A CONSTRUCTIVE PROOF OF THE CODING THEOREM FOR DISCRETE MEMORYLESS CHANNELS WITH FEEDBACK

RUDOLF AHLSWEDE

COLUMBUS and URBANA

0. INTRODUCTION

In the present paper we study discrete memoryless channels (d.m.c.) with noiseless feedback. A d.m.c. with noiseless feedback will be abbreviated as d.m.c.f. When we talk about feedback we shall always mean noiseless feedback.

In [8] Shannon proved that feedback does not increase the capacity of a d.m.c. Kemperman [5] and Kesten (oral communication) improved on this result by showing that also the strong converse to the coding theorem holds. Even though feedback has no effect on the value of the capacity of a memoryless channel it provides new possibilities for the actual construction of codes. The first attempt in this direction was made by Horstein. In [4] he introduced a *sequential* (varying block length) coding scheme for the binary symmetric channel with feedback (b.s.c.f.). However, the scheme is fairly complicated and – what is more important – Horstein does *not* rigorously prove that for any rate below channel capacity the decoding error probability for his scheme tends to 0. It seems that no worker in the field of coding theory understands his reasoning or can give a proof. Some mathematicians believe that his method is wrong. A completely different approach was taken by Schalkwijk [6], [7] and Kailath [6]. They found for the Gaussian channel with feedback and with an energy constraint a sequential coding scheme which performs at any rate below the capacity with a *double* exponentially descreasing error probability. This coding scheme makes heavy use of some of the properties of the Gaussian channel and nobody has succeeded in carrying the basic idea over to the d.m.c.f. – perhaps, because it is impossible. The result stands as an isolated "break-through".

It is clear from what we said earlier that the coding theorem for the d.m.c.f. is an immediate consequence of the coding theorem for the d.m.c.

The known proofs of the coding theorem for d.m.c. use either a random coding method (Shannon [7]) or a maximal coding method (Feinstein [3], Wolfowitz [10]). The presence of feedback enables us to give a new proof of the coding theorem for *block* codes, which is *not* based on random coding or maximal coding ideas.

Our proof has two parts. In the first step we reduce the set of all messages to subset of suitable size. This is made possible by the elementary lemmas 1, 2, 3 section 1. Here we use the idea of "generated sequences" (see [10]), but there e also other ways to obtain the reduction. In this first step we do not use feed-ck. The second step consists of an iteration of the earlier procedure, we iterate til we come up with the message sent. The iteration is possible because we have dback.

In section 3 we give an alternate scheme for the b.s.c.f. That scheme differs from one outlined above in that it makes use of feedback already in the first step. e second step is the same as before.

Our approach provides the following advantages:

1. For every block length we give an explicit coding scheme, which can easily implemented. The maximal coding method yields a code construction only for d block length. If one changed the block length one would have to repeat the nstruction. The same difficulty arises if we select codes at random according to the dom coding method.

2. Our approach turns out to be very useful in solving coding problems for more nplex channels with feedback. For the channels treated in [1], for instance, dom coding and maximal coding methods seem to fail.

The extension of our result to the case of infinite alphabets is straight forward.

1. DEFINITIONS AND AUXILIARY RESULTS

et $X = \{1, \ldots, a\}$ and $Y = \{1, \ldots, b\}$ be finite sets, which serve as input and put alphabets of the channel described below. Write $X^t = X$ and $Y^t = Y$ for 1, 2, ... By $X_n = \prod_{t=1}^{n} X^t$ denote the set of input n-sequences (words of length n) by $Y_n = \prod_{t=1}^{n} Y^t$ denote the set of output n-sequences.

et $w(\cdot|\cdot)$ be an $a \times b$ -stochastic matrix, that is,

$$1 \geqq w(j \mid i) \geqq 0 \quad \text{for} \quad i \in X, j \in Y$$

$$\sum_{j=1}^{b} w(j \mid i) = 1 \quad \text{for} \quad i \in X .$$

he transmission probabilities of a discrete memoryless channel (d.m.c) are

defined by

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

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

We introduce now a d.m.c. with feedback (d.m.c.f.). By this is meant that there exists a return channel which sends back from the receiving point to the transmitting point the element of Y actually received. It is assumed that this information is received at the transmitting point before the next letter is sent, and can therefore be used for choosing the next letter to be sent.

