Channels without Synchronization

R. Ahlswede; J. Wolfowitz

Advances in Applied Probability, Vol. 3, No. 2. (Autumn, 1971), pp. 383-403.

Adv. Appl. Prob. 3, 383–403 (1971)
Printed in Israel
© *Applied Probability Trust* 1971

CHANNELS WITHOUT SYNCHRONIZATION

R. AHLSWEDE, *Ohio State University and University of Illinois*
J. WOLFOWITZ, *University of Illinois*

Summary

Let $X = \{1,\ldots,a\}$ and $Y = \{1,\ldots,a\}$ be the input and output alphabets, respectively. In the (unsynchronized) channels studied in this paper, when an element of X is sent over the channel, the receiver receives either nothing or a sequence of k letters, each a member of Y, where k, determined by chance, can be 1, or 2, or ... or L, a given integer. The channel is called unsynchronized because the sequence received (for each letter sent) is not separated from the previous sequence or the following sequence, so that the receiver does not know which letters received correspond to which letter transmitted.

In Sections 1 and 2 we give the necessary definitions and auxiliary results. In Section 3 we extend the results of Dobrushin [2] by proving a strong converse to the coding theorem[1] and making it possible to compute the capacity to within any desired accuracy.

In Section 4 we study the same channel *with feedback*, prove a coding theorem and strong converse, and give an algorithm for computing the capacity.

In Section 5 we study the unsynchronized channel where the transmission of each word is governed by an arbitrary element of a set of channel probability functions. Again we obtain the capacity of the channel, prove a coding theorem and strong converse, and give an algorithm for computing the capacity.

In Section 6 we apply results of Shannon [4] and supplement Dobrushin's results on continuous transmission with a fidelity criterion.

1. Introduction

Let $X = \{1,\cdots,a\}$ be the "input alphabet" and $Y = \{1,\cdots,a\}$ be the "output alphabet". Write $X^t = X$ and $Y^t = Y$ for $t = 1,2,\cdots$. By $X_n = \Pi_{t=1}^n X^t$ denote the set of input n-sequences (words of length n) and by $Y_n = \Pi_{t=1}^n Y^t$ denote the set of output n-sequences. Let Y_0 denote the empty set, and $\bar{Y} = \bigcup_{n=0}^{\infty} Y_n$ the set of all finite output sequences. Define $\bar{Y}_n = \Pi_{t=1}^n \bar{Y}^t$ for $n = 1,2,\cdots$, where $\bar{Y}^t = \bar{Y}$ for $t = 1,2,\cdots$.

Let $w(\cdot\,|\,\cdot)$ be a stochastic matrix, i.e.,

$$(1.1)\quad w(\bar{y}\,|\,x) \geqq 0 \text{ for every } x \in X,\ \bar{y} \in \bar{Y}, \text{ and } \sum_{\bar{y} \in Y} w(\bar{y}\,|\,x) = 1 \text{ for every } x \in X.$$

Received in revised form 15 December 1970. Research supported by the Air Force Office of Scientific Research, Office of Aerospace Research, U.S. Air Force, under AFOSR Grant No. 68–1472.

[1] The paper by O. K. Kozlov, Strong converse of Shannon's theorem for channels without memory and with errors of synchronization, *Problemy Peredači Informacii* 7(1971) 102–105 (in Russian), has just reached us. The author proves the strong converse of Section 3 in a manner somewhat different from ours.

The transition probabilities of a synchronized memoryless channel (s.m.c.) are defined by

$$(1.2)\qquad P(\bar{y}_n | x_n) = \prod_{t=1}^{n} w(\bar{y}^t | x^t)$$

for every $x_n = (x^1,\cdots,x^n) \in X_n$ and every $\bar{y}_n = (\bar{y}^1,\cdots,\bar{y}^n) \in \bar{Y}_n$, $n = 1,2,\cdots$. We denote by $\bar{y}^1\bar{y}^2\cdots\bar{y}^t$ an element of $\bar{Y}$ obtained by writing the terms of the sequence $\bar{y}^1,\bar{y}^2,\cdots,\bar{y}^t$ consecutively in their natural order. With this convention we define the transition probabilities of an unsynchronized memoryless channel (u.m.c.) U by

$$(1.3)\qquad P(\bar{y} | x_n) = \sum_{\bar{y}^1\bar{y}^2\cdots\bar{y}^n = \bar{y}} P((\bar{y}^1,\cdots,\bar{y}^n) | x_n)$$

for every $x_n \in X_n$, $\bar{y} \in \bar{Y}$, $n = 1,2,\cdots$.

We now define an unsynchronized compound channel (u.c.c.). Let S be an arbitrary index set and let $\Gamma = \{w(\cdot | \cdot | s) | s \in S\}$ be a set of matrices which satisfy (1.1). For every $s \in S$ define an u.m.c. $P(\cdot | \cdot | s)$ by

$$(1.4)\qquad P(\bar{y} | x_n | s) = \sum_{\bar{y}^1\bar{y}^2\cdots\bar{y}^n = \bar{y}} P((\bar{y}^1,\cdots,\bar{y}^n) | x_n | s)$$

for every $x_n \in X_n$, $\bar{y} \in \bar{Y}$, $n = 1,2,\cdots$.

Consider the class of channels

$$(1.5)\qquad \bar{S} = \{P(\cdot | \cdot | s) | s \in S\}.$$

If we are interested in the simultaneous behavior of all these channels we call this indexed set of channels the unsynchronized compound channel $\bar{S}$. (For compound channels see [6] and [8].)

A code (n,N) is a system

$$(1.6)\qquad \begin{gathered} \{(u_i,A_i) | i = 1,\cdots,N\}, \text{ where } u_i \in X_n,\ A_i \subset \bar{Y}, \\ A_i \cap A_j = \varnothing \quad \text{for } i \neq j. \end{gathered}$$

Let $0 < \lambda < 1$. A code (n,N) is a λ-code (n,N,λ)

(1.7)

(a) for the u.m.c. U, if $P(A_i | u_i) \geqq 1 - \lambda$ for $i = 1,\cdots,N$;

(b) for the u.c.c. $\bar{S}$, if $P(A_i | u_i | s) \geqq 1 - \lambda$ for $i = 1,\cdots,N$ and for every $s \in S$.

We now introduce the unsynchronized memoryless channel with feedback (u.c.f.). By this we mean that there exists a return channel which sends back from the receiving point to the transmitting point the element of $\bar{Y}$ 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 $\bar{N} = \{1, \cdots, N\}$, one of which will be presented to the sender for transmission. Message $i \in \bar{N}$ is encoded by an encoding (vector valued) function

$$(1.8) \qquad f_n(i) = [f_i^1, f_i^2(Z^1), \cdots, f_i^t(Z^1, \cdots, Z^{t-1}), \cdots, f_i^n(Z^1, \cdots, Z^{n-1})],$$

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

$$(1.9) \qquad P(\bar{y} \mid f_n(i)) = \sum_{\bar{y}^1 \cdots \bar{y}^n = \bar{y}} P((\bar{y}^1, \cdots, \bar{y}^n) \mid f_n(i)).$$

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

$$(1.10) \qquad \{(f_n(i), A_i) \mid i = 1, \cdots, N\},$$

where the $f_n(i)$ are as defined in (1.8), $A_i \subset \bar{Y}$ for $i = 1, \cdots, N$, $A_i \cap A_j = \emptyset$ for $i \neq j$, and $P(A_i \mid i) \geqq 1 - \lambda$ for $i = 1, \cdots, N$.

The study of unsynchronized channels up to 1963 was summarized in [3]. The present model of an unsynchronized memoryless channel is due to Dobrushin [2]. The other channels described above have not been treated before. Dobrushin proved (Theorem 1 of [2]) a coding theorem and a weak converse for the u.m.c. under the additional assumption (1.11) on $w(\cdot \mid \cdot)$, which we now describe.

Let $l(\bar{y})$ equal the number of components of $\bar{y}$, i.e., the length of the "letter" $\bar{y}$. There exist constants $c_1 < \infty$, $c_2 > 0$, such that

$$(1.11) \qquad c_2 \leqq \sum_{\bar{y} \in \bar{Y}} l(\bar{y}) w(\bar{y} \mid x) \leqq c_1 \text{ for every } x \in X.$$

The capacity C obtained by him is

$$(1.12) \qquad C = \lim_{n \to \infty} C_n,$$

where[2]

$$(1.13) \quad C_n = \sup_{p_n} n^{-1} \sum_{x_n \in X_n} \sum_{\bar{y} \in \bar{Y}} p_n(x_n) P(\bar{y} \mid x_n) \log \frac{P(\bar{y} \mid x_n)}{\sum_{x_n \in X_n} p_n(x_n) P(\bar{y} \mid x_n)},$$

the supremum being taken over the set of all probability distributions on X_n.

