Reprinted from JOURNAL OF COMBINATORICS,
INFORMATION & SYSTEM SCIENCES

Vol. 5, No. 1, 10–35 (1980)
Printed in India

A Method of Coding and an Application to Arbitrarily Varying Channels

RUDOLF AHLSWEDE

Fakultät für Mathematik,
Universität Bielefeld, Universitätsstr, 1
4800 Bielefeld 1

1. INTRODUCTION

In the standard Shannon random coding method ([1], [2]) one derives bounds on the *expected average* error probability and then concludes that at least one code must be as good as the ensemble average. For high rates this leads to asymptotically optimal results (see [15]) and therefore in this case "most" codes in the ensemble must be close to the optimum. In the study of complex channel systems such as arbitrarily varying channels (AVC) it is useful to have good estimates on the proportion of codes in the ensemble which are good.

A first step in this direction was mentioned on page 174 of [5], where the following property of random coding is stated:

Selecting a code with positive rate below capacity at random for the discrete memoryless channel (DMC) fails to lead to a code with *average* error probability bounded by any constant λ, $0 < \lambda < 1$, with a probability which decreases double exponentially in the block length. Since our present results, which are for maximal errors, go considerably beyond this, we leave the proof of this earlier result as an advanced exercise.

For standard channels such as the DMC a simple application of the pigeon hole principle leads from an *average* error code to a *maximal* error code of essentially the same rate. This argument fails for big systems of channels such as the AVC and for the most robust model of such a channel ([9], [11]) one is forced to use the *maximal* error concept (see the discussion in Section 8 of [5]). It is therefore not just of academic interest to have a random coding method which leads *directly* to a code with small *maximal* error probability.

In this paper we present such a method (Theorem 1).

The key idea is to set up the random experiment in such a way that it is guaranteed that any two code words have at least a certain Hamming distance. This can be done in many ways and for the DMC there is also some freedom in the choice of the decoding rule. Here we use *threshold*

decoding (TD), because we need it anyhow for the treatment of the AVC. For those channels we use TD in *conjunction with other two-code word decoding rules* such as maximum likelihood decoding (MLD) or, more generally, maximum support decoding (MSD), a rule which we introduce.

TD makes certain saddle point arguments applicable, which enable us to determine the *maximal* error capacity for the AVC under a disjointness condition (S) (Theorem 2). Originally we used an I-divergence type regularity condition (L) (Theorem 3).

We can now derive Theorem 3 from Theorem 2. Both Theorems include the only known result in this direction ([11]) and they both completely settle the binary input, arbitrary output case. Without any regularity condition on the AVC the maximal error capacities are known ([17], [4]) for list codes of relatively small list size and in the presence of complete feedback. The results of [8], [10], [16], [5] are for average errors or randomized coding.

We draw attention to the solution of the famous zero-error capacity problem in the case of the pentagon ([12]) and the observation of [7] that this is a special case of the AVC coding problem for maximal errors.

Here our regularity conditions do not hold.

We would also like to draw attention to the forthcoming paper [6]. There it is emphasized that Information Theory deals essentially only with packing (channel coding), coloring (source coding), and covering (rate-distortion theory) of edges in certain hypergraphs. This way of looking at the basic combinatorial structures in abstract also enabled us to find the present method.

2. Statement of Results

§ 1 The Coding Method

Let us consider a DMC with alphabets $\mathcal{X}$, $\mathcal{Y}$ and transmission matrix w with *distinct row vectors.*

$I(p, w) = \sum_{x, y} p(x)w(y \mid x) \log \frac{w(y \mid x)}{q(y)}$, $q = p \cdot w$, denotes the mutual information, and p^* is a probability distribution (PD) on $\mathcal{X}$ such that $I(p^*, w) = \max_p I(p, w) = C$.

Let $U_1, \ldots, U_M$ be i.i.d. RV's with distribution

$$\text{Prob}\,(U_i = x^n) = p^*(x^n) = \prod_{t=1}^{n} p^*(x_t),$$

where $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n$.

Those U_i's describe the selection of code words in Shannon's random coding method ([1], [2]).

They are now used to define a new sequence $V_1, \ldots, V_M$ of RV's as follows

$$V_i = \begin{cases} U_i & \text{if } d(U_i, U_j) \geqslant \epsilon n \text{ for all } j \neq i \\ x_\infty & \text{otherwise.} \end{cases} \tag{2.1}$$

Here $\epsilon > 0$ and d denotes the Hamming distance. The equation $V_i = x_\infty$ is to be interpreted as "the i-th word selected is dropped".

Define for $x^n \in \mathcal{X}^{*n} = \{x^n : x^n \in \mathcal{X}^n, p^*(x^n) > 0\}$ and $\gamma > 0$

$$B(x^n) = \left\{y^n : \frac{1}{n} \log \frac{w(y^n \mid x^n)}{q^*(y^n)} \geqslant C - \gamma\right\}. \tag{2.2}$$

$\mathcal{M} = \{i : V_i \neq x_\infty, 1 \leqslant i \leqslant M\}$ is a random set of indices.

We choose $\{V_i : i \in \mathcal{M}\}$ as set of code words and define the decoding sets $\{D_i : i \in \mathcal{M}\}$ by the rule

$$D_i = D_i(U_1, \ldots, U_M) = B(V_i) - \bigcup_{j \in \mathcal{M} - \{i\}} B(V_j). \tag{2.3}$$

Finally for $i \in \mathcal{M}$

$$\lambda_i = \lambda_i(U_1, \ldots, U_M) = w(D_i^c \mid V_i) \tag{2.4}$$

denotes the error probability for code word V_i.

THEOREM 1. *There exists an $\epsilon_0 = \epsilon_0(w) > 0$, an positive function $\gamma_0(\epsilon)$, and a positive function $\tau(\epsilon, \gamma)$ with $\lim_{\epsilon, \gamma \to 0} \tau(\epsilon, \gamma) = 0$ such that the coding method described in* (2.1)–(2.4), *with M specified by*

$$M = \exp\{(C - \tau(\epsilon, \gamma))n\},$$

has the property

$$\text{Prob}\left(|\mathcal{M}| \geqslant \tfrac{1}{2}M, \max_{i \in \mathcal{M}} \lambda_i \leqslant \lambda\right) \geqslant 1 - \exp\{-e^{\rho(\epsilon, \gamma, \lambda)n}\} \tag{2.5}$$

for $0 < \epsilon \leqslant \epsilon_0$, $0 < \gamma \leqslant \gamma_0(\epsilon)$.