A code (n, N, λ) for this channel is described as follows:

There is given a finite set of messages $M = \{1, \ldots, N\}$, one of which will be presented to the sender for transmission. Message $m \in M$ is encoded by an encoding (vector valued) function

$$f_n(m) = [f_m^1, f_m^2(Z^1), \ldots, f_m^t(Z^1, \ldots, Z^{t-1}), \ldots, f_m^n(Z^1, \ldots, Z^{n-1})], \tag{1.3}$$

where f_m^t is defined on Y^{t-1} for $t > 1$ and takes values in X^t, and $Z^1, Z^2, \ldots, Z^{t-1}$ are the chance received elements of Y (known to the sender before he sends $f_m^t(Z^1, \ldots, Z^{t-1})$); f_m^1 is an element of X^1. The distribution of the random variables Z^t $(t = 1, \ldots, n)$ is determined by $f_m^1, \ldots, f_m^{t-1}$, and $w(\cdot \mid \cdot)$. We denote the probability of receiving $y_n \in Y_n$, if m is thus encoded, by $P(y_n \mid f_n(m))$.

A code (n, N, λ) for the d.m.c.f. is a system

$$\{(f_n(m), A_m) \mid m = 1, \ldots, N\}, \tag{1.4}$$

where the $f_n(m)$ are as defined in (1.3), $A_m \subset Y_n$ for $m = 1, \ldots, N$; $A_m \cap A_{m'} = \emptyset$ for $m \neq m'$, and $P(A_m \mid f_n(m)) \geqq 1 - \lambda$ for $m = 1, \ldots, N$.

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

$$H(p) = -\sum_{i=1}^{c} p_i \log p_i . \tag{1.5}$$

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

$$R(\pi, w(\cdot \mid \cdot)) = H(q) - \sum_i \pi_i H(w(\cdot \mid \cdot)), \tag{1.6}$$

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

For π and $w(\cdot \mid \cdot)$ define a $b \times a$-stochastic matrix $w^*(\cdot \mid \cdot)$ by

$$w^*(i \mid j) = \frac{\pi_i \, w(j \mid i)}{q_j}, \quad j = 1, \ldots, b; \; i = 1, \ldots, a . \tag{1.7}$$

It is well known and easy to verify that

$$R(\pi, w(\cdot|\cdot)) = H(\pi) - \sum_j q_j H(w^*(\cdot|j)) . \tag{1.8}$$

The capacity C of our channel is given by

$$C = \max_{\pi} R(\pi, w(\cdot|\cdot)) . \tag{1.9}$$

For $u \in X_l$ define $N(i \mid u)$ as the number which counts how often i occurs as a component of u. Similarily define $N(j \mid v)$ for $v \in Y_l$. $N(i, j \mid u, v)$ shall count the number of components in which u has an i and v has a j; $i = 1, \ldots, a$; $j = 1, \ldots, b$. For a probability distribution π on X define the set $X_l(\pi)$ by

$$X_l(\pi) = \{x_l \mid x_l \in X_l, |\pi_i l - N(i \mid x_l)| \leqq 1 \text{ for } i = 1, \ldots, a\} . \tag{1.10}$$

LEMMA 1.

a) $$|X_l(\pi)| = \exp\{H(\pi)\, l + 0(\log l)\} ,$$

b) $$|X_l(\pi)| \geqq \exp\{H(\pi)\, l - f(a, \pi) \log l\} \quad \textit{for} \quad l = 1, 2, \ldots$$

$f(a, \pi)$ can be given explicitly.

This Lemma follows immediately from definition (1.10) and Stirling's formula. For $u \in X_l(\pi)$ and $\varepsilon > 0$ define $Y_l(u, \varepsilon, \pi, w)$ by

$$Y_l(u, \varepsilon, \pi, w) = \{v \mid v \in Y_l, |N(i, j \mid u, v) - w(j \mid i) N(i \mid u)| \leqq \varepsilon l \text{ for } i = 1, \ldots, a; j = 1, \ldots, b\} . \tag{1.11}$$

LEMMA 2. *For $u \in X_l(\pi)$; $l = 1, 2, \ldots$:*

$$P(Y_l(u, \varepsilon, \pi, w) \mid u) \geqq 1 - e^{-E(\varepsilon, \pi, w) l} ,$$

where $E(\varepsilon, \pi, w)$ is positive and can be given explicitly.