[2] All logarithms in this paper are to the base 2. The expression "exp { }" means "exp$_2$ { }".

Instead of Assumption (1.11) we shall make, throughout this paper, the assumption:

(1.14) If $w(\bar{y} \mid x) > 0$ for some $x \in X$, then $0 \leqq l(\bar{y}) \leqq L$, where L is a constant.

This assumption is a little weaker than (1.11) with respect to the lower bound on $l(\bar{y})$ and certainly stronger than (1.11) with respect to the upper bound on $l(\bar{y})$. We prove a coding theorem and a strong converse under Assumption (1.14) by using the method of "generated sequences", which was developed in [5]. We also give an estimate on the speed of approximation in (1.12), which makes it possible actually to compute C to within any desired accuracy, i.e., makes C "computable". We show by an example that this estimate is the best possible to within a constant factor. The method of "generated sequences" turns out to be particularly suited to the treatment of unsynchronized channels with feedback (Section 4) and unsynchronized compound channels (Section 5). We prove coding theorems and strong converses for these channels and make the capacities computable. These results could also be extended to other channels, such as unsynchronized non-stationary memoryless channels (cf. [1]), and unsynchronized finite-state channels ([8], Chapter 6).

2. Auxiliary results

We now repeat the definition of generated sequences and the basic lemmas about them. For proofs we refer to [5] or [8], Chapter 2.

Let $w(\cdot \mid \cdot)$ be an $a \times a$-stochastic matrix, which will be called a "channel probability function" (c.p.f.). For an n-sequence $u = (u^1, \cdots, u^n) \in X_n$ and an n-sequence $v = (v^1, \cdots, v^n) \in Y_n$ define

$$(2.1) \qquad N(i \mid u) = \left| \{ t \mid 1 \leqq t \leqq n,\ u^t = i \} \right| \text{ for } i = 1, \cdots, a.$$

$$(2.2) \qquad N(i,j \mid u,v) = \left| \{ t \mid 1 \leqq t \leqq n,\ u^t = i,\ v^t = j \} \right| \quad \text{for } i = 1, \cdots, a;\ \ j = 1, \cdots, a.$$

Let $\pi = (\pi_1, \cdots, \pi_a)$ be a probability distribution on X. The sequence $u \in X_n$ is called a π-sequence if

$$(2.3) \qquad \left| N(i \mid u) - n\pi_i \right| \leqq 2[an\pi_i(1 - \pi_i)]^{\frac{1}{2}}, \quad i = 1, \cdots, a.$$

The sequence $v \in Y_n$ is said to be generated by $u \in X_n$ if

$$(2.4) \qquad \left| N(i,j \mid u,v) - N(i \mid u)w(j \mid i) \right| \leqq \delta[N(i \mid u)w(j \mid i)(1 - w(j \mid i))]^{\frac{1}{2}}$$

for all $i,j = 1, \cdots, a$, and a $\delta > 2a$ to be chosen later. We denote the set of π-sequences by $X_n(\pi)$, the set of sequences generated by $u \in X_n$ by $G(u)$ and the set of sequences generated by any π-sequence by $G(\pi)$.

We also need the following definitions.

(2.5) $\pi' = (\pi'_1, \cdots, \pi'_a)$, where $\pi'_j = \sum_i \pi_i w(j \mid i), \quad j = 1, \cdots, a.$

(2.6) π_n (respectively π'_n) denotes the independent product distribution on X_n (respectively Y_n) with 1-dimensional marginal-distributions π (respectively π').

(2.7) $B(w \mid u) = \left| G(u) \right|$,

(2.8) $B(w \mid \pi) = \left| G(\pi) \right|$.

(2.9) H is the entropy function; for a probability vector $p = (p_1, \cdots, p_b)$, $H(p) = - \sum_{i=1}^{b} p_i \log p_i$.

Lemma 2.1. $\pi_n(X_n(\pi)) \geqq \frac{3}{4}$.

Lemma 2.2. $P(G(u) \mid u) \geqq 1 - \varepsilon'$ for every $u \in X_n$, where $\varepsilon' \leqq a^2\delta^{-2} < \frac{1}{4}$, so that $\varepsilon' \to 0$ as $\delta \to \infty$.

Lemma 2.3. If $v \in G(\pi)$, then $\exp\{-nH(\pi') - K_1(a, \delta)\sqrt{n}\} < \pi'_n(v) < \exp\{-nH(\pi') + K_1(a,\delta)\sqrt{n}\}$. The function K_1 is independent of v, n, π and w.

Lemma 2.4. $\exp\{nH(\pi') - K_2(a,\delta)\sqrt{n}\} < B(w \mid \pi) < \exp\{nH(\pi') + K_2(a,\delta)\sqrt{n}\}$. The function K_2 is independent of n, π, and w.

Lemma 2.5. Let $u \in X_n(\pi)$, then

$$\exp\{n \sum_i \pi_i H(w(\cdot \mid i)) - K_3(a,\delta)\sqrt{n}\} < B(w \mid u)$$
$$< \exp\{n \sum_i \pi_i H(w(\cdot \mid i)) + K_3(a,\delta)\sqrt{n}\}.$$

The function K_3 does not depend on u, n, π or w.

These five lemmas suffice for the proof of the coding theorem and its strong converse for the d.m.c. (see [8], Chapter 3), which we now state. Define the constant C by

$$C = \max_\pi (H(\pi') - \sum_i \pi_i H(w(\cdot \mid i))).$$

One can explicitly give a positive function $K(\lambda)$ of λ, $0 < \lambda < 1$, such that;

Coding theorem. For any n and λ there exists a code (n, N, λ) with

$$N > \exp\{nC - a^4 K(\lambda)\sqrt{n}\}.$$

Strong converse. For any n and λ there does not exist a code (n, N, λ) such that

$$N > \exp\{nC + a^4 K(\lambda)\sqrt{n}\}.$$

The function $K(\cdot)$ does not depend on the c.p.f. $w(\cdot\,|\,\cdot)$ or upon the alphabet size a.

If one weakens Lemmas 2.3, 2.4 and 2.5 by replacing $K_1(a,\delta)\sqrt{n}$, $K_2(a,\delta)\sqrt{n}$, $K_3(a,\delta)\sqrt{n}$ by $o(n)$, one obtains the coding theorem and its strong converse with an accuracy $o(n)$ instead of $K\sqrt{n}$. In Section 3 we will give a definition of generated sequences for the u.m.c. such that these lemmas hold with an accuracy $o(n)$. The proof of the strong converse of the coding theorem for the d.m.c. will then carry over verbatim to the u.m.c. The proof of the coding theorem for the u.m.c. will require an additional argument.

3. Unsynchronized memoryless channels

In (1.2) we defined a synchronized memoryless channel (s.m.c.). We now introduce a channel J with less synchronization to serve only as a tool. Let $I(n)$ be a function of n $(n = 1,2,\cdots)$ with positive integer values, such that $I(n) = o(n)$, and $I(n) \to \infty$ as $n \to \infty$. Define $s(n) = [n/I(n)] - 1$, where [ ] denotes the smallest integer larger than the number in the brackets. Let

$$(X_{I(n)})^t = \prod_{s=I(n)(t-1)+1}^{I(n).t} X^s \text{ for } t = 1,2,\cdots$$

and let

$$(X_{I(n)})_m = \prod_{t=1}^{m} (X_{I(n)})^t.$$

We write for convenience $\tilde{X}$ instead of $X_{I(n)}$ and we define $\tilde{Y}$ as

$$\tilde{Y} = \{\bar{y}\,|\,\bar{y} = \bar{y}^1\cdots\bar{y}^{I(n)},\quad 0 \leqq l(\bar{y}^s) \leqq L,\quad \text{for } s = 1,\cdots,I(n)\}.$$

The transition matrix of the synchronized channel J is defined by

$$\text{(3.1)}\qquad \begin{aligned} w^*(\bar{y}\,|\,\tilde{x}) &= \sum_{\bar{y}^1\cdots\bar{y}^{I(n)}=\bar{y}} P((\bar{y}^1,\cdots,\bar{y}^{I(n)})\,|\,(x^1,\cdots,x^{I(n)})) \\ &\qquad\qquad \text{for every } \tilde{x} = (x^1,\cdots,x^{I(n)}),\quad \bar{y}\in\bar{Y}. \end{aligned}$$

For every $\tilde{x}_m = (\tilde{x}^1,\cdots,\tilde{x}^m) \in (X_{I(n)})_m$ the transition probabilities of J are defined by

$$\text{(3.2)}\qquad Q(\bar{y}_m\,|\,\tilde{x}_m) = \prod_{t=1}^{m} w^*(\bar{y}^t\,|\,\tilde{x}^t)$$

for every $\bar{y}_m = (\bar{y}^1,\cdots,\bar{y}^m) \in \bar{Y}_m$ and every $m = 1,2,\cdots$. J is a d.m.c. with input alphabet $\tilde{X}$ of size $a^{I(n)}$ and output alphabet $\tilde{Y} \subset \bar{Y}$ of size $A < a^{(L+1)\cdot I(n)}$. We now treat the u.m.c. U by comparing it with the d.m.c. J.