Here, $\rho(\epsilon, \gamma, \lambda) > 0$ for any $\lambda \in (0, 1)$.

Remark. In order to achieve the superexponential bound (2.5) one cannot use the U_i's as code words, because already the probability that two code words are equal is only exponentially small.

§ 2 An Application to Arbitrarily Varying Channels

A discussion about the channel model can be found in Section 8 of [5], in chapter 6 of [14] and also in [13]. Here we consider the most robust model of an AVC in which the choice of states may depend on the word sent. In this case one is forced to consider the *maximal* decoding error

(see [5]). A first attempt to find the capacity of this channel was made in [9]. There a necessary and sufficient condition for the capacity to be positive is given in case there are finitely many states (see Lemma 8 in Section 4). In [11] the binary output case was completely solved, that is, a computable formula for the capacity was found.

We give now the formal description of the channel, the coding problem and the results.

Let $\mathcal{S}$ be an arbitrary set and let $\mathcal{C} = \{w(\cdot|\cdot|s) : s \in \mathcal{S}\}$ be a set of stochastic $|\mathcal{X}| \times |\mathcal{Y}|$-matrices. For every $s^n = (s_1, \ldots, s_n) \in \mathcal{S}^n = \prod_1^n \mathcal{S}$ we define transmission probabilities $W(\cdot|\cdot|s^n)$ by

$$W(y^n|x^n|s^n) = \prod_{t=1}^{n} w(y_t|x_t|s_t) \tag{2.6}$$

for all $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n$, $y^n \in \mathcal{Y}^n$, and all $n = 1, 2, \ldots$.

If $\mathcal{C}^n = \{W(\ |\cdot|s^n) : s^n \in \mathcal{S}^n\}$, then we call the sequence $\left(\mathcal{C}^n\right)_{n=1}^{\infty}$ an arbitrarily varying channel (AVC) and denote it by $\mathcal{A}$.

An (n, N, λ)-code for $\mathcal{A}$ is a system $\{(u_i, D_i) : 1 \leqslant i \leqslant N\}$, where $u_i \in \mathcal{X}^n$, $D_i \subset \mathcal{Y}^n$ for $i = 1, 2, \ldots, N$; $D_i \cap D_j = \emptyset$ for $i \neq j$, and

$$\max_{1 \leqslant i \leqslant N} \max_{s^n \in \mathcal{S}^n} W(D_i^c\ |u_i|s^n) \leqslant \lambda. \tag{2.7}$$

A number C is the capacity of $\mathcal{A}$, if for any $\epsilon > 0$, any λ, $0 < \lambda < 1$, and for all sufficiently large n there exists an $(n, \exp\{(C - \epsilon)n\}, \lambda)$-code and there does not exist an $(n, \exp\{(C + (\epsilon)n\}, \lambda)$ code.

$\mathcal{R}(x)$, $x \in \mathcal{X}$, denotes the closed convex hull of the set $\{w(\cdot|x|s) : s \in \mathcal{S}\}$ of probability distributions (PD) on $\mathcal{Y}$. The set of matrices

$$\bar{\bar{\mathcal{C}}} = \{(w(y \mid x))_{x \in \mathcal{X},\, y \in \mathcal{Y}} : w(\cdot \mid x) \in \mathcal{R}(x),\ x \in \mathcal{X}\} \tag{2.8}$$

is called the row-convex closure of the set $\mathcal{C}$. It shall be parametrized by an index set $\bar{\bar{\mathcal{S}}}$.

Denote by $\mathcal{P}$ the set of all PD's on $\mathcal{X}$. Since $\mathcal{P}$ and $\bar{\bar{\mathcal{C}}}$ are compact in the supremum-norm topology and $I(p, w)$ is continuous in both variables the definition

$$\bar{\bar{C}} = \max_{p \in \mathcal{P}} \min_{w \in \bar{\bar{\mathcal{C}}}} I(p, w) \tag{2.9}$$

is meaningful. Also, since $I(p, w)$ is convex in p and concave in w by the Min-Max Theorem

$$\bar{\bar{C}} = \min_{w \in \bar{\bar{\mathcal{C}}}} \max_{p \in \mathcal{P}} I(p, w). \tag{2.10}$$

Let $(p^*, w^*) \in \mathcal{P} \times \bar{\bar{C}}$ be a pair for which $I(p^*, w^*) = \bar{\bar{C}}$.

THEOREM 2. *If the AVC $\mathcal{A}$ satisfies the condition*

$$\mathcal{R}(x) \cap \mathcal{R}(x') = \emptyset \text{ for all } x, x' \in \mathcal{X} \text{ with } x \neq x', \tag{S}$$

then its capacity C equals $\bar{\bar{C}}$.

One readily verifies that in case $\mathcal{X} = \{1, 2\}$ $\mathcal{R}(1) \cap \mathcal{R}(2) \neq \emptyset$ implies $C = \bar{\bar{C}} = 0$. This and Theorem 2 imply

COROLLARY. *The capacity of an AVC with binary input alphabet equals $\bar{\bar{C}}$.*

THEOREM 3. *If the AVC $\mathcal{A}$ satisfies the condition*

$$\min_{w \in \bar{\bar{C}}} \sum_y w(y \mid x) \log \frac{w^*(y \mid x)}{w^*(y \mid x')} > 0 \text{ for all } x, x' \in \mathcal{X}^* \text{ with } x \neq x', \tag{L}$$

then its capacity C is given by

$$C = \bar{\bar{C}}. \tag{2.11}$$

Observe that (L) implies

$$\mathcal{R}(x) \cap \mathcal{R}(x') = \emptyset \text{ for all } x, x' \in \mathcal{X}^* \text{ with } x \neq x'. \tag{S*}$$

Since the letters $\mathcal{X}^*$ determine already $\bar{\bar{C}}$ and since $C \leqslant \bar{\bar{C}}$ (see [11]), substitution of $\mathcal{X}$ in Theorem 2 by $\mathcal{X}^*$ gives Theorem 3.

Remarks. 1. We have not completely analysed condition (L), but we have found three important AVC's for which it holds:

(a) For a w with distinct row vectors and $\epsilon = \epsilon(w)$ sufficiently small set $\mathcal{C} = \{w' : |w(y \mid x) - w'(y \mid x)| \leqslant \epsilon, x \in \mathcal{X}, y \in \mathcal{Y}\}$.