This Lemma can easily be verified by using Chebyshev's inequality. Define $Y_l(\varepsilon, \pi, w)$ by

$$Y_l(\varepsilon, \pi, w) = \bigcup_{u \in X_l(\pi)} Y_l(u, \varepsilon, \pi, w) . \tag{1.12}$$

Finally, define for a $v \in Y_l$ a probability distribution q^* on Y by

$$q_j^* = N(j \mid v)\, l^{-1} \quad \text{for} \quad j = 1, \ldots, b \tag{1.13}$$

and a set $X_l(v, \varepsilon, \pi, w)$ by

$$X_l(v, \varepsilon, \pi, w) = \{u \mid u \in X_l(\pi), |N(i, j \mid u, v) - w^*(i \mid j) N(j \mid v)| \leqq \leqq (a + 1)\, \varepsilon l \text{ for } j = 1, \ldots, b; i = 1, \ldots, a\} . \tag{1.14}$$

LEMMA 3.

$$|X_l(v, \varepsilon, \pi, w)| \leqq \exp\{\sum_j q_j^* H(w^*(\cdot \mid j))\, l + g(\varepsilon)\, l\},$$

where $\lim_{\varepsilon \to 0} g(\varepsilon) = 0$ *and* $g(\varepsilon)$ *is a known function of* ε.

The Lemma follows from (1.14) and Chebyshev's inequality.

Definitions (1.10), (1.11) and (1.12) were used (in a slightly different form) in [10].

2. DESCRIPTION OF OUR CODING SCHEME FOR THE D.M.C.F. AND PROOF OF THE CODING THEOREM

Let l be a positive integer and let $M_1 = \{1, 2, \ldots, a^l\}$ be a set of $N = a^l$ messages. Choose π such that $R(\pi, w) = C$ and let l_1 be the smallest integer such that $|X_{l_1}(\pi)| \geqq a^l$. It follows from Lemma 1 that

$$l_1 = \frac{\log a}{H(\pi)} l + h(l), \tag{2.1}$$

where $h(l)$ can be given explicitly and $h(l) = O(\log l)$. We map now M_1 one to one into $X_{l_1}(\pi)$ and call the image $\overline{X}_{l_1}(\pi)$. Let $u^* = (f_m^1, \ldots, f_m^{l_1})$ be the image of m, $m \in M_1$.

For $m \in M_1$ and $t = 1, \ldots, l_1$ we define now $f_m^t(Z^1, \ldots, Z^{t-1})$ by

$$f_m^t(Z^1, \ldots, Z^{t-1}) = f_m^t. \tag{2.2}$$

Suppose the sender is sending message m and he has sent already the letters $f_m^1, \ldots, f_m^{l_1}$. The receiver has received a sequence $v = (v^1, \ldots, v^{l_1}) \in Y_{l_1}$, which is known to the sender, because we have a channel with feedback. Lemma 2 implies that the probability λ_1 that v is not contained in $Y_{l_1}(u^*, \varepsilon, \pi, w)$ satisfies

$$\lambda_1 \leqq e^{-E(\varepsilon, \pi, w) l_1}. \tag{2.3}$$

v is therefore contained in $Y_{l_1}(\varepsilon, \pi, w)$ with a probability larger than $1 - \lambda_1$. The set $Y_{l_1}(\varepsilon, \pi, w)$ is known to the sender and to the receiver. If $v \notin Y_{l_1}(\varepsilon, \pi, w)$, we count this as a decoding error. In this case it is irrelevant how the sender continuous the transmission (over the fixed block length). Let us assume now that v is contained in $Y_{l_1}(\varepsilon, \pi, w)$ and define $X_{l_1}(\varepsilon, \pi, w)$ as in (1.14) and q^* as in (1.13). v is actually contained in $Y_{l_1}(u^*, \varepsilon, \pi, w)$ with a probability greater than $1 - \lambda_1$. For v in $Y_{l_1}(u^*, \varepsilon, \pi, w)$ we have by (1.11)

$$|N(i, j \mid u^*, v) - w(j \mid i)\, N(i \mid u^*)| \leqq \varepsilon l_1 \quad \text{for} \quad i = 1, \ldots, a;\ j = 1, \ldots, b. \tag{2.4}$$