Recalling the definition of C_n given in (1.13) we see that J has capacity $I(n)\cdot C_{I(n)}$. If, in Section 2, we replace X by $\tilde{X}$, Y by $\tilde{Y}$, n by m, and a by A, then the five lemmas of Section 2 hold for channel J. Now let $m = s(n)$

and $n' = s(n) \cdot I(n)$. Obviously, n' differs from n by less than $I(n)$. We now use the enlarged alphabet $\tilde{X}$ in the u.m.c. U. After these preparations we are ready to state the basic definition:

(3.3) $\bar{y} \in \bar{Y}$ is *generated* by $u \in \tilde{X}_{s(n)}$ with respect to the u.m.c. U if there exist $\tilde{y}^1, \cdots, \tilde{y}^{s(n)} \in \tilde{Y}$ such that $\bar{y} = \tilde{y}^1 \cdots \tilde{y}^{s(n)}$ and $(\tilde{y}^1, \cdots, \tilde{y}^{s(n)})$ is generated by u with respect to the channel J in the sense of (2.4).

Define $G(u)$, $G(\pi)$, π'_m, $B(w^* \mid \pi)$, $B(w^* \mid u)$ for channel J as in Section 2. The corresponding quantities for channel U will be denoted by $G^*(u)$, $G^*(\pi)$, $\pi^{*\prime}_m$, $B^*(w^* \mid \pi)$, $B^*(w^* \mid u)$. Then we have

Lemma 3.1. $\pi_{s(n)}(\tilde{X}_{s(n)}(\pi)) \geqq \frac{3}{4}$.

(This is Lemma 2.1 restated.)

Lemma 3.2. $P(G^*(u) \mid u) \geqq 1 - \varepsilon'$ for every $u \in \tilde{X}_{s(n)}$, where $\varepsilon' \leqq A^2\delta^{-2} < \frac{1}{4}$, so that $\varepsilon' \to 0$ as $\delta \to \infty$.

This is obvious from Lemma 2.2 and Definition 3.3.

Define M by

$$M = \log \binom{n' \cdot L + s(n)}{s(n)}. \tag{3.4}$$

We now prove the following lemmas.

Lemma 3.3. If $v \in G^*(u)$, then $\exp\{-s(n)H(\pi^{*\prime}) - K_1(A, \delta)[s(n)]^{\frac{1}{2}}\} < \pi^{*\prime}_{s(n)}(v)$.

Lemma 3.4. $\exp\{s(n)H(\pi^{*\prime}) - K_2(A, \delta)[s(n)]^{\frac{1}{2}} - M\} < B^*(w^* \mid \pi) < \exp\{s(n)H(\pi^{*\prime}) + K_2(A, \delta)[s(n)]^{\frac{1}{2}}\}$.

Lemma 3.5. If $u \in \tilde{X}_{s(n)}(\pi)$, then

$$\exp\{s(n) \sum_i \pi_i H(w^*(\cdot \mid i)) - K_3(A, \delta)[s(n)]^{\frac{1}{2}} - M\} < B^*(w^* \mid u)$$
$$< \exp\{s(n) \sum_i \pi_i H(w^*(\cdot \mid i)) + K_3(A, \delta)[s(n)]^{\frac{1}{2}}\}.$$

We prove Lemma 3.5 first. It follows from Definition (3.3) that

$$B^*(w^* \mid u) \leqq B(w^* \mid u). \tag{3.5}$$

In order to get a lower bound on $B^*(w^* \mid u)$ we have to count in how many ways $\bar{y}$ can be written as $\tilde{y}^1, \cdots, \tilde{y}^{s(n)}$. This number is bounded by

$$\binom{n' \cdot L + s(n)}{s(n)}.$$

Therefore,

$$B(w^*|u)\binom{n'L+s(n)}{s(n)}^{-1} \leqq B^*(w^*|u). \tag{3.6}$$

Similarly,

$$B(w^*|\pi)\binom{n'L+s(n)}{s(n)}^{-1} \leqq B^*(w^*|\pi) \leqq B(w^*|\pi). \tag{3.7}$$

Lemma 3.5 now follows from Lemma 2.5, (3.5) and (3.6). Lemma 3.4 follows from Lemma 2.4 and (3.7). Lemma 3.3 is a consequence of Lemma 2.3 and the Definition (3.3).

Using Lemmas 3.1—3.5, instead of Lemmas 2.1—2.5, enables us to carry over verbatim the proof of the strong converse of the coding theorem, given for a d.m.c. in [8], Chapter 3, to the u.m.c. U.

One also can prove the strong converse of the coding theorem for channel U by using the strong converse of the coding theorem for the d.m.c. J. Obviously, an upper bound on the length of a code for channel J is *a fortiori* an upper bound on the length of the "corresponding" code for the u.m.c. U.

To prove the coding theorem we need an additional argument, because Lemma 3.3 is not the exact analogue of Lemma 2.3. We now supply this additional argument and prove the coding theorem. Without loss of generality we take $\lambda < \frac{1}{2}$. We begin the construction of a code for channel U as in [8], Theorem 3.2.1. Choose $\delta > 2A$ so that $Q(G(u)|u) \geqq 1-\frac{1}{4}\lambda$ for every $u \in \tilde{X}_{s(n)}$. Then also $P(G^*(u)|u) \geqq 1-\frac{1}{4}\lambda$ for every $u \in \tilde{X}_{s(n)}$. Let $\{(u_i, A_i)|i = 1,\cdots,N\}$ be a code for channel U such that the following conditions are satisfied:

(a) $u_i \in \tilde{X}_{s(n)}(\pi)$, $i = 1,\cdots,N$.

(b) $A_i = G^*(u_i) - \bigcup_{j=1}^{i-1} A_j$.

(c) $P(A_i|u_i) \geqq 1-\lambda$, $i = 1,\cdots,N$.

(d) The code is maximal in the sense that it is impossible to add an element (u_{N+1}, A_{N+1}) such that (u_{N+1}, A_{N+1}) satisfies (a), (b) and (c) for $i = N+1$.

From Lemmas 3.1—3.5 we see that the proof of [8] would go through verbatim if we could only give the proper lower bound on the number of sequences in $\bigcup_{i=1}^{N} A_i$. In order to obtain this lower bound we proceed as follows. We define sets $B_i, i = 1,\cdots,N$, by $B_i = \{(\bar{y}^1,\cdots,\bar{y}^{s(n)})|\bar{y}^1\cdots\bar{y}^{s(n)} \in A_i\}$ and prove that $\{(u_i, B_i)|i = 1,\cdots,N\}$ has the following two properties.

(e) It is a code with maximal probability of error λ for the d.m.c. J.

(f) The code is maximal in the sense that it is impossible to add a pair (u_{N+1}, B_{N+1}), where $u_{N+1} \in \tilde{X}_{s(n)}(\pi)$ and $B_{N+1} \subset G(u_{N+1})$, such that $\{(u_1, B_1),\cdots,(u_{N+1}, B_{N+1})\}$ is a code with maximal probability of error λ for channel J.