(b) $\mathcal{X} = \{1, 2\}$ and $\mathcal{C}$ such that $\mathcal{R}(1) \cap \mathcal{R}(2) = \emptyset$.

(c) $\bar{\bar{C}}$ contain an "information-extremal" matrix w_I, that is, $I(p, w_I) \leqslant I(p, w)$ for all $p \in \mathcal{P}$, $w \in \bar{\bar{C}}$; and w_I has positive capacity. Notice that by the data-processing theorem this includes the case where one matrix in $\bar{\bar{C}}$ is a degraded version of all the others.

2. In [11] the binary output case was solved. The case $|\mathcal{X}| \geqslant 2$ was reduced to the case $|\mathcal{X}| = 2$ by the elementary idea of using 2 "extremal letters". The idea works because of the special geometric properties of PD's on a 2 element set. Then the case $|\mathcal{X}| = |\mathcal{Y}| = 2$ was solved with the idea of a worst channel. This case is now a special case of the Corollary.

Theorem 3 also directly settles the case $|\mathcal{X}| \geqslant 2$, $|\mathcal{Y}| = 2$ because there exists a p^*, $I(p^*, w^*) = \bar{\bar{C}}$, concentrated on the 2 extremal letters.

It is interesting that the coding scheme of [11] is *universal*, that is, the decoding sets for a given set of code words can be defined without using the channel ([19]). For our more general cases no universal decoding method exists. The proof for this negative result is somewhat lengthy and therefore omitted here.

3. For the AVC $\mathcal{A}$ C is positive exactly when the following condition holds: (SP) there exist $x, x' \in \mathcal{X}$ with $\mathcal{R}(x) \cap \mathcal{R}(x') = \emptyset$.

This was proved in [9] for the case $|C| < \infty$. This restriction can easily be removed (see Lemma 8 in Section 4).

We give now 2 examples of channels for which (SP) holds, but (2.11) doesn't.

EXAMPLE 1. $\mathcal{X} = \mathcal{Y} = \{0, 1, 2, 3, 4\}$, $C = \{w(\cdot|\cdot|1), w(\cdot|\cdot|2)\}$, where

$$w(y|x|1) = 1 \Leftrightarrow y = x$$

$$w(y|x|2) = 1 \Leftrightarrow y = x + 1 \bmod 5.$$

It was shown in [7] that for this channel $C = C_0$, the zero-error capacity of the pentagon. Lovász has shown in [12] that $C_0 = \frac{1}{2} \log 5$. By an easy calculation $\overline{\overline{C}} = \log \frac{5}{2} > \frac{1}{2} \log 5$. That this is not just a pathology occurring for 0-1-matrices can be seen as follows:

Replace every output letter by two letters carrying the same total probability as the original letter.

The following observation is more interesting: $\overline{\overline{C}}$, for the C above, is defined by the sets $\mathcal{R}(i) = \{p : p(j) = 0 \text{ for } j \neq i, i + 1 \bmod 5\}$, $i = 0, 1, \ldots, 4$.

If one replaces them by $\tilde{\mathcal{R}}(i) = \{p : p \in \mathcal{R}(i), p(j) < 1 - \epsilon_j \text{ for all } j\}$ where $0 < \epsilon_j < \frac{1}{2}$, then for this new channel (S) holds and $C = \log \frac{5}{2}$.

EXAMPLE 2 (G. Dueck). $|\mathcal{X}| = |\mathcal{Y}| = 6$, $|\mathcal{S}| = 4$.

For $w_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $w_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ let

$$C = \left\{ \begin{pmatrix} w_j & 0 \\ 0 & w_i \end{pmatrix} : i, j \in \{1, 2\} \right\}$$

(SP) is satisfied, $\overline{\overline{C}} = 2 \log \frac{3}{2} > C = \log 2$.

The structure of condition (S) indicates that AVS's are to be classified according to the intersection properties of the sets $\mathcal{R}(x)$, $x \in \mathcal{X}$. One classification would be by a graph with vertex set $\mathcal{X}$, where x and x' are connected iff $\mathcal{R}(x) \cap \mathcal{R}(x') \neq \emptyset$. Example 1 thus leads to the pentagon. A finer classification would also consider common intersections of more

than 2 sets. In case $\bigcap_{x \in \mathcal{X}} \mathcal{R}(x) \neq \emptyset$ for instance $\overline{\overline{C}} = 0$, and vice versa. All this is still to be understood and may be called *general* zero-error capacity theory. A nice starting point might be

EXAMPLE 3.

$$\mathcal{C} = \left\{ \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \right\}.$$

Here $\mathcal{R}(1) \cap \mathcal{R}(2) \neq \emptyset$, $\mathcal{R}(2) \cap \mathcal{R}(3) \neq \emptyset$, $\mathcal{R}(1) \cap \mathcal{R}(3) = \emptyset$.

3. PROOF OF THEOREM 1

§ 1 AUXILIARY INEQUALITIES

For the analysis of the coding scheme described in Section 2 we use Bernstein's versions of Chebyshev's inequality. We give the proofs, because they are so brief.

LEMMA 1. *Let $T_1, \ldots, T_k$ be a sequence of discrete RV's, then*

(a) $$\text{Prob}\left\{\frac{1}{K}\sum_{i=1}^{K} T_i \geqslant a\right\} \leqslant e^{-\alpha a K} \prod_{i=1}^{K} \max_{t^{i-1}} E(\exp\{\alpha T_i\} \mid t^{i-1}).$$

Here $(T_{i-1} = t_{i-1}, \ldots, T = t_1)$ is abbreviated as t^{i-1} and $\alpha > 0$.

(b) $$\text{Prob}\left\{\frac{1}{K}\sum_{i=1}^{K} T_i \geqslant a\right\} \leqslant e^{-aK/2b} \prod_{i=1}^{K} \max_{t^{i-1}} E(1 + b^{-1}T_i \mid t^{i-1})$$

if $T_1, \ldots, T_K$ take values in $[0, b]$.

(c) $$\text{Prob}\left\{\frac{1}{K}\sum_{i=1}^{K} (T_i - ET_i) \geqslant a\right\} \leqslant \exp[(-\alpha a + \alpha^2 b^2)K]$$

for $0 < \alpha \leqslant \min\left(1, \frac{b^2}{2} e^{-2b}\right)$, if $T_1, \ldots, T_K$ are independent and take values in $[-b, b]$. Furthermore, the exponent is negative, if $\alpha < ab^{-2}$.

