(x^{n-1}) * k\Big| + \sum_{i \in I - \{x_n\}} |\mathcal{Y}_i(x^{n-1}) * x^n|$$
$$\le N(n) + N(n, r-1).$$

Therefore, in both cases

$$\sum_{i \in I} |\mathcal{Y}_i(x^n)| \le N(n) + N(n, r-1)$$

and thus

$$N(n+1, r) \le N(n) + N(n, r-1). \qquad (3.10)$$

For $r > 1$ by induction hypothesis,

$$N(n+1, r) \le N(n) + (N(n-1) + \cdots + N(n+1-r))$$

and for $r = 1$ also from (3.10),

$$N(n+1, 1) \le N(n) + N(n, 0) = N(n).$$

Lemma 4: $N(n+1) \le N(n, \alpha - 1) + N(n)$.

Proof: We have the following:

$$|\mathcal{Y}(x^n)| = \sum_{j \in \mathcal{X}} |\mathcal{Y}_j(x^n)|$$
$$= \sum_{j \ne x_n} |\mathcal{Y}_j(x^n)| + |\mathcal{Y}_{x_n}(x^n)|$$
$$= \sum_{j \ne x_n} |\mathcal{Y}_j(x^{n-1})| + |\mathcal{Y}(x^{n-1})| \qquad \text{(by (3.9))}$$
$$\le N(n, \alpha - 1) + N(n).$$

Theorem 2:[1] Let $\alpha \ge 2$ and $\beta = 1$. Then for block-length $n - 1$ the maximal codelength $N(n)$ for the PR channel satisfies, for $n = 1, 2, \cdots$, the recursion

$$\text{(F)} \quad N(n) = N(n-1) + N(n-2) + \cdots + N(n - \alpha)$$

with the convention $N(1) = N(0) = 1$ and $N(-m) = 0$ for $m \in \mathbb{N}$.

Proof: Notice that for $n \ge 2$ by Lemma 4,

$$N(n) \le N(n-1, \alpha - 1) + N(n-1).$$

This and Lemma 3 yield

$$N(n) \le N(n-1) + N(n-2) + \cdots + N(n-\alpha). \qquad (3.11)$$

This inequality holds also for $n = 1$.

We prove equality in (3.11) by explicitly giving a coding strategy leading to codelengths which satisfy the same recurrence relation as claimed for N in (F). The result is the following lemma.

Lemma 5: For $s_0 = 1$ let $s_0 z^\infty = s_0 2\ 3 \cdots \alpha 1\ 2\ 3 \cdots$ be a periodic sequence. Then with $f(0) = f(I) = 1$, $f(-m) = 0$, and $f(n+1) \triangleq |\mathcal{Y}(z^n)|$ we have

$$f(n+1) = f(n) + \cdots + f(n - \alpha), \qquad \text{for all } n = 1, 2, \cdots.$$

Proof: Again by (3.9),

$$f(n+1) = |\mathcal{Y}(z^n)| = \sum_{j \ne z_n} |\mathcal{Y}_j(z^n)| + |\mathcal{Y}(z^{n-1})|$$
$$= \sum_{j \ne z_n} |\mathcal{Y}_j(z^n)| + f(n).$$

Since $z_{n-1} \ne z_n$, we have

$$\mathcal{Y}_{z_{n-1}}(z^n) = \mathcal{Y}_{z_{n-1}}(z^{n-1}) * z_n$$

and thus

$$|\mathcal{Y}_{z_{n-1}}(z^n)| = |\mathcal{Y}_{z_{n-1}}(z^{n-1})|$$
$$\text{(by (3.9))} = \Big|\bigcup_{j \in \mathcal{X}} \mathcal{Y}_j(z^{n-2})\Big| = |\mathcal{Y}(z^{n-2}) = f(n-1).$$

Therefore,

$$f(n+1) = \sum_{j \ne z_n, z_{n-1}} |\mathcal{Y}_j(z^n)| + f(n) + f(n-1), \qquad (3.12)$$

[1] In [7] Kobayashi extended Theorem 2 to all $\beta \ge 1$.

and repetition of this argument gives

$$|\mathscr{Y}_{z_{n-2}}(z^n)| = |\mathscr{Y}_{z_{n-2}}(z^{n-2}) * z_{n-1} * z_n| = f(n-2).$$

We thus have

$$f(n+1) = \sum_{j \neq z_n, z_{n-1}, z_{n-2}} |\mathscr{Y}_j(z^n)| + f(n) + f(n-1) + f(n-2). \quad (3.13)$$

This procedure stops after α steps, when the z_n, $z_{n-1}, \cdots, z_{n-\alpha+1}$ have exhausted $\mathscr{X}$. Q.E.D.