Property (e) follows from the definitions of U and J. Suppose Property (f) did not hold. Then we could find a pair (u_{N+1}, B_{N+1}), with $u_{N+1} \in \tilde{X}_{s(n)}(\pi)$

and $B_{N+1} \subset G(u_{N+1})$ such that $\{(u_1,B_1),\cdots,(u_{N+1},B_{N+1})\}$ is a code with maximal probability of error λ for channel J. Define

$$A_{N+1} = \{\bar{y} \,|\, \bar{y} = \bar{y}^1 \cdots \bar{y}^{s(n)} \text{ for } (\bar{y}^1,\cdots,\bar{y}^{s(n)}) \in B_{N+1}\}.$$

Because $B_{N+1} \cap B_i = \varnothing$ for $i = 1,\cdots,N$, and because of the way in which the B_i, $i = 1,\cdots,N$, were constructed from their respective A_i, $A_{N+1} \cap A_i = \varnothing$ for $i = 1,\cdots,N$. Since $B_{N+1} \subset G(u_{N+1})$, $A_{N+1} \subset G^*(u_{N+1})$. From this and the fact that $P(A_{N+1} \,|\, u_{N+1}) \geqq Q(B_{N+1} \,|\, u_{N+1}) \geqq 1 - \lambda$, we obtain a contradiction to (d) and thus prove (f).

From Properties (e) and (f) of the code $\{(u_i,B_i) \,|\, i = 1,\cdots,N\}$ for channel J we have that, for every $u \in \tilde{X}_{s(n)}(\pi)$,

$$Q\left(G(u) \cap \bigcup_{i=1}^{N} B_i \,\middle|\, u\right) > \lambda - \tfrac{1}{4}\lambda = \tfrac{3}{4}\lambda,$$

and hence that, for every $u \in \tilde{X}_{s(n)}(\pi)$,

$$Q\left(\bigcup_{i=1}^{N} B_i \cap G(\pi) \,\middle|\, u\right) > \tfrac{3}{4}\lambda.$$

Hence

$$\pi'_{s(n)}\left(\bigcup_{i=1}^{N} B_i \cap G(\pi)\right) > \tfrac{3}{4}\cdot\tfrac{3}{4}\lambda = 9\lambda/16.$$

Applying Lemma 2.3 for channel J we obtain that the set $\bigcup_{i=1}^{N} B_i \cap G(\pi)$ contains at least $\exp\{s(n)H(\pi^{*\prime}) - K_1'(A,\delta)[s(n)]^{\frac{1}{2}}\}$ sequences. Hence $\bigcup_{i=1}^{N} A_i$ contains at least $\exp\{s(n)H(\pi^{*\prime}) - K_1'(A,\delta)[s(n)]^{\frac{1}{2}} - M\}$ sequences.

Let $N(n,\lambda)$ be the maximal N for which a code (n,N,λ) exists. We obtain the following result. For any λ, $0 < \lambda < 1$, one can give explicitly a positive function $K(\lambda)$ such that

$$N(n,\lambda) > \exp\{nC_{I(n)} - (\log a)\cdot I(n) - K(\lambda)A^4[s(n)]^{\frac{1}{2}} - M\}. \tag{3.8}$$

We have already obtained the result that

$$N(n,\lambda) < \exp\{nC_{I(n)} + (\log a)\cdot I(n) + K(\lambda)A^4[s(n)]^{\frac{1}{2}}\}. \tag{3.9}$$

We now choose $I(n)$ such that

$$A^4\sqrt{n} < a^{4(L+1)I(n)}\sqrt{n} \leqq n^{\frac{3}{4}}. \tag{3.11}$$

We can choose

$$I(n) = T\cdot \log n, \tag{3.12}$$

with

$$T = [16\cdot(L+1)\log a]^{-1}. \tag{3.13}$$

For this choice of $I(n)$ we obtain, by Stirling's approximation, the following bound on M. For $n \geqq 4$,

$$M \leqq c(a,L)\cdot n\cdot\frac{\log\log n}{\log n}, \tag{3.14}$$

where $c(a,L)$ can easily be computed explicitly, and does not depend on the transition matrix of the channel. Define $T'(\lambda,a,L) = T^{-\frac{1}{2}}K(\lambda) + \log a + c(a,L)$. From (3.8), (3.9), (3.11), and (3.14) we obtain that for any λ, $0<\lambda<1$, and for $n \geqq 4$,

$$N(n,\lambda) > \exp\left\{n\cdot C_{T\log n} - T'\cdot\frac{n\log\log n}{\log n}\right\} \tag{3.15}$$

and

$$N(n,\lambda) < \exp\left\{nC_{T\log n} + T'\cdot\frac{n\log\log n}{\log n}\right\}. \tag{3.16}$$

We now prove that $\lim_{n\to\infty}C_n$ exists and give an estimate on the speed with which C_n approaches this limit. Choosing $I(n)=\frac{1}{2}T\cdot\log n$ with T as in (3.13), we obtain, by the same argument as before, that we can compute a $T''(\lambda,a,L)$ such that for any λ, $0<\lambda<1$,

$$N(n,\lambda) > \exp\left\{nC_{\frac{1}{2}T\log n} - T''\cdot n\cdot\frac{\log\log n}{\log n}\right\} \tag{3.17}$$

and

$$N(n,\lambda) < \exp\left\{nC_{\frac{1}{2}T\log n} + T''\cdot n\cdot\frac{\log\log n}{\log n}\right\}. \tag{3.18}$$

Fix λ at any value between 0 and 1. It follows from (3.15)–(3.18) that

$$\left|C_{T\log n} - C_{\frac{1}{2}T\log n}\right| \leqq T'''\frac{\log\log n}{\log n}, \tag{3.19}$$

where

$$T''' = T' + T''.$$

Choose $n = 2^{2^h}, 2^{2^{h+1}}, \cdots$, successively. We obtain that

$$\begin{aligned} \left|C_{T2^h} - C_{T2^{h-1}}\right| &\leqq T'''\frac{h}{2^h}, \\ \left|C_{T2^{h+1}} - C_{T2^h}\right| &\leqq T'''\frac{h+1}{2^{h+1}}, \end{aligned} \tag{3.20}$$

and so on. (3.20) implies that $\lim_{h\to\infty}C_{T2^h}$ exists. Denote this limit by C. Then

$$(3.21)\qquad \begin{aligned} |C_{T2^{h-1}} - C| &\leqq T''' \sum_{v=0}^{\infty} \frac{h+v}{2^{h+v}} = T''' \frac{h}{2^h} \sum_{v=0}^{\infty} \frac{1+(v/h)}{2^v} \\ &\leqq T^{(4)} \frac{h}{2^h}, \text{ with } T^{(4)} = T''' \sum_{v=0}^{\infty} \frac{1+v}{2^v}, \end{aligned}$$

for $h = 2,3,\cdots$. Choosing $I(n) = \alpha T \log n$ with α between $\frac{1}{2}$ and 1, we obtain (3.17) and (3.18) with $C_{\frac{1}{2}T\log n}, T''$ replaced by $C_{\alpha T \log n}, T(\alpha)$, where $T' \leqq T(\alpha) \leqq T''$. Therefore

$$(3.22)\qquad |C_{T\log n} - C_{\alpha T \log n}| \leqq T''' \frac{\log\log n}{\log n},$$

and

$$|C_{T2^h} - C_{\alpha T 2^h}| \leqq T''' \frac{h}{2^h}.$$

For any n there exist h, α ($\frac{1}{2} \leqq \alpha \leqq 1$) such that $n = \alpha T 2^h$. Then (3.21), (3.22) imply that

$$(3.23)\qquad \begin{aligned} |C_n - C| &\leqq (T''' + T^{(4)}) \frac{h}{2^h} \\ &= (T''' + T^{(4)}) \frac{\log n - \log \alpha T}{n} \cdot \alpha T. \end{aligned}$$

As a consequence of (3.23) we obtain

Theorem 3.1. There exists a positive $T^*(a,L)$, which depends only upon a and L and can be explicitly computed by application of our argument above, such that

$$|C_n - C| \leqq T^* \frac{\log n}{n}$$

for $n \geqq 2$.

The following examples show that Theorem 3.1 is, to within the constant T^*, the best possible.

Example 1. Let $X = \{0,1\}$, $Y = \{0\}$ and w such that $w(0|0) = 1$, $w(00|1) = 1$. We have $C_n = n^{-1}\log(n+1)$, $C = 0$.

Example 2. T^* has to depend on the alphabet length a. Let $X = \{1,2,\cdots,a\}$, $Y = \{0\}$ and w be such that

$$w(0|1) = 1,\ w(00|2) = 1, \cdots, w(00\cdots 0|a) = 1.$$

Then $C_n = n^{-1}\log(n(a-1)+1)$, $C = 0$.