(d) $$\text{Prob}\left\{\frac{1}{K}\sum_{i=1}^{K} (T_i - ET_i) \leqslant a\right\} \leqslant \exp[(-\alpha a + \alpha^2 b^2)K]$$

for $\alpha < 0$, $|\alpha| \leqslant \min\left(1, \frac{b^2}{b} e^{-2b}\right)$, if $T_1, \ldots, T_k$ are independent and take values in $[-b, b]$. The exponent is negative if $a < 0$ and $|\alpha| < |a|b^{-2}$.

Proof. $$\text{Prob}\left\{\frac{1}{K}\sum_{i=1}^{K} T_i \geqslant a\right\} \leqslant e^{-\alpha a K} E\left(\exp\left\{\alpha \sum_{i=1}^{K} T_i\right\}\right)$$

$$\leqslant e^{-\alpha a K} \prod_{i=1}^{K} \max_{t^{i-1}} E(\exp\{\alpha T_i\} \mid t^{i-1}).$$

(b) is now readily established by using the expansion

$$E(\exp\{\alpha T_i\} \mid t^{i-1}) = \sum_{s=0}^{\infty} \frac{\alpha^s}{s!} E(T_i^s \mid t^{i-1}).$$

Since $T_i^s \leqslant b^{s-1}T_i$ for $s \geqslant 1$ we conclude that for $\alpha = (2b)^{-1}$

$$\sum_{s=0}^{\infty} \frac{\alpha^s}{s!} E(T_i^s \mid t^{i-1}) \leqslant 1 + (2b)^{-1} \sum_{s=1}^{\infty} \frac{1}{s!} E(T_i \mid t^{i-1})$$
$$\leqslant 1 + b^{-1}E(T_i \mid t^{i-1}) = E(1 + b^{-1}T_i \mid t^{i-1}).$$

We show now (c). Starting with

$$\text{Prob}\left\{\sum_{i=1}^{K} (T_i - ET_i) \geqslant aK\right\} \leqslant e^{-\alpha aK} E \exp\left\{\alpha \sum_{i=1}^{K} (T_i - ET_i)\right\}$$

we upper bound next

$$E \exp\{\alpha(T_i - ET_i)\} = \sum_{s=0}^{\infty} \frac{\alpha^s}{s!} E(T_i - ET_i)^s \leqslant 1+0+ \frac{\alpha^2}{2} b^2 + \sum_{s=3}^{\infty} \frac{\alpha^s}{s!}(2b)^s$$

$$\leqslant 1 + \frac{\alpha^2}{2} b^2 + \alpha^3 e^{2b\alpha} \quad \text{for } \alpha \leqslant 1$$

$$\leqslant 1 + \alpha^2 b^2 \quad \text{for } \alpha \leqslant \min\left(1, \frac{b^2}{2} e^{-2b}\right).$$

Since the T_i's are independent we get

$$\text{Prob}\left\{\sum_{i=1}^{K} (T_i - ET_i) \geqslant aK\right\} \leqslant e^{-\alpha aK}(1 + \alpha^2 b^2)^K \leqslant \exp\left[-\alpha a + \alpha^2 b^2)K\right].$$

The proof for (d) is almost identical. Just observe that for $\alpha < 0$

$$\text{Prob}\left\{\sum_{i=1}^{K} (T_i - ET_i) \leqslant aK\right\} \leqslant e^{-\alpha aK} E \exp\left\{\alpha \sum_{i=1}^{K} (T_i - ET_i)\right\}$$

and replace in the argument above α^3 by $|\alpha^3|$.

LEMMA 2 (Shannon). *Let p^* be such that $I(p^*, w) = C$, $q^* = p^* \cdot w$, then*

$$\sum_y w(y \mid x) \log \frac{w(y \mid x)}{q^*(y)} \begin{cases} = C & \text{for } x \text{ with } p^*(x) > 0 \\ \leqslant C & \text{otherwise.} \end{cases}$$

§ 2. THE RATE

We derive now an upper bound on Prob $(|\mathcal{M}| \leqslant \frac{1}{2}M)$, where $\mathcal{M}$ is as defined in Section 2.

Define

$$\hat{V}_i = \begin{cases} 1 & \text{if} \quad V_i = x_\infty \\ 0 & \text{if} \quad V_i \neq x_\infty \end{cases} \tag{3.1}$$

and notice that

$$\text{Prob}\,(|\mathcal{M}| \leqslant \tfrac{1}{2}M) = \text{Prob}\left(\sum_{i=1}^{M} \hat{V}_i \geqslant \tfrac{1}{2}M\right). \tag{3.2}$$

Let $S_\epsilon(x^n)$ be a Hamming sphere with radius ϵn and centre x^n. Clearly,

$$\hat{V}_i = 1 \Leftrightarrow U_i \in \bigcup_{j \neq i} S_\epsilon(U_j).$$

One can view the selection of code words as an iterative prodecure and upper bound $\sum_{i=1}^{M} \hat{V}_i$ by $\sum_{i=1}^{M} T_i + \sum_{i=1}^{M} 0_i$, where

$$T_i = \begin{cases} 1 & \text{if } U_i \in \bigcup_{j<i} S_\epsilon(U_j) \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

$$0_i = \begin{cases} 1 & \text{if } U_i \in \bigcup_{j>i} S_\epsilon(U_j) \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

Since $\text{Prob}\left(\sum_{i=1}^{M} \hat{V}_i \geqslant \frac{1}{2}M\right) \leqslant \text{Prob}\left(\sum_{i=1}^{M} T_i \geqslant \frac{1}{4}M\right) + \text{Prob}\left(\sum_{i=1}^{M} 0_i \geqslant \frac{1}{4}M\right)$ it suffices to upper bound $\text{Prob}\left(\sum_{i=1}^{M} T_i \geqslant \frac{1}{4}M\right)$, because $\text{Prob}\left(\sum_{i=1}^{M} T_i \geqslant \frac{1}{4}M\right)$ $= \text{Prob}\left(\sum_{i=1}^{M} 0_i \geqslant \frac{1}{4}M\right)$. For the Hamming spheres we have

$$|S_\epsilon(x^n)| \leqslant \binom{n}{n\epsilon} |\mathcal{X}|^{n\epsilon} \leqslant \exp\{(h(\epsilon) + \epsilon \log |\mathcal{X}|)\, n\}, \tag{3.5}$$

and if we choose ϵ such that

$$C - \tau(\epsilon, \gamma) + h(\epsilon) + \epsilon \log |\mathcal{X}| < H(p^*) - \epsilon, \tag{3.6}$$

then for any $\{u_j : 1 \leqslant j \leqslant M\} \subset \mathcal{X}^n$,

$$\left| \bigcup_{j=1}^{M} S_\epsilon(u_j) \right| \leqslant \exp\{(H(p^*) - \epsilon)n\}. \tag{3.7}$$

This and the source coding theorem imply

$$p^*\left(\bigcup_{j=1}^{M} S_\epsilon(u_j) \right) \leqslant \exp\{-g(\epsilon)n\}, \tag{3.8}$$

where $g(\epsilon) > 0$ for $\epsilon > 0$.