Using the definition of $\overline{X}_{l_1}(\pi)$ we obtain from (2.4)

$$|N(i,j \mid u^*, v) - w(j \mid i)\,\pi_i l_1| \leqq \varepsilon l_1 \quad \text{for} \quad i = 1, \ldots, a;\ j = 1, \ldots, b\,, \tag{2.5}$$

and

$$|N(i,j \mid u^*, v) - w^*(i \mid j)\,q_j l_1| \leqq \varepsilon l_1 \quad \text{for} \quad i = 1, \ldots, a;\ j = 1, \ldots, b\,. \tag{2.6}$$

Since $N(j \mid v) = \sum_{i=1}^{a} N(i,j \mid u^*, v)$ we obtain from (2.5) that

$$|N(j \mid v) - q_j l_1| \leqq a\varepsilon l_1 \quad \text{for} \quad j = 1, \ldots, b\,. \tag{2.7}$$

(2.7) and the definition of q^* imply that

$$|q_j^* - q_j| \leqq a\varepsilon \quad \text{for} \quad j = 1, \ldots, b\,. \tag{2.8}$$

It follows from (2.8) and (2.6) that

$$|N(i,j \mid u^*, v) - w^*(i \mid j)\,q_j^* l_1| \leqq (a+1)\,\varepsilon l_1 \quad \text{for} \quad i = 1, \ldots, a;\ j = 1, \ldots, b\,. \tag{2.9}$$

This and definition (1.14) imply that u^* is contained in $X_{l_1}(v, \varepsilon, \pi, w)$. Since v is contained in $Y_{l_1}(u^*, \varepsilon, \pi, w)$ with a probability greater than $1 - \lambda_1$, u^* is contained in $X_{l_1}(v, \varepsilon, \pi, w)$ with a probability greater than $1 - \lambda_1$.

Define now M_2 by

$$M_2 = X_{l_1}(v, \varepsilon, \pi, w)\,. \tag{2.10}$$

Since v is known to sender and receiver, M_2 is also known to them. If u^* is not in M_2, we count this as a decoding error. The sender may then continue the transmission over the fixed block length in any way he wants. If u^* is in M_2 we have reduced the number N of possible messages to a number $|M_2|$ of possible messages. We give now an upper bound on $|M_2|$.

Lemma 3 and (2.8) yield that

$$|M_2| \leqq \exp\{\sum_{j=1}^{b} q_j\, H(w^*(\cdot \mid j))\, l_1 + f(\varepsilon)\, l_1\}\,, \tag{2.11}$$

where $f(\varepsilon)$ is a known function and $\lim_{\varepsilon \to 0} f(\varepsilon) = 0$. Abbreviate $H(\pi)$ as H and $\sum_{j=1}^{b} q_j\, H(w^*(\cdot \mid j))$ as $\overline{H}$. We iterate now our procedure. Let l_2 be the smallest integer such that $|X_{l_2}(\pi)| \geqq |M_2|$.

It follows from Lemma 1 and (2.11) that one can give explicitly a function $\bar{f}(\varepsilon)$, $\lim_{\varepsilon \to 0} \bar{f}(\varepsilon) = 0$, such that

$$l_2 \leqq \frac{\overline{H}}{H}\, l_1 + \frac{\bar{f}(\varepsilon)}{H}\, l_1\,. \tag{2.12}$$

We map now M_2 one to one into $X_{l_2}(\pi)$ and call the image $\bar{X}_{l_2}(\pi)$. Let $(f_m^{l_1+1}, \ldots, f_m^{l_1+l_2})$ be the image of $(f_m^1, \ldots, f_m^{l_1}) \in M_2$. For $m \in M$ and $t = l_1 + 1, \ldots, l_1 + l_2$ we define now $f_m^t(Z^1, \ldots, Z^{t-1})$ by

$$f_m^t(Z^1, \ldots, Z^{t-1}) = f_m^t . \tag{2.13}$$

We apply now the same procedure, which we applied above to the set M_1, to the set M_2. After l_2 letters have been sent we come up with a set M_3, defined analogously to M_2. The image of m is contained in M_3 with a probability $1 - \lambda_2 \geqq \geqq 1 - e^{-E(\varepsilon,\pi,w)l_2}$. Since $0 \leqq \bar{H}/H < 1$, we constantly reduce the number of messages by iterating the procedure. However, since $l_1 > l_2 > l_3 > \ldots$, the decoding errors $\lambda_1, \lambda_2, \ldots$ are increasing. We iterate the procedure only $d(l) = c \log l$ times, where c is a constant to be choosen later. For the remaining steps we need only a few — relatively to l — letters, because the l_s's decrease quickly. We achieve small error probabilities for the steps s, $s > c \log l$, by *repetition* of these steps.