From (3.17), (3.18) and Theorem 3.1 we obtain

Theorem 3.2 (Coding theorem and strong converse for the u.m.c.) For any λ, $0 < \lambda < 1$,

$$N(n,\lambda) > \exp\left\{nC - T^{**}(\lambda, a, L)\frac{n \log\log n}{\log n}\right\}$$

and

$$N(n,\lambda) < \exp\left\{nC + T^{**}(\lambda, a, L)\frac{n \log\log n}{\log n}\right\}$$

for $n \geqq 4$. T^* can be explicitly computed.

4. Unsynchronized memoryless channels with feedback

We define $I(n)$, $s(n)$, $\tilde{X}$, $\tilde{Y}$ as in Section 3. Let

$$(4.1) \qquad \tilde{F} = \{f_{I(n)} | f_{I(n)} = [f^1, f^2(Z^1), \cdots, f^{I(n)}(Z^1, \cdots, Z^{I(n)-1})]\}$$

be the set of all possible vector valued encoding functions with $I(n)$ components. Using Definition (1.9) for $P(\bar{y}|f)$, we define the transition matrix of a synchronized channel J as follows.

$$(4.2) \qquad w^*(\bar{y}|f) = \sum_{\bar{y}^1 \times \cdots \times \bar{y}^{I(n)} = \bar{y}} P((\bar{y}^1, \cdots, \bar{y}^{I(n)})|f)$$

for every $f \in \tilde{F}$, $\bar{y} \in \tilde{Y}$. For every $\tilde{f}_m = (\tilde{f}^1, \cdots, \tilde{f}^m) \in \tilde{F}_m = \Pi_1^m \tilde{F}$ the transition probabilities of J are defined by

$$(4.3) \qquad Q(\bar{y}_m | \tilde{f}_m) = \prod_{t=1}^{m} w^*(\bar{y}^t | \tilde{f}^t) \quad \text{for every } \bar{y}_m = (\bar{y}^1, \cdots, \bar{y}^m) \in \bar{Y}_m$$

and every $m = 1, 2, \cdots$. J is a d.m.c. with input alphabet $\tilde{F}$ of size less than $(a^{a^{(L+1)I(n)}})^{I(n)}$ and output alphabet $\tilde{Y}$ of size A less than $a^{(L+1)I(n)}$

The following lemma, due to Muroga [9], allows us to restrict ourselves to a suitable subset $\tilde{F}^*$ of $\tilde{F}$, where $|\tilde{F}^*| < a^{(L+1)I(n)}$.

Lemma 4.1. Let w be an $a \times b$-stochastic matrix, which serves as the transition matrix of a d.m.c., and let ρ be the rank of w. Then there exists a $\rho \times b$-submatrix w' of w such that the corresponding d.m.c. has the same capacity.

Replacing $\tilde{F}$ by $\tilde{F}^*$, we can proceed as in Section 3 to obtain a lower bound on $N^f(n, \lambda)$, the maximal N for which an (n, N, λ)-code exists for the u.c.f. Denoting the capacity of J by $C^f_{I(n)}$ we obtain, as in (3.8), that

$$(4.4) \qquad N^f(n,f) > \exp\{nC^f_{I(n)} - K(\lambda)A^4[s(n)]^{\frac{1}{2}} - I(n)(\log a) - M\}.$$

In order to give an upper bound on $N^f(n,\lambda)$ we make use of two results, which we state as lemmas.

Lemma 4.2. [(a) is due to Shannon, (b) is due to Kemperman and Kesten. The topic is discussed and referenced in [8], pages 49–55.]

(a) The capacity of a d.m.c. remains unchanged under feedback.

(b) The maximal length of an (n,N,λ)-code for a discrete memoryless channel with feedback and capacity C^* is smaller than $\exp\{nC^* + \bar{K}(\lambda,w)\sqrt{n}\}$ for $n = 1,2,\cdots$, where $\bar{K}(\lambda)$ depends on λ and the c.p.f. w (strong converse for d.m.c.f. [8], Theorem 4.9.1).

Lemma 4.3. For an u.c.f. it is sufficient to consider only encodir g functions $f_n = (f^1, f^2(Z^1), \cdots, f^t(Z^1,\cdots,Z^{t-1}), \cdots, f^n(Z^1,\cdots,Z^{n-1}))$ for which $f^t(Z^1,\cdots,Z^{t-1}) = g^t(Z^1Z^2\cdots Z^{t-1})$ for $t = 1,2,\cdots,n$. More explicitly, for any code $\{(f_n(i),A_i) \mid i \cdots, =1,N\}$ we can find a set $\{g_n(i) \mid i = 1,\cdots,N\}$ such that $P(A_i \mid g_n(i)) \geqq P(A_i \mid f_n(i))$ for $i = 1,\cdots,N$.

Proof. Let $A \subset \bar{Y}$ and f_n^* be such that $P(A \mid f_n^*) \geqq P(A \mid f_n)$ for all $f_n \in F_n$. We shall find a $g_n = (g^1, g^2(Z^1), \cdots, g^n(Z^1\cdots Z^{n-1}))$ such that $P(A \mid g_n) = P(A \mid f_n^*)$. Denote by $\bar{y}A$ the sequences in A which start with $\bar{y}$. The optimality of f_n^* implies that

$$P(\bar{y}^1\cdots\bar{y}^{n-1}A \mid f^{*n}(\bar{y}^1,\cdots,\bar{y}^{n-1})) \text{ is maximal for every } \bar{y}^1,\cdots,\bar{y}^{n-1}. \tag{4.5}$$

Consequently, we can find a $g^n(\bar{y}^1\cdots\bar{y}^{n-1})$ such that $P(\bar{y}^1\cdots\bar{y}^{n-1}A \mid g^n(\bar{y}^1\cdots\bar{y}^{n-1}))$ is maximal. Now f^{*n-1} is such that

$$\begin{aligned} &\sum_{\bar{y}^{n-1}} P(\bar{y}^1\cdots\bar{y}^{n-1}A \mid f^{*n}(\bar{y}^1,\cdots,\bar{y}^{n-1}))P(\bar{y}^{n-1} \mid f^{*n-1}(\bar{y}^1,\cdots,\bar{y}^{n-2})) \\ &\text{is maximal for every } \bar{y}^1,\cdots,\bar{y}^{n-2}. \end{aligned} \tag{4.6}$$

Repeated iteration of the earlier argument yields Lemma 4.3.

We now give an upper bound on $N^f(n,\lambda)$. Let $s^*(n) = [n/I(n)]$ and $n^* = s^*(n)I(n)$. Obviously $n^* - n \leqq I(n)$. An (n,N,λ)-code for a channel can be modified into an (n^*,N,λ)-code for the channel. Map the encoding functions

$$g_{n^*}(i) = (g^1, g^2(Z^1),\cdots, g^{n^*}(Z^1\cdots Z^{n^*-1}))$$

into

$$\tilde{g}_{s^*(n)}(i) = ([g^1, g^2(Z^1),\cdots, g^{I(n)}(\cdots)],[g^{I(n)+1}(\cdots),\cdots, g^{2I(n)}(\cdots)],\cdots)$$

and define

$$A_i^* = \{(\bar{y}^1,\cdots,\bar{y}^{s^*(n)}) \mid \bar{y}^1\cdots\bar{y}^{s^*(n)} \in A_i\}.$$

It follows from Lemma 4.3 that

$$(4.7) \qquad \{(\tilde{g}_{s^*(n)}(i), A_i^*) \mid i = 1, \cdots, N\}$$

is an $(s^*(n), N, \lambda)$-code for channel J with feedback. It follows from (4.7) and Lemma 4.2 that

$$(4.8) \qquad N^f(n,\lambda) < \exp\{nC^f_{I(n)} + I(n)(\log a) + \bar{K}(\lambda, w^*)[s^*(n)]^{\frac{1}{2}}\}.$$

Comparing (4.8) with its analog (3.9) we notice one essential difference. In (3.9) $K(\lambda)$ does not depend on $I(n)$, but in (4.8) $\bar{K}$ depends on $I(n)$ because w^* depends on $I(n)$. We now prove that C^f_n converges to a limit C^f and give an estimate on the speed of approximation without using (4.4) and (4.8). The arguments we are using can also be used to give a different proof for Theorem 3.1.