It follows from (3.8) that

$$\text{Prob}\,(T_i = 1 \mid T_{i-1} = t_{i-1}, \ldots, T_1 = t_1) \leqslant \exp\{-g(\epsilon)n\} \tag{3.9}$$

Now we apply Lemma 1 (a) with $\alpha = 1$ and obtain

$$\begin{aligned} \text{Prob}\left(\sum_{i=1}^{M} T_i \geqslant \tfrac{1}{4}M\right) &\leqslant \exp\,(-\tfrac{1}{4}M)(1 + \exp\{-g(\epsilon)(n+1)\})^M \\ &\leqslant \exp\{-\tfrac{1}{4} + \exp\{-g(\epsilon)(n+1)\}M\}. \end{aligned} \tag{3.10}$$

Since $g(\epsilon) > 0$, for $n > \dfrac{3}{g(\epsilon)}$ this is smaller than $\exp\{-\frac{1}{8}M\}$ and we have derived the superexponential bound

$$\text{Prob}\,(|\mathcal{M}| \leqslant \tfrac{1}{2}M) \leqslant 2 \cdot \exp\{-\tfrac{1}{8}M\} \tag{3.11}$$

§ 3. Two code word error probabilities for threshold decoding

The following two Lemmas play a key role in the analysis of our coding scheme. For the proof of Theorem 2 one needs analogous types of results.

LEMMA 3. *Let U be a RV with Prob $(U = x^n) = p^*(x^n)$. For $\mathcal{X}^{*n}$ and $B(x^n)$, defined in* (2.2),

(a) $\quad w(B(x^n) \mid x^n) \geqslant 1 - \exp\{-C_1(\gamma)n\}$ *for* $x^n \in \mathcal{X}^{*n}$,

where $C_1(\gamma) > 0$ *for* $\gamma > 0$.

(b) $Ew(B(U) \mid x^n) \leqslant \exp\{-(C - \gamma)n\}$, $x^n \in \mathcal{X}^{*n}$.

Actually, also the inequality $Ew(B(x^n) \mid U) \leqslant \exp\{-(C - \gamma)n\}$, $x^n \in \mathcal{X}^{*n}$, holds, but this is not used here.

Proof. (a) For $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^{*n}$ consider the independent RV's $Y_1, \ldots, Y_n$, where Prob $(Y_t = y) = w(y \mid x_t)$. The RV's

$$Z_t = \begin{cases} \log \dfrac{w(Y_t \mid x_t)}{q^*(Y_t)} & \text{if } w(Y_t \mid x_t) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.12}$$

are again independent and uniformly bounded by a constant b, say, where

$$b = \max(-\log w_{\min}, -\log q^*_{\min}),$$

$w_{\min} = \min\{w(y \mid x) : w(y \mid x) > 0\}$, $q^*_{\min} = \min\{q^*(y) : q^*(y) > 0\}$.

By Lemma 2 Prob $\left\{\sum_{t=1}^n Z_t \leqslant (C - \gamma)n\right\} = \text{Prob}\left\{\sum_{t=1}^n (Z_t - EZ_t) \leqslant -\gamma n\right\}$ and for $\alpha < 0$, $|\alpha| \leqslant \min\left(1, \dfrac{b^2}{2} e^{-2b}, \gamma b^{-2}\right)$ Lemma 1(d) implies that this probability is smaller than $\exp\{(\alpha\gamma + \alpha^2 b^2)n\}$ and that the exponent is negative.

(b) $Ew(B(U) \mid x^n) = \sum p^*(u) w(B(u) \mid x^n)$

$$= \sum_{y^n}^{u} w(y^n \mid x^n) \sum_u p^*(u) 1_{B(u)}(y^n).$$

It suffices to show that for every $y^n \in \mathcal{Y}^n$

$$\sum_u p^*(u) 1_{B(u)}(y^n) \leqslant \exp[-(C - \gamma)n].$$

Define $A(y^n) = \left\{u : \dfrac{1}{n} \log \dfrac{w(y^n \mid u)}{q^*(y^n)} \geqslant C - \gamma\right\}$ and notice that

$$\sum_u p^*(u) 1_{B(n)}(y^n) = p^*(A(y^n)). \tag{3.14}$$

Let now $\tilde{Z}_1, \ldots, \tilde{Z}_n$ be independent RV's with distribution

$$\text{Prob}\left(\tilde{Z}_i = \log \frac{w(y_t \mid x)}{q^*(y_t)}\right) = p^*(x).$$

Then

$$p^*(A(y^n)) = \text{Prob}\left(\sum_{t=1}^n \tilde{Z}_t \geqslant (C - \gamma)n\right)$$

$$\leqslant e^{-(C-\gamma)n} \prod_{t=1}^{n} E \exp\{\tilde{Z}_t\}$$

$$= e^{-(C-\gamma)n} \prod_{t=1}^{n} \sum_{x} p^*(x) \frac{w(y_t \mid x)}{q^*(y_t)}.$$

Since $\sum_x p^*(x)w(y_t \mid x) = q^*(y_t)$ the product is 1 and the result follows.

Remark. Inspection of the proof shows that $Q(B(U)) \leqslant e^{-(C-\gamma)n}$ for *any* PD Q on $\mathcal{Y}^n$.