The decoding error probability after $d = d(l) = c \log l$ iterations of the procedure is bounded by $\sum_{s=1}^{d} \lambda_s$, which is smaller than $d \cdot \exp\{-E(\varepsilon, \pi, w)(K(\varepsilon))^{d-1} l_1\}$, if we set

$$K(\varepsilon) = \frac{\bar{H}}{H} + \frac{\bar{f}(\varepsilon)}{H} .$$

By choosing $d(l) = \frac{1}{2} \log (K(\varepsilon))^{-1} \cdot \log l$, we obtain that

$$(K(\varepsilon))^{d(l)} \cdot l = l^{1/2} \tag{2.14}$$

and that

$$\sum_{s=1}^{d} \lambda_i \leqq \exp\{-\tfrac{1}{2} E(\varepsilon, \pi, w)\, l^{1/2}\} \tag{2.15}$$

for $l \geqq l^*(\varepsilon, \pi, w)$, a known function.

Let now D be the smallest integer such that

$$a^{l_D} \leqq a^{(K(\varepsilon))^{D-1} l_1} < 2 . \tag{2.16}$$

Obviously, D is an upper bound on the number of steps needed and satisfies

$$D \leqq f^*(\varepsilon) \log l , \tag{2.17}$$

where $f^*(\varepsilon)$ is a known function.

From the fact that for small ε, $0 \leqq K(\varepsilon) < 1$ and from (2.14) we conclude that

$$l_s \leqq l^{1/2} \quad \text{for} \quad s = d(l), \ldots, D . \tag{2.18}$$

For every s between d and D we repeat the *same* procedure $[l^{1/4}]$ times. The total amount of letters needed is less than $l^{1/2}\, l^{1/4} f^*(\varepsilon) \log l$.

To be more specific, let us assume that at instance $s = d$ we are dealing with the set M_{d+1}. After l_{d+1} letters have been sent we come up with the set $M_{d+2} = M_{d+2}(1)$. Now we repeat the same procedure of sending the l_{d+1} letters

$$f_m^{l_1+\ldots+l_d+1}, \ldots, f_m^{l_1+\ldots+l_d+l_{d+1}}$$

$[l^{1/4}]$ times. We thus obtain sets

$$M_{d+2}(r)\,; \quad r = 1, \ldots, [l^{1/4}]\,. \tag{2.19}$$

Define now $\overline{M}_{d+2}$ by

$$\overline{M}_{d+2} = \{u \mid u \in \overline{X}_{l_{d+1}}(\pi),\ u \in M_{d+2}(r) \text{ for more than } \tfrac{1}{2}[l^{1/4}] \text{ of the } r\text{'s}\}\,. \tag{2.20}$$

Obviously,

$$|\overline{M}_{d+2}| \leqq 2 \max_r M_{d+2}(r)\,. \tag{2.21}$$

$$u^* = (f_m^{l_1+\ldots+l_d+1}, \ldots, f_m^{l_1+\ldots+l_d+l_{d+1}})$$

is contained in every one of the sets (2.19) with a probability greater than $1 - \lambda_{d+2} = \alpha$, $\alpha > \frac{1}{2}$, if $l \geqq l_0(\alpha, \varepsilon, \pi)$, a known function. Since the channel is memoryless we obtain that the probability for u^* to be in $\overline{M}_{d+2}$ is greater than

$$\sum_{r=[l^{1/4}/2]}^{[l^{1/4}]} \binom{[l^{1/4}]}{r} \alpha^r (1-\alpha)^{[l^{1/4}]-r} \geqq 1 - \exp\{-H(\alpha, 1-\alpha)\, l^{1/4}\}\,. \tag{2.22}$$

We repeat now the same procedure for $s = d + 2, \ldots, \min(D', D) \leqq D$. D' is the largest integer such that

$$\exp\{-E(\varepsilon, \pi, w)\, l_{D'}\} \leqq \alpha', \text{ a constant smaller than } \tfrac{1}{2}\,. \tag{2.23}$$