We make use of

Lemma 4.4. Let X, Y, Z be finite sets, ϕ a mapping from Y onto Z and w a c.p.f. with input alphabet X and output alphabet Y. Define $w^{**}(z \mid x) = \sum_{\phi(y)=z} w(y \mid x)$ and denote by $S_\phi(z)$, $z \in Z$, the number of solutions of $\phi(y) = z$, $y \in Y$. The following inequality holds:

$$\left| \max_p \sum_{x \in X} \sum_{y \in Y} p(x) w(y \mid x) \log \frac{w(y \mid x)}{\sum_{x \in X} p(x) w(y \mid x)} \right.$$
$$\left. - \max_p \sum_{x \in X} \sum_{z \in Z} p(x) w^{**}(z \mid x) \log \frac{w^{**}(z \mid x)}{\sum_{x \in X} p(x) w^{**}(z \mid x)} \right|$$
$$\leqq \max_{z \in Z} \log S_\phi(z).$$

For a proof see [2], (4.1)–(4.5).

With $n' = s(n) \cdot I(n)$ we have $n' + I(n) \geqq n \geqq n'$. This, together with the obvious inequality $nC^f_n \geqq n'C^f_{n'}$ implies

$$(4.9) \qquad \left| nC^f_n - n'C^f_{n'} \right| \leqq I(n) \log a.$$

It follows from Lemma 4.4 that

$$(4.10) \qquad \left| n'C^f_{n'} - n'C^f_{I(n)} \right| \leqq \log \binom{(L+1)n' + s(n)}{s(n)}.$$

Choosing $I(n) = [\log n]$ we get

$$(4.11) \qquad \left| n'C^f_{n'} - [\log n] s(n) C^f_{[\log n]} \right| \leqq K(a,L) \cdot n \cdot \frac{\log \log n}{\log n}$$

for some constant $K(a,L)$. We obtain from (4.9) and (4.11) that

$$(4.12)\qquad \left|C_n^f - C_{[\log n]}^f\right| \leqq K_1(a,L)\frac{\log\log n}{\log n}$$

for $n = 4,5,\cdots$ and some constant $K_1(a,L)$. Choosing $l(n) = 2[\log n]$ we obtain

$$(4.13)\qquad \left|C_n^f - C_{2[\log n]}^f\right| \leqq K_2(a,L)\frac{\log\log n}{\log n}$$

for $n = 4,5,\cdots$ and some constant K_2. From (4.12), (4.13) we obtain

$$(4.14)\qquad \left|C_{[\log n]}^f - C_{2[\log n]}^f\right| \leqq (K_1 + K_2)\cdot\frac{\log\log n}{\log n}.$$

Repeating the argument used in (3.19)–(3.23) we get that $\lim_{n\to\infty} C_n^f$ exists and that the following theorem holds.

Theorem 4.1. Let $C^f = \lim_{n\to\infty} C_n^f$. There exists a $K(a,L)$ such that

$$\left|C_n^f - C^f\right| \leqq K\,\frac{\log n}{n}$$

for $n = 2,\cdots$.

From (4.4), (4.8) and Theorem 4.1 we obtain

Theorem 4.2. (Coding theorem and strong converse (within ε), for the u.c.f.) For any λ, $0 < \lambda < 1$, and any $n \geqq 4$,

$$N^f(n,\lambda) > \exp\left\{nC^f - K^*(\lambda,a,L)\cdot n\frac{\log\log n}{\log n}\right\}.$$

For any λ, $0 < \lambda < 1$, and any $\varepsilon > 0$, one can compute $n_0(\lambda,\varepsilon)$ such that, for all $n \geqq n_0(\lambda,\varepsilon)$,

$$N^f(n,\lambda) < \exp\{n(C^f + \varepsilon)\}.$$

5. Unsynchronized cômpound channels

Recall the definition of an u.c.c. given in (1.4), (1.5), (1.6). We also make the following assumption.

(5.1) If $\bar{y} \in \bar{Y}$ is such that $w(\bar{y}\,|\,x\,|\,s) > 0$ for some $x \in X$ and some $s \in S$, then $0 \leqq l(\bar{y}) \leqq L$.

Define

$$(5.2)\qquad \bar{C}_n = n^{-1}\sup_{p_n}\inf_{s\in S}\sum_{x_n\in X_n}\sum_{\bar{y}\in\bar{Y}} p_n(x_n)P(\bar{y}\,|\,x_n\,|\,s)$$

$$\times \log\frac{P(\bar{y}\,|\,x_n\,|\,s)}{\sum_{x\ \in X_n} p_n(x_n)P(\bar{y}\,|\,x_n\,|\,s)},$$

the supremum being taken over the set of all probability distributions p_n on X_n. Denote the maximal length of an (n,N,λ)-code for $\bar{S}$ by $\bar{N}(n,\lambda)$. Our aim is to prove the following two theorems.

Theorem 5.1. One can give explicitly a function $T^*(a,L)$ such that $|\bar{C}_n - \bar{C}| \leqq T^* n^{-1} \log n$ for $n = 2,\cdots$.

Theorem 5.2. (Coding theorem and strong converse for the u.c.c.) For any λ, $0 < \lambda < 1$, and $n \geqq 4$,

$$N(n,\lambda) > \exp\left\{n\bar{C} - T^{**}(\lambda,a,L)\,\frac{n \log\log n}{\log n}\right\}$$

and

$$N(n,\lambda) < \exp\left\{n\bar{C} + T^{**}(\lambda,a,L)\,\frac{n \log\log n}{\log n}\right\}.$$

T^{**} can be explicitly computed and $\bar{C} = \lim_{n\to\infty} \bar{C}_n$.

The proofs are based on a combination of the ideas used in the proofs of Theorems 3.1, 3.2 and the ideas used for the proof of the coding theorem for compound channels ([6] and [8], Chapter 4).

We define $I(n)$, $s(n)$, $\tilde{X}$, $\bar{Y}$ as in Section 3. The transition matrix for the synchronized channel $J(s)$, $s \in S$, is given by

$$(5.3) \qquad w^*(\bar{y}\,|\,\tilde{x}\,|\,s) = \sum_{\bar{y}^1\cdots\bar{y}^{I(n)} = \bar{y}} P((\bar{y}^1,\cdots,\bar{y}^{I(n)})\,|\,(x^1,\cdots,x^{I(n)})\,|\,s)$$

for every $\tilde{x} = (x^1,\cdots,x^{I(n)})$, $\bar{y} \in \bar{Y}$. For every $\tilde{x}_m = (\tilde{x}^1,\cdots,\tilde{x}^m) \in \tilde{X}_m$ the transition probabilities of $J(s)$ are defined by

$$(5.4) \qquad Q(\bar{y}_m\,|\,\tilde{x}_m\,|\,s) = \prod_{t=1}^{m} w^*(\bar{y}^t\,|\,\tilde{x}^t\,|\,s)$$

for every $\bar{y}_m = (\bar{y}^1,\cdots,\bar{y}^m) \in \bar{Y}_m$ and every $m = 1,2,\cdots;\ s \in S$.

We now treat the u.c.c. $\bar{S}$ by comparison with the compound channel $\bar{J} = \{J(s)\,|\,s \in S\}$. $\bar{J}$ has capacity $I(n)\cdot\bar{C}_{I(n)}$ ([8], Theorem 4.3.1, 4.4.1).

First we give an upper bound on $\bar{N}(n,\lambda)$. Let $s^*(n) = [n/I(n)]$ and $n^* = s^*(n)I(n)$, as in Section 3. Obviously $n^* - n \leqq I(n)$. An (n,N,λ)-code for $\bar{S}$ can be modified into an (n^*,N,λ) code for $\bar{S}$. Denote this code by

$$\{(u_i = (u_i^1,\cdots,u_i^{n^*}), A_i)\,|\,i = 1,\cdots,N\}.$$

Define

$$v_i = (v_i^1,\cdots,v_i^{s^*(n)}),$$

where

$$v_i^t = (u_i^{I(n)(t-1)+1},\cdots,u_i^{I(n)t}) \in \tilde{X}^t$$

for

$$t = 1, \cdots, s^*(n),$$

and

$$B_i = \{(\bar{y}^1, \cdots, \bar{y}^{s^*(n)}) \,|\, \bar{y}^1 \cdots \bar{y}^{s^*(n)} \in A_i\}.$$

Then, $\{(v_i, B_i) \,|\, i = 1, \cdots, N\}$ is an $(s^*(n), N, \lambda)$-code for $\bar{J}$. The strong converse for compound channels gives

$$(5.5) \qquad \bar{N}(n, \lambda) < \exp\{n\bar{C}_{I(n)} + (\log a) \cdot I(n) + K(\lambda)A^4\sqrt{n}\}.$$