LEMMA 4. *If w has distinct row vectors and if x^n, $x'^n \in \mathcal{X}^{*n}$ satisfy $d(x^n, x'^n) \geqslant \epsilon n$, then there exists a $\gamma_0(\epsilon)$: for $\gamma \leqslant \gamma_0$, $W(B(x'^n) \mid x^n) \leqslant \exp\{-C_2(\gamma)\epsilon n\}$, where $C_2(\gamma) > 0$ for $\gamma > 0$.*

Proof. Let $\beta \in (0, 1)$ be a number to be further specified below. Consider the sets $A_t = \{y : w(y \mid x'_t) = 0, \quad w(y \mid x_t) > 0\}$. Since for $y \in A_t$, $q^*(y) > 0$ we can conclude

$$B(x'^n) \subset \prod_{t=1}^{n} A_t^c. \tag{3.15}$$

Write $\{1, \ldots, n\} = I \cup J \cup K$, where

$$I = \{t : 1 \leqslant t \leqslant n, A_t \neq \emptyset\}, J = \{t : 1 \leqslant t \leqslant n, x_t = x'_t\},$$

and $$K = \{t : 1 \leqslant t \leqslant n, A_t = \emptyset, x_t \neq x'_t\}.$$

Clearly, if $|I| \geqslant \beta\epsilon n$, then by (3.15)

$$w(B(x'^n) \mid x^n) \leqslant (1 - w_{\min})^{\beta\epsilon n}. \tag{3.16}$$

We can therefore assume that $|I| < \beta\epsilon n$.

With the RV's $Y_1, \ldots, Y_n$ used in the proof of (a) of Lemma 3 we define now the RV's

$$\bar{Z}_t = \begin{cases} \log \dfrac{w(Y_t \mid x'_t)}{q^*(Y_t)} & \text{if } w(Y_t \mid x'_t) > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.17}$$

They are again independent and bounded by b.

For $0 < \alpha \leqslant \min\left(1, \dfrac{b^2}{2} e^{-2b}\right)$ it follows from Lemma 1(c) that

$$\text{Prob}\left\{\sum_{t=1}^{n} \bar{Z}_t \geqslant (C-\gamma)n\right\} = \text{Prob}\left\{\sum_{t=1}^{n} (\bar{Z}_t - E\bar{Z}_t) \geqslant \sum_{t=1}^{n} (C - \gamma - E\bar{Z}_t)\right\}$$

$$\leqslant \exp\left\{-\alpha \sum_{t=1}^{n} (C - \gamma - E\bar{Z}_t) + \alpha^2 b^2 n\right\}. \tag{3.18}$$

We consider now the quantity $\sum_{t=1}^{n} (C - \gamma - E\bar{Z}_t)$.

Since $E\bar{Z}_t = \sum_{y : w(y \mid x_t') > 0} w(y \mid x_t) \log \frac{w(Y \mid x_t')}{q^*(y)}$, we have

$$E\bar{Z}_t \leqslant b \text{ for } t \in I, \text{ because } |\bar{Z}_t| \leqslant b. \tag{3.19}$$

$$E\bar{Z}_t = C \text{ for } t \in J, \text{ by Lemma 2} \tag{3.20}$$

and finally, for $t \in K$,

$$\begin{aligned} E\bar{Z}_t &= \sum_y w(y \mid x_t) \log \frac{w(y \mid x_t')}{q^*(y)} \\ &= \sum_y w(y \mid x_t)\left(\log \frac{w(y \mid x_t)}{q^*(y)} - \log \frac{w(y \mid x_t)}{w(y \mid x_t')}\right). \end{aligned} \tag{3.21}$$

If D denotes the I-divergence and if we set $D_{\min} = \min_{x \neq x'} D(w(\cdot \mid x) \| w(\cdot \mid x'))$, which is positive because w has distinct row vectors, we derive from (3.21) and Lemma 2

$$E\bar{Z}_t \leqslant C - D_{\min} \quad \text{for } t \in K. \tag{3.22}$$

The inequalities (3.18), (3.19), and (3.22) imply

$$\sum_{t=1}^{n} (C - \gamma - E\bar{Z}_t) \geqslant (C - \gamma)n - b \cdot |I| - C|J| - (C - D_{\min})|K|. \tag{3.23}$$

Since $b \geqslant C \geqslant C - D_{\min}$ we conclude that

$$\begin{aligned} \sum_{t=1}^{n} (C - \gamma - E\bar{Z}_t) &\geqslant (C - \gamma)n - \beta\epsilon nb - (1 - \epsilon)nC - (1 - \beta)\epsilon n(C - D_{\min}) \\ &= (-\gamma - \beta\epsilon(b - C) + (1 - \beta)\epsilon D_{\min})n \\ &= f_2(\gamma, \beta, \epsilon)n, \end{aligned}$$

where $f_2(\gamma, \beta, \epsilon) > 0$ if γ and β are sufficiently small. Now for α sufficiently small (3.18) and (3.16) give the result.

Remark. Instead of this Lemma one could use Lemma 8 in Section 4 in a very special case ($|\mathcal{Q}| = |\mathcal{P}| = 1$) or the result of Section 5.3 in [15], which is for MLD and has a more elegant proof. This way we also would not be forced to worry about the right choices of ϵ and γ (γ small in Lemma 4, ϵ small in (3.6)), which makes a separate treatment of the cases $C = H(p^*)$ and $C < H(p^*)$ necessary. We decided to use Lemma 4 in order to show that for the DMC everything can be based on threshold decoding alone, which is not the case for the AVC.

§ 4. The double-exponential bound

Since there are only exponentially many code words it suffices to show that the probability to select a particular code word, which is decoded incorrectly with a probability $\geqslant \lambda$, is superexponentially small.

The idea which leads to this result is the following:

If $V_1 = v_1$, then v_1 has a "territory" $B(v_1)$ such that

$$w(B(v_1) \mid v_1) \geqslant 1 - \exp\,[-C_1(\gamma)n]$$

by Lemma 3 (a). The other code words cause losses in this territory, which together result in a sum of RV's $\sum_{j=2}^{M} T_j$, say. Those RV's have nice properties:

$E(T_i \mid T_{i-1} = t_{i-1}, \ldots, T_2 = t_2) \leqslant \exp\,[-(C - \gamma)n]$ by Lemma 3 (b) (1)

and $T_i \leqslant \exp\,[-C_2(\gamma)\epsilon n]$ by Lemma 4. (2)

Lemma 1 (b) gives the superexponential bound.

We now make this precise.

Case I. $C = H(p^*)$.

In this case for any γ, $0 < \gamma < C$, $B(x^n) \cap B(x'^n) = \emptyset$ for x^n, $x'^n \in \mathcal{X}^{*n}$, $x^n \neq x'^n$; $w(B(x^n) \mid x^n) = 1$ for $x^n \in \mathcal{X}^{*n}$; and therefore paragraph 2 contains the proof of the Theorem.

Case II. $C < H(p^*)$.