$l_{D'}$ depends on ε, but is *independent* of l. $l_{D'}$ satisfies

$$\frac{|\log \alpha'|}{E(\varepsilon, \pi, w)} + 1 \geqq l_{D'} \geqq \frac{|\log \alpha'|}{E(\varepsilon, \pi, w)}\,. \tag{2.24}$$

We thus come finally up with a set $\overline{M}_{D'}$ of messages, where

$$|\overline{M}_{D'}| \leqq a^{l_{D'}}\,. \tag{2.25}$$

$C > 0$ implies that at least two row vectors of $w(\cdot|\cdot)$ are different. One can therefore easily construct a code $(n_0(l_{D'}), a^{l_{D'}}, \alpha')$ for our channel. $n_0(l_{D'})$ depends only on α' and ε. If we send every code word $[l^{1/4}]$ times we decrease the error probability to $\lambda_{D'} \leqq \exp\{-H(\alpha', 1-\alpha')\, l^{1/4}\}$. Thus we reduce the set $\overline{M}_{D'}$ to a set with one

element. The probability λ that this is not an image of message m satisfies

$$(2.26)\quad \lambda \leqq \sum_{s=1}^{d} \lambda_s + \sum_{s=d+1}^{D'-1} \lambda_s + \lambda_{D'} \leqq \exp\{-\tfrac{1}{2}E(\varepsilon, \pi, w)\, l^{1/2}\} + \\ + f^*(\varepsilon) \log l \,.\, \exp\{-H(\alpha', 1-\alpha')\, l^{1/4}\} + \exp\{-H(\alpha', 1-\alpha')\, l^{1/4}\}\,.$$

The total number n of letters sent is less than

$$H^{-1} \log a(1 + (K(\varepsilon) + K(\varepsilon))^2 + \ldots)\, l + f^*(\varepsilon)\, l^{3/4} \log l + n_0(l_{D'})\,.$$

We therefore have:

$$(2.27)\qquad n \leqq H^{-1} \log a \left(\frac{1}{1 - K(\varepsilon)}\right) l + f^*(\varepsilon)\, l^{3/4} \log l + n_0(l_{D'})\,.$$

Since $K(\varepsilon) = \bar{H}/H + \bar{f}(\varepsilon)/H$, we obtain from (2.27) that

$$(2.28)\qquad l \geqq \log a^{-1} \,.\, (H - \bar{H} - \bar{f}(\varepsilon))\, n - \bar{g}(\varepsilon, n)\,,$$

where $\bar{g}(\varepsilon, n)$ is a known function and equals $o(n)$. (2.28) and $N = a^l$ imply

$$(2.29)\qquad N = \exp\{l \,.\, \log a\} \geqq \exp\{(H - \bar{H} - \bar{f}(\varepsilon))\, n - \bar{g}(\varepsilon, n)\}\,.$$

It follows now from the definitions of π, H, $\bar{H}$ and from (1.8) that $N \geqq \exp\{Cn - \bar{f}(\varepsilon)\, n - \bar{g}(\varepsilon, n)\}$.

We thus have proved the

THEOREM (Coding theorem for d.m.c.f.). *Given R, $0 < R < C$, then one can compute an $E(R)$ such that for every n $(n = 1, 2, \ldots)$ one can give explicitly a code of length $N = e^{Rn}$ such that the decoding error probability λ is smaller than $e^{-E(R)n^{1/4}}$.*

REMARK. We were not concerned about the problem to find the best possible bound on the error probability λ. One easily can improve on our bound by refining our estimates and our coding scheme.

3. AN ALTERNATE CODING SCHEME FOR THE B.S.C.F.

Let now $X = Y = \{0, 1\}$ and let $w(\cdot\,|\,\cdot)$ be a 2×2-stochastic matrix satisfying:

$$w(0\,|\,0) = w(1\,|\,1) = q > \tfrac{1}{2}\,,$$
$$w(1\,|\,0) = w(0\,|\,1) = p = 1 - q\,;$$

$w(\cdot\,|\,\cdot)$ is the transmission matrix of a b.s.c. It is well-known that for base 2 the

capacity C of the b.s.c. – and therefore according to [9] also the capacity of the b.s.c.f. – is given by

$$C = H(\tfrac{1}{2}, \tfrac{1}{2}) - H(q, p) = 1 + q \log_2 q + p \log_2 p \,. \tag{3.1}$$

Let l be a positive integer and let $M = \{1, 2, \ldots, 2^l\}$ be a set of $N = 2^l$ messages. We describe now our encoding procedure.