We now derive a lower bound on $\bar{N}(n,\lambda)$ to be given in (5.21). Consider the set of all stochastic matrices with $a^{I(n)}$ rows and $A = a^{(L+1)I(n)}$ columns with elements which are integral multiples of $2^{-[s(n)]^{\frac{1}{2}}}$ except for perhaps the last element in every row. Let Γ' be any subset of this set of matrices, with index set S' and of minimal cardinality, which satisfies the following:

$$(5.6) \qquad \begin{aligned} &\text{For any } w^*(\cdot \,|\, \cdot \,|\, s),\ s \in S, \text{ there exists a } w^*(\cdot \,|\, \cdot \,|\, s') \in \Gamma' \\ &\text{such that } |w^*(j \,|\, i \,|\, s) - w^*(j \,|\, i \,|\, s')| \leqq A \cdot 2^{-[s(n)]^{\frac{1}{2}}} \\ &\text{for } i = 1, \cdots, a^{I(n)};\ j = 1, \cdots, A. \end{aligned}$$

We shall say that $w^*(\cdot \,|\, \cdot \,|\, s')$ approximates $w^*(\cdot \,|\, \cdot \,|\, s)$. Define $\pi'(s)$ by

$$(5.7) \qquad \pi'_j(s) = \sum_i \pi_i w^*(j \,|\, i \,|\, s) \text{ for } s \in S \text{ and } s \in S'.$$

Lemma 5.1. If $w^*(\cdot \,|\, \cdot \,|\, s')$ approximates $w^*(\cdot \,|\, \cdot \,|\, s)$, then, for any π,

$$|H(\pi'(s)) - H(\pi'(s'))| \leqq A^2 \cdot 2^{-[s(n)]^{\frac{1}{2}}/2}$$

$$\Big| \sum_i \pi_i H(w^*(\cdot \,|\, i \,|\, s) - \sum_i \pi_i H(w^*(\cdot \,|\, i \,|\, s'))\Big| \leqq A^2 \cdot 2^{-[s(n)]^{\frac{1}{2}}/2}.$$

This follows from an easy calculation (see for instance the proof of Lemma 4.2.1 in [8], Chapter 4).

Lemma 5.2. Let b be a positive constant. Choose $I(n) = K \cdot \log n$, K such that $a^{(L+1)K \log n} \leqq n^{1/8}$ and let $w^*(\cdot \,|\, \cdot \,|\, s')$ approximate $w^*(\cdot \,|\, \cdot \,|\, s)$. Let $u \in \tilde{X}_{s(n)}$ and $V \subset \tilde{Y}_{s(n)}$, such that $Q(V \,|\, u \,|\, s) > b$. Then

$$\left| \frac{Q(V \,|\, u \,|\, s')}{Q(V \,|\, u \,|\, s)} - 1 \right| < a_n,$$

where $a_n \to 0$ as $n \to \infty$, and a_n depends only on b and n and not on u, V, or $w^*(\cdot \,|\, \cdot \,|\, s)$.

This is a slight modification of Lemma 4.2.2 of [8]. There the alphabet was kept fixed, while here its size increases with n, but so slowly that the proof given there carries over to the present case without any essential change.

We now introduce an u.m.c. $F(s')$ with transition probabilities $F(\cdot \mid \cdot s')$ given by

$$F(\bar{y} \mid \tilde{x}_{s(n)} \mid s') = \sum_{\bar{y}^1 \cdots \bar{y}^{s(n)} = \bar{y}} \prod_{t=1}^{s(n)} w^*(\bar{y}^t \mid \tilde{x}^t \mid s'). \tag{5.8}$$

Let v be the image of $\tilde{x}_{s(n)}$ under the canonical mapping from $\tilde{X}_{s(n)}$ onto $X_{n'}$, where $n' = s(n) \cdot I(n)$ as in Section 3. For $V \subset \bar{Y}$ define

$$B = \{(\bar{y}^1, \cdots, \bar{y}^{s(n)}) \mid \bar{y}^1 \cdots \bar{y}^{s(n)} = \bar{y} \in V\},$$

then

$$F(V \mid \tilde{x}_{s(n)} \mid s') = Q(B \mid \tilde{x}_{s(n)} \mid s').$$

Suppose $Q(B \mid \tilde{x}_{s(n)} \mid s') > b$, and $w^*(\cdot \mid \cdot \mid s')$ approximates $w^*(\cdot \mid \cdot \mid s)$. Then, according to Lemma 5.2, $Q(B \mid \tilde{x}_{s(n)} \mid s) > b$ for n sufficiently large and therefore also

$$P(V \mid v \mid s) > b. \tag{5.9}$$

It suffices therefore to prove a coding theorem for $\{F(\cdot \mid \cdot \mid s') \mid s' \in S'\}$.

We shall say that $\bar{y} \in \bar{Y}$ is *generated* by $u \in \tilde{X}_{s(n)}$ with respect to the u.m.c. $F(\cdot \mid \cdot \mid s')$, if there exist $\bar{y}^1, \cdots, \bar{y}^{s(n)} \in \tilde{Y}$ such that $\bar{y} = \bar{y}^1 \cdots \bar{y}^{s(n)}$ and $(\bar{y}^1, \cdots, \bar{y}^{s(n)})$ is generated by u with respect to the channel $J(s')$ in the sense of (2.4). Define $G(u \mid s')$, $G(\pi \mid s')$, $\pi_m'(s')$, $B(w^*(s') \mid \pi)$, $B(w^*(s') \mid u)$ for channel $J(s')$ and $G^*(u \mid s')$, $G^*(\pi \mid s')$, $\pi_m^{*\prime}(s')$, $B^*(w^*(s') \mid \pi)$, $B^*(w^*(s') \mid u)$ for channel $F(\cdot \mid \cdot \mid s')$ as in Section 3.

Lemma 5.3. If $v \in G(\pi \mid s_0) \cap G(\pi \mid s_{00})$, then

$$\left| H(\pi'(s_0)) - H(\pi'(s_{00})) \right| < \frac{K_4 A^4 (1+\delta)^2}{[s(n)]^{\frac{1}{2}}}, \tag{5.10}$$

with K_4 a constant.

A proof is given in [8], Chapter 4, (4.3.10) to (4.3.18).

Now we are ready to prove the coding theorem. Without loss of generality we assume that $\lambda < \frac{1}{2}$. Let $\lambda' < \lambda$, $\lambda' > 0$. It follows from Lemma 5.2 and (5.9) that, for some $n \geqq n_0(\lambda, \lambda')$, a code with probability of error $\leqq \lambda'$ for $\bar{F} = \{F(\cdot \mid \cdot \mid s') \mid s' \in S'\}$ corresponds to a code with probability of error $\leqq \lambda$ for $\bar{S}$. Let $\delta > 2A$ be sufficiently large so that Lemma 3.2 holds with $\varepsilon' \leqq \frac{1}{4}\lambda'$. Let $\{(u_i, A_i) \mid i = 1, \cdots, N\}$ be a code for $\bar{F}$ such that the following conditions are satisfied:

(a) $u_i \in \tilde{X}_{s(n)}(\pi)$, $i = 1, \cdots, N$.

(b) $A_i = \bigcup_{s \in S'} G^*(u_i \mid s) - \bigcup_{j=1}^{i-1} A_j$.

(c) $F(A_i \mid u_i \mid s') \geqq 1 - \lambda'$, $i = 1, \cdots, N$ and $s' \in S'$.

(d) The code is maximal in the sense that it is impossible to add an element (u_{N+1}, A_{N+1}) such that (u_{N+1}, A_{N+1}) satisfies (a), (b), and (c) for $i = N + 1$.