Now the γ in (3.6) can be omitted, we choose ϵ so small that

$$C + h(\epsilon) + \epsilon \log |\mathcal{X}| < H(p^*) - \epsilon \tag{3.6'}$$

and then we choose γ as small as required in Lemma 4.

We make the following *conventions*:

$$D_i^c = B(V_i)^c = \mathcal{Y}^n \text{ if } V_i = x_\infty,\ w(F \mid x_\infty) = 0 \text{ for every } F \subset \mathcal{Y}^n. \tag{3.25}$$

By symmetry it suffices to consider $\lambda_1(U_1, \ldots, U_M)$ which we upper bound as follows:

$$\lambda_1(U_1, \ldots, U_M) = w(D_1^c \mid V_1) \leqslant \sum_{j=2}^{M} w(B(V_1) \cap B(V_j) \mid V_1) + w(V_1)^c \mid V_1) \tag{3.26}$$

This and Lemma 3 (a) imply

$$\lambda_1(U_1, \ldots, U_M) \leqslant \sum_{j=2}^{M} w(B(V_j) \mid V_1) + \exp\,[-C_1(\gamma)n] \tag{3.27}$$

Define now RV's $\hat{V}_1, \ldots, \hat{V}_m$ by

$$\hat{V}_i = \begin{cases} U_i & \text{if } U_i \notin \bigcup_{j<i} S_\epsilon(U_j) \\ x_\infty & \text{otherwise} \end{cases} \tag{3.28}$$

and finally RV's $T_2, \ldots, T_M$ by

$$T_i = w(B(\hat{V}_i) \mid \hat{V}_1). \tag{3.29}$$

Notice that

$$\sum_{j=2}^{M} w(B(V_j) \mid V_1) \leqslant \sum_{j=2}^{M} T_j,\ \hat{V}_1 = U_1 \tag{3.30}$$

and that

$$E(T_j \mid T_{j-1} = t_{j-1}, \ldots, T_2 = t_2) \leqslant Ew(B(U_j) \mid U_1). \tag{3.31}$$

It therefore follows from Lemma 3 (b) that

$$E(T_j \mid T_{j-1} = t_{j-1}, \ldots, T_2 = t_2) \leqslant \exp\,[-(C - \gamma)n] \tag{3.32}$$

Since by definition (3.28) $d(\hat{V}_j, \hat{V}_1) \geqslant \epsilon n$, Lemma 4 and (3.29) yield

$$T_j \leqslant \exp\,[-C_2(\gamma)\epsilon n] \qquad \text{for } j = 2, \ldots, M. \tag{3.33}$$

It is clear from (3.27), (3.29), and (3.30) that the proof is complete once we have a superexponential bound on

$$\text{Prob}\left(\sum_{j=2}^{M} T_j \geqslant n\right), 0 < \eta < 1.$$

But this is now a straightforward consequence of (3.32), (3.33), and Lemma 1(b).

$$\text{Prob}\left(\sum_{j=2}^{M} T_j \geqslant \eta\right) \leqslant \exp\left[-\frac{\eta}{2}\exp\,(C_2(\gamma)\epsilon n)\right]$$

$$\times \prod_{j=2}^{M} \max_{t^{j-1}} E(1 + \exp\,[C_2(\gamma)\epsilon n]T_j \mid t^{j-1})$$

Using (3.32) we get

$$E(1 + \exp\,[C_2(\gamma)\epsilon n]T_j \mid t^{j-1}) \leqslant 1 + \exp\,[-(C - \gamma - C_2(\gamma)\epsilon)n]$$

and with the choices $\tau(\epsilon, \gamma) = \gamma + C_2(\gamma)\epsilon$, $M = \exp\,[(C - \tau(\epsilon, \gamma))]n$ therefore

$$\prod_{j=2}^{M} \max_{t^{j-1}} E(1 + \exp\,[C_2(\gamma)\epsilon n]T_j \mid t^{j-1}) \leqslant (1 + M^{-1})^M \leqslant e$$

and hence

$$\text{Prob}\left(\sum_{j=2}^{M} T_j \geqslant \eta\right) \leqslant \exp\left[-\frac{\eta}{2}\exp\,\{C_2(\gamma)n\} + 1\right]. \tag{3.34}$$

Q.E.D.

4. The Proof of Theorem 2

§ 1 Saddle points of $I(p, w)$ for classes of channels

We establish now two auxiliary results which go beyond Lemma 2.

Lemma 5. *Let $\bar{\mathcal{C}}$ be the closed convex hull of the set of matrices $\mathcal{C}$. Fix a PD P on $\mathcal{X}$ and let $\bar{w} \in \mathcal{C}$ be such that $\min_{w \in \bar{\mathcal{C}}} I(p, w) = I(p, \bar{w})$ Set. $\bar{q} = p \cdot \bar{w}$. Then*

$$\sum_{x, y} p(x)w(y \mid x) \log \frac{\bar{w}(y \mid x)}{\bar{q}(y)} \geqslant I(p, \bar{w}) \quad \textit{or all } w \in \bar{\mathcal{C}} f \tag{4.1}$$

Proof. For every $w \in \bar{C}$ and α, $0 \leqslant \alpha < 1$, $I(p, \alpha w + (1-\alpha)\overline{w}) \geqslant I(p, \overline{w})$ and therefore

$$\lim_{\alpha \to 0} \frac{\partial}{\partial \alpha} I(p, \alpha w + (1-\alpha)\overline{w}) \geqslant 0.$$

By an elementary calculation

$$\frac{\partial}{\partial \alpha} I(p, \alpha w + (1-\alpha)\overline{w})$$

$$= \sum_{x,y} p(x)(w(y \mid x) - \overline{w}(y \mid x)) \log \frac{\alpha w(y \mid x) + (1-\alpha)\overline{w}(y \mid x)}{\alpha q(y) + (1-\alpha)\overline{q}(y)}$$

and hence the result.