In a first step both, sender and receiver, partition M into 2 sets of equal size:

$$M_0^0 = \{1, 2, \ldots, 2^{l-1}\} \tag{3.2}$$

and

$$M_1^0 = \{2^{l-1} + 1, \ldots, 2^l\} \,.$$

Suppose the sender is going to send message i, $i \in M$. If $i \in M_0^0$, he sends at the first instant a 0 and if $i \in M_1^0$, he sends a 1. The receiver receives – no matter what the sender has sent – a 0 or a 1. Since we have a channel with feedback, the letter received by the receiver is also known to the sender. If 0 is received, sender and receiver count this as 1 success for each message in M_0^0, and if 1 is received, they count it as 1 success for each message in M_1^0. Let S_0^1 be the set of messages, which had no success at the first instant, and let S_1^1 be the set of messages, which had a success. Obviously, $|S_0^1| = |S_1^1| = 2^{l-1}$. We partition S_0^1 into 2 sets of equal size, $\dot{S}_0^1$ and $\dot{S}_0^1$, say. $\dot{S}_0^1$ shall contain the smaller (message) numbers and $\dot{S}_0^1$ shall contain the larger numbers of S_0^1. Similarily we define $\dot{S}_1^1$ and $\dot{S}_1^1$. A sub-point and a super-point shall have always this meaning in the sequel.

(This device to partition S_0^1 and S_1^1 into two sets of equal size could be replaced by any other device, which is known to both, sender and receiver.) Define now M_0^1 and M_1^1 by

$$M_0^1 = \dot{S}_0^1 \cup \dot{S}_1^1 \quad \text{and} \quad M_1^1 = \dot{S}_0^1 \cup \dot{S}_1^1 \,. \tag{3.3}$$

The sender sends now a 0 or a 1 depending on whether $i \in M_0^1$ or $i \in M_1^1$. If a 0 is received, sender and receiver count this as a success for every message in M_0^1 and if a 1 is received, they count this as a success for each message in M_1^1. Let now S_0^2 be the set of messages with no success, S_1^2 be the set of messages with 1 success, and S_2^2 be the set of messages with 2 successes. Our procedure is such that $|S_0^2| = |S_2^2| = 2^{l-2}$ and $|S_1^2| = 2^{l-1}$.

Now we partition S_0^2 into two sets of equal size, $\dot{S}_0^2$ and $\dot{S}_0^2$, say. Similarly, we partition S_1^2 into the sets $\dot{S}_1^2$, $\dot{S}_1^2$ and S_2^2 into the sets $\dot{S}_2^2$, $\dot{S}_2^2$. Define M_0^2 and M_1^2 by

$$M_0^2 = \dot{S}_0^2 \cup \dot{S}_1^2 \cup \dot{S}_2^2 \tag{3.4}$$

and

$$M_1^2 = \dot{S}_0^2 \cup \dot{S}_1^2 \cup \dot{S}_2^2 \,. \tag{3.5}$$

The sender sends now a 0 or a 1 depending on whether $i \in M_0^2$ or $i \in M_1^2$.

By iteration we obtain sets S_k^t $(0 \leqq k \leqq t;\ t = 1, 2, \ldots, l)$, where S_k^t contains the messages with k successes after t letters have been sent. What the elements of S_k^t are is a matter of chance, however, the cardinality $|S_k^t|$ of S_k^t satisfies a simple recursion formula.

Define for convenience $S_{-1}^s = \emptyset$ for $s = 0, 1, 2, \ldots, l$, $S_k^t = \emptyset$ for $k > t$ and $S_0^0 = M$. Then we have

$$|S_k^t| = \tfrac{1}{2}|S_{k-1}^{t-1}| + \tfrac{1}{2}|S_k^{t-1}| \tag{3.6}$$

for $k = 0, 1, \ldots, t;\ t = 1, 2, \ldots, l$. Since $|S_0^0| = |M| = 2^l$, the solution of (2.6) is given by

$$|S_k^t| = \frac{1}{2^l}\binom{t}{k} \cdot 2^l \tag{3.7}$$

for $k = 0, 1, 2, \ldots, t;\ t = 1, 2, \ldots, l$. In particular we have

$$|S_k^l| = \binom{l}{k} \quad \text{for} \quad k = 0, 1, \ldots, l\,. \tag{3.8}$$