In order to obtain a lower bound on $\left|\bigcup_{i=1}^{N} A_i\right|$ we proceed as in the proof of the coding theorem in Section 3. We define sets B_i, $i = 1,\cdots,N$, as in Section 3. Then $\{(u_i,B_i)\,|\, i = 1,\cdots,N\}$ is a code with maximal error probability λ' for channel $\bar{J}' = \{J(s')\,|\, s' \in S'\}$. As in Section 3 one can show that it is impossible to prolong the code by (u_{N+1},B_{N+1}), where $u_{N+1} \in \tilde{X}_{s(n)}(\pi)$, $B_{N+1} \subset \bigcup_{s' \in S'} G(u_{N+1}\,|\, s')$, and such that the resulting code has maximal error probability λ' for channel $\bar{J}'$. This implies that for every $u \in \tilde{X}_{s(n)}(\pi)$ there exists an $s_0 \in S'$ such that

$$Q\left(\bigcup_{i=1}^{N} B_i \cap G(u\,|\, s_0)\,\middle|\, u\,\middle|\, s_0\right) > \tfrac{3}{4}\lambda'.$$

Since $|S'| \leqq 2^{A^2\sqrt{n}}$, Lemma 2.1 implies the existence of an $s_{00} \in S'$ and a set $R \subset \tilde{X}_{s(n)}(\pi)$ such that

$$\pi_{s(n)}(R) \geqq \tfrac{3}{4} a^{-A^2\sqrt{n}}, \tag{5.11}$$

and such that, for every $u \in R$,

$$Q\left(\bigcup_{i=1}^{N} B_i \cap G(u\,|\, s_{00})\,\middle|\, u\,\middle|\, s_{00}\right) > \tfrac{3}{4}\lambda'. \tag{5.12}$$

Then (5.11), (5.12) imply that

$$\pi'_{s(n)}\left(\bigcup_{i=1}^{N} B_i \cap G(\pi\,|\, s_{00})\,\middle|\, s_{00}\right) > \frac{9\lambda'}{16} 2^{-A^2\sqrt{n}}\ . \tag{5.13}$$

It follows from (5.13) and Lemma 2.3 that

$$\left|\bigcup_{i=1}^{N} B_i \cap G(\pi\,|\, s_{00})\right| \geqq \frac{9\lambda'}{16}\cdot 2^{-A^2\sqrt{n}}\cdot \exp\{s(n)H(\pi^*(s_{00})) - K_1'(A,\delta)[s(n)]^{\frac{1}{2}}\}$$

and therefore

$$\left|\bigcup_{i=1}^{N} A_i \cap G^*(\pi|\, s_{00})\right| \geqq \frac{9\lambda'}{16}\cdot 2^{-A^2\sqrt{n}}\cdot \exp\{s(n)H(\pi^{*\prime}(s_{00})) - K_1'(A,\delta)\,[s(n)]^{\frac{1}{2}} - M\}. \tag{5.14}$$

We now give an upper bound on $\left|\left(\bigcup_{i=1}^{N} A_i\right) \cap G^*(\pi\,|\, s_{00})\right|$. Certainly,

$$\begin{aligned}\left|\left(\bigcup_{i=1}^{N} A_i\right) \cap G^*(\pi|\, s_{00})\right| &\leqq \left|\left(\bigcup_{i=1}^{N} B_i\right) \cap G(\pi\,|\, s_{00})\right| \\ &\leqq \sum_{i=1}^{N} \left|B_i \cap G(\pi|\, s_{00})\right|.\end{aligned} \tag{5.15}$$

If $v \in B_i \cap G(\pi\,|\, s_{00})$ then either $v \in G(u_i\,|\, s_{00})$ or there exists an $s^* \neq s_{00}$ such that $v \in G(u_i | s^*)$ and therefore $v \in G(\pi | s_{00}) \cap G(\pi | s^*)$. As a consequence of Lemma 5.3

$$|H(\pi'(s^*)) - H(\pi'(s_{00}))| < \frac{K_4 A^4 (1+\delta)^2}{[s(n)]^{\frac{1}{2}}}. \tag{5.16}$$

Let s^*_{00} be such that

$$\sum_i \pi_i H(w^*(\cdot | i | s^*_{00})) = \max\{\sum_i \pi_i H(w^*(\cdot | i | s)) \,|\, G(\pi | s) \cap G(\pi | s_{00}) \neq \emptyset\}. \tag{5.17}$$

(5.15), (5.16), (5.17), and Lemma 3.5 imply that

$$\begin{aligned} &\left| \left(\bigcup_{i=1}^{N} A_i \right) \cap G^*(\pi | s_{00}) \right| \\ &\leqq N \cdot 2^{A^2\sqrt{n}} \exp\{s(n) \sum_i \pi_i H(w^*(\cdot | i | s^*_{00})) + K_3(A,\delta)[s(n)]^{\frac{1}{2}}\}. \end{aligned} \tag{5.18}$$

(5.14) and (5.18) imply that

$$\begin{aligned} N &> \tfrac{3}{8}\lambda' 2^{-2A^2\sqrt{n}} \exp\{s(n)[H(\pi'(s_{00}) - \sum_i \pi_i H(w^*(\cdot | i | s^*_{00}))] \\ &\quad - \sqrt{n}(K_1(A,\delta) + K_3(A,\delta)) - M\} \\ &> \tfrac{3}{8}\lambda' \exp\{s(n)[H(\pi'(s^*_{00})) - \sum_i \pi_i H(w^*(\cdot | i | s^*_{00}))] \\ &\quad - \sqrt{n}[2A^2 + K_1(A,\delta) + K_3(A,\delta) + K_4 A^4 (1+\delta)^2] - M\} \end{aligned} \tag{5.19}$$

and with an optimal choice of π and $I(n) = K_5 \log n$, K_5 suitable

$$\bar{N}(n,\lambda) > \exp\left\{n \bar{C}_{K_5 . \log n} - T(a,L,\lambda) \frac{n \log\log n}{\log n}\right\}. \tag{5.20}$$

From (5.5) and (5.20) one can now derive Theorem 5.1 and Theorem 5.2 by the same arguments as those used in (3.19)–(3.23).

6. Continuous transmission with a fidelity criterion

In this section we give the improvement of Dobrushin's Theorem 3 which can readily be obtained by applying the results of Shannon [4] (see also Wolfowitz [7]) to Dobrushin's argument. Since the application is straightforward we give only the results. We assume familiarity with the definitions and notation of [2] as they relate to its Theorem 3, so as not to repeat several pages of a complex system. We replace Dobrushin's R by S to avoid confusion with Shannon's distortion function $R(\cdot)$, familiarity with which we assume.

Let n information digits be transmitted by a block code of word length n' over the unsynchronized channel of capacity C. Write $n = n'S$. For arbitrary positive ε and sufficiently large n the 2^n information sequences are spanned (see [7]) by $2^{nR(D)}$ sequences of the same space to within distortion $D+\varepsilon$, i.e., to within average Hamming distance $\leqq n(D+\varepsilon)$. Let D^* be the minimum of all D such that

$$\frac{C}{S} \geqq R(D).$$

Then one can generalize Theorem 3 and (2.10) of [2] as follows. For arbitrary positive ε there exists a transmission method such that

$$E(\gamma(s,t)\,|\,\bar{z}) \leqq (t-s)(D^*+\varepsilon),$$
$$E(\tilde{\gamma}(s,t)\,|\,\bar{z}) \leqq (t-s)(D^*+\varepsilon).$$

For n sufficiently large, the expected proportion of information symbols wrongly decoded, among the first n transmitted, is not less than

$$R^{-1}\left(\frac{C}{S}\right) - \varepsilon.$$

Acknowledgement

The authors thank Mr. Michael Ulrey for conscientious reading of the manuscript and helpful suggestions.

References

[1] Ahlswede, R. (1968) Beiträge zur Shannonschen Informationstheorie im Falle nichtstationärer Kanale. *Z. Wahrscheinlichkeitsth.* **10**, 1–42.

[2] Dobrushin, R. L. (1967) Shannon's theorems for channels with synchronization errors. *Problemy Peredači Informacii* **3**, 18–36. (*Problems of Information Transmission* 11–26).

[3] Golomb, S. W., Davey, J. R., Reed, J. S., Trees, H. L. and Stiffler, J. J. (1963) Synchronization. *IEEE Trans. Comm. Syst.* **11**, 481–491.

[4] Shannon, C. E. (1960) Coding theorems for a discrete source with a fidelity criterion. *Information and Decision Processes.* Ed. by Robert E. Machol. McGraw-Hill, New York.

[5] Wolfowitz, J. (1957) The coding of messages subject to chance errors. *Illinois J. Math.* **1**, 591–606.

[6] Wolfowitz, J. (1959) Simultaneous channels. *Arch. Rational Mech. Anal.* **4**, 371–386.

[7] Wolfowitz, J. (1965) Approximation with a fidelity criterion. *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob.* **1**, 565–573.

[8] Wolfowitz, J. (1964) *Coding Theorems of Information Theory.* Springer, Berlin–Heidelberg–New York.

[9] Muroga, S. (1953) On the capacity of a discrete channel. *Research and Development Data No.* 5, Nippon Telegraph and Telephone Public Corporation, Tokyo.