Asymptotic solutions of (F) are well-known (cf. [5]). For $\alpha = 2$ we get the famous Fibonacci sequence.

Corollary: $C_R(2,1) = \log(1 + \sqrt{5})/2 \approx 0.69$.

IV. Directions for Further Investigations

We formulate here channel models, which are in the spirit of those considered and, in particular, include multiuser aspects. Several open problems are stated.

1) Theorems 1 and 2 give the capacities $C_J(\alpha, \beta, S_0)$ and $C_R(\alpha, \beta, S_0)$ for certain values of α and β. What are the general solutions? For $\alpha = 2$ and $\beta = 2$ we can show that for the PR channel $N(n+2) \geq N(n+1) + 2N(n) - N(n-1)$, independent of the value of S_0. We conjecture that equality holds in this recursion. The capacity can then be found from known solutions of linear recurrence relations.

2) For the PJ channel a more robust assumption than the one used is that none of the participants knows S_0 and that S_0 can vary arbitrarily from message to message. Here the capacity of $C_J(\alpha, \beta)$, say, is smaller than $C_J(\alpha, \beta, S_0)$. Notice that $C_J(2,1,S_0)$ (respectively, $C_J(2,1)$) is the zero-error capacity of the classical trapdoor channel for a known (respectively, unknown) initial state. We guess, for instance, that $C_J(2,1) = 1/3$.

3) The PI$\mathscr{O}$ Channel: Suppose that both I and $\mathscr{O}$ try to send messages over the nonprobabilistic permuting channel to D. What is the *region* of achievable pairs of rates? Theorem 2 gives one point on the boundary of this region for the case of a known initial state and (α, β) as specified. Other problems arise if the initial state is unknown to all or some participants.

4) Life Channel: Consider the relay problem with the difference that instead of letting I choose a fixed sequence x^n, now a sequence of independent identically distributed RV's $X^n = (X_1, \cdots, X_n)$ feeds the channel. They provide the environment (or the life conditions) in which $\mathscr{O}$ tries to transmit messages to D. The initial state may be chosen by the same random process. What is the capacity $C_L(\alpha, \beta, X)$?

5) Multiple-Access Channel (MAC): Specify positive integers α, β, and γ and assume that there are now participants $I_1, \cdots, I_\gamma$ who put one ball each into the box at a time instant. Those balls are mixed with the β balls in the box, and then $\mathscr{O}$ chooses any γ balls and gives them to the decoder D. (Notice that for this nonprobabilistic permuting MAC the causality in the sense of [3] is preserved.)

The notions of PJ channels, PR channels, and PI$\mathscr{O}$ channels extend to those of PJ MAC, PR MAC, and PI$\mathscr{O}$ MAC. What are the capacity regions? A multitude of further problems arises if, for instance, $\mathscr{O}$ can send balls to *several decoders* (interference channel). In short, channel models familiar from multiuser information theory (such as the broadcast channel, wire-tape channel, etc.) can all be reformulated for nonprobabilistic permuting channels.

6) Feedback: All channels mentioned so far can also be studied if *feedback links* are present.

7) Sources: One can even formulate models for permuting multisources, where the correlation of sources is defined by restrictions on the joint outputs.

V. Concluding Remarks

The theory of nonprobabilistic permuting multiuser channels has the advantage of being purely combinatorial in nature. Error probabilities do not arise. Because of the memory in the channels, the combinatorial techniques (see the proofs of the theorems) differ from those usually encountered in information theory. A further study may give new insight into, or at least new ideas about, the harder unsolved problems in multiuser information theory.

Acknowledgment

We wish to thank T. Berger for stimulating discussions.

References

[1] D. Blackwell, L. Breiman, and J. Thomasian, "Proof of Shannon's transmission theorem for finite-state indecomposable channels," *Ann. Math. Statist.*, vol. 29, pp. 1209–1220, 1958.

[2] D. Blackwell, "Information theory," in *Modern Mathematics for the Engineer*, 2nd series, E. F. Beckenbach, Ed. New York: McGraw-Hill, 1961.

[3] T. W. Benjamin, "Coding for a noisy channel with permutation errors," Ph.D. dissertation, Cornell Univ., Ithaca, NY, 1975.

[4] R. Ahlswede, "Coloring hypergraphs: A new approach to multi-user source coding," Part I, *J. Combinatorics Inform. Syst. Sci.*, vol. 4, pp. 76–115, 1979; Part II, vol. 5, pp. 220–268, 1980.

[5] D. E. Knuth, *The Art of Computer Programming*, vol. 1. Reading, MA: Addison-Wesley, 1973.

[6] R. Ash, *Information Theory*. New York: Wiley, 1965.

[7] K. Kobayashi, "The capacity of permuting relay channels," *IEEE Trans. Inform. Theory*, to be published.