Where is message i after the first l letters have been sent? For every letter sent the probability of a success for message i is q. Therefore i will be with high probability in one of the sets

$$S_{[ql-\varepsilon l]}^l, S_{[ql-\varepsilon l]+1}^l, \ldots, S_l^l\,,$$

where ε is a fixed number between 0 and $q - \frac{1}{2}$. Denote the union of these sets by $S(l, q, \varepsilon)$. The probability that i is in $S(l, q, \varepsilon)$ equals $\sum_{k=[ql-\varepsilon l]}^{l} \binom{l}{k} q^k p^{l-k}$. It is well-known (see for instance inequality (A.6) on page 246 in Peterson's book "Error correcting codes") that

$$\sum_{k=[ql-\varepsilon l]}^{l} \binom{l}{k} q^k p^{l-k} \geqq 1 - 2^{-E(\varepsilon,q)l}\,, \tag{3.9}$$

where

$$E(\varepsilon, q) = +(q - \varepsilon)\log_2 \frac{q - \varepsilon}{q} + (p + \varepsilon)\log_2 \frac{p + \varepsilon}{p}\,.$$

$E(\varepsilon, q)$ is positive. The cardinality of $S(l, q, \varepsilon)$ can be estimated by

$$|S(l, q, \varepsilon)| = \sum_{k=[ql-\varepsilon l]}^{l} \binom{l}{k} \leqq (q - \varepsilon)^{-(q-\varepsilon)l}(p + \varepsilon)^{-(p+\varepsilon)l}\,. \tag{3.10}$$

(See Peterson, inequality (A.8)). Denoting the entropy of the probability vector $(q - \varepsilon, p + \varepsilon)$ by $H(q - \varepsilon, p + \varepsilon)$ we obtain therefore that

$$|S(l, q, \varepsilon)| \leqq 2^{H(q-\varepsilon, p+\varepsilon)l}\,. \tag{3.11}$$

The equations $N = 2^{H(1/2,1/2)l}$, (3.9) and (3.11) are now full substitutes for the lemmas 1, 2, 3. We can now iterate the procedure in exactly the same way as in section 2 and thus obtain a non-sequential coding scheme for the b.s.c.f. The scheme is optimal in the sense that we can achieve any rate below the capacity with an arbitrary small decoding error probability.

REMARK. After this paper was finished E. Berlekamp pointed out to me that he used the idea to partition the messages already in his very interesting paper [2]. However, the combination of this idea with the idea of an iterative procedure, as described in section 2, seems to be new. For this reason and also because it may be interesting to compare the different approaches taken in section 2 and section 3, we did not exclude the later section.

REFERENCES

[1] AHLSWEDE, R.: The capacity of channels with arbitrarily varying channel probability functions in the presence of feedback. To appear in Zeitschrift f. Wahrsch. u. verw. Geb.

[2] BERLEKAMP, E. R.: Block coding for the binary symmetric channel with noiseless delayless feedback. Proceedings of the Symposium on Error Correcting Codes, University of Wisconsin, 1968.

[3] FEINSTEIN, A.: A new basic theorem of information theory. Trans. IRE *PGIT* (1954), 2—22.

[4] HORSTEIN, M.: Sequential transmission using noiseless feedback. IEEE Trans. Information Theory *IT-9* (1963), 136—143.

[5] KEMPERMAN, J. H. B.: Strong converses for a general memoryless channel with feedback. Presented at the Sixth Prague Conference on Inf. Th., Random Processes and Statistical Decision Functions. This volume 375—409.

[6] SCHALKWIJK, J. P. M., KAILATH, P.: A coding scheme for additive noise channels with fedback — part I. IEEE Trans. Information Theory *IT-12* (1966), 172--182.

[7] SCHALKWIJK, J. P. M.: A coding scheme for additive noise channels with feedback-part II. IEEE Trans. Information Theory *IT-12* (1966), 183—189.

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

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

[10] WOLFOWITZ, J.: The coding of messages subject to chance errors. Illinois J. Math. (1957), 4, 591—606.

THE OHIO STATE UNIVERSITY

and

UNIVERSITY OF ILLINOIS