LEMMA 6. *Consider the row-convex hull $\bar{\bar{C}}$ of C. If p^*, $w^* \in \bar{\bar{C}}$ are such that $I(p^*, w^*) = \max_p \min_{w \in \bar{C}} I(p, w) = \bar{\bar{C}}$, then for $w \in \bar{\bar{C}}$:*

$$\sum_y w(y \mid x) \log \frac{w^*(y \mid x)}{q^*(y)} \geqslant \bar{\bar{C}} \quad \text{for } x \in \mathcal{X}^*. \tag{4.2}$$

Proof. It follows from Lemma 5 that

$$\sum_{x,y} p^*(x) w(y \mid x) \log \frac{w^*(y \mid x)}{q^*(y)} \geqslant I(p^*, w^*) = \bar{\bar{C}} \tag{4.3}$$

and from Lemma 2 that

$$\sum_y w^*(y \mid x) \log \frac{w^*(y \mid x)}{q^*(y)} = \bar{\bar{C}} \quad \text{for } x \in \mathcal{X}^*. \tag{4.4}$$

Suppose that (4.2) does not hold for some $x' \in \mathcal{X}^*$, then consider the matrix

$$\hat{w}(\cdot \mid x) = \begin{cases} w(\cdot \mid x') & \text{for } x = x' \\ w^*(\cdot \mid x) & \text{for } x \neq x' \end{cases},$$

which lies in $\bar{\bar{C}}$, because $\bar{\bar{C}}$ is row-convex closed. It follows from our supposition and from (4.4) that

$$\sum_{x,y} p^*(x) \hat{w}(y \mid x) \log \frac{w^*(y \mid x)}{q^*(y)} < \bar{\bar{C}},$$

which contradicts (4.3).

§ 2 TWO CODE WORD ERROR PROBABILITIES

We establish now two results for the AVC, which play the same role as Lemmas 3, 4 for the DMC. Notice however that Lemma 4 has no direct analogue, instead of TD we have to use MSD now.

LEMMA 7. *With p^*, $w^* \in \bar{\bar{C}}$, q^*, $\mathcal{X}^*$ and $\bar{\bar{C}}$ as previously defined, for*

$x^n \in \mathcal{X}^{*n}$ *and* $\gamma > 0$

$$\bar{\bar{B}}(x^n) = \left\{y^n : \frac{1}{n} \log \frac{(y^n \mid x^n)}{q^*(y^n)} \geqslant \bar{\bar{C}} - \gamma\right\}.$$

Let U be a RV with Prob $(U = x^n) = p^*(x^n)$. Then for $s^n \in \mathcal{S}^n$:

(a) $w(\bar{\bar{B}}(x^n)|x^n|s^n) \geqslant 1 - \exp[-\bar{\bar{C}}_1(\gamma)n]$, $x^n \in \mathcal{X}^{*n}$, *where*

$$\bar{\bar{C}}_1(\gamma) > 0 \text{ for } \gamma > 0.$$

(b) $Ew(\bar{\bar{B}}(U)|x^n|s^n) \leqslant \exp[-(\bar{\bar{C}} - \gamma)n]$.

Proof. (a) Using Lemma 6 instead of Lemma 2 the proof of Lemma 3(a) can be used here with small modifications. For $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^{*n}$, $s^n = (s_1, \ldots, s_n)$ consider the independent RV's $Y_1, \ldots, Y_n$, where Prob $(Y_t = y) = w(y|x_t|s_t)$. The RV's

$$Z_t = \begin{cases} \log \dfrac{w^*(Y_t \mid x_t)}{q^*(Y_t)} & \text{if } w^*(Y_t \mid x_t) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.5}$$

are again independent and uniformly bounded by a constant b^*, say, where

$$b^* = \max(-\log w^*_{\min}, -\log q^*_{\min}). \tag{4.6}$$

By Lemma 6

$$\text{Prob}\left\{\sum_{t=1}^n Z_t \leqslant (\bar{\bar{C}} - \gamma)n\right\} \leqslant \text{Prob}\left\{\sum_{t=1}^n (Z_t - EZ_t) \leqslant -\gamma n\right\},$$

and by Lemma 1 (d) this is smaller than $\exp\{(-\alpha\gamma + \alpha^2 b^{*2})n\}$ for $0 < \alpha < \min\left(1, \frac{b^2}{2} e^{-2b}, \gamma b^{-2}\right)$ and the exponent is negative.

(b) Just replace w in the proof of Lemma 3 (b) by w^* and make use of the Remark at the end of that proof.

LEMMA 8. *Let Q and $\mathcal{P}$ be two disjoint convex compact sets of PD's on a finite set $\mathcal{Z}$.* Let $\theta_0 = \frac{1}{2|\mathcal{Z}|} \min_{p \in \mathcal{P},\, q \in Q} \sum_{z \in \mathcal{Z}} |p(z) - q(z)|$ *and*

$$f(\theta) = \min_{0 \leqslant \beta \leqslant 1} \max_{\alpha > 0} [(\beta + \theta)\alpha - \log(\beta e^\alpha + 1 - \beta)], f(\theta) > 0 \quad \text{for } \theta > 0.$$

One can give a sequence $\left(S_n\right)_{n=1}^\infty$, $S_n \subset \mathcal{Z}^n$, *with* $p^n(S_n) \geqslant 1 - |\mathcal{Z}| \exp\{-f(\theta_0)n\}$, $q^n(S_n^c) \geqslant 1 - |\mathcal{Z}| \exp\{-f(\theta_0)n\}$ *for all* $p^n \in \{p_1 \times, \ldots, \times p_n : p_t \in \mathcal{P}, 1 \leqslant t \leqslant n\}$, $q^n \in \{q_1 \times \ldots \times q_n : q_t \in Q, 1 \leqslant t \leqslant n\}$.

Remark. In [9] a slightly weaker form was proved. There Q and $\mathcal{P}$ have only finitely many extremal points and the term $\exp\{-f(\theta_0)n\}$ is replaced by a constant λ, $0 < \lambda < 1$.

Proof. Let $n(z \mid z^n)$ count how often z occurs as component in $z^n = (z_1, \ldots, z_n)$.

For a PD p on $\mathcal{Z}$ define the set of sequences

$$\mathcal{Z}(p, \theta_0, n) = \{z^n \in \mathcal{Z}^n : |n(z \mid z^n) - p(z)n| < \theta_0 n, z \in \mathcal{Z}\}.$$

Recalling the definition of θ_0 we notice that

$$\mathcal{Z}(p, \theta_0, n) \cap \mathcal{Z}(q, \theta_0, n) = \emptyset \quad \text{for } p \in \mathcal{P}, q \in Q.$$

Therefore also

$$\Big(\bigcup_{p\in\mathcal{P}} \mathcal{Z}(p, \theta_0, n)\Big) \cap \Big(\bigcup_{q\in Q} \mathcal{Z}(q, \theta_0, n)\Big) = \emptyset.$$

Choose $S_n = \bigcup_{p\in\mathcal{P}} \mathcal{Z}(p, \theta_0, n)$.

By symmetry it suffices to show that for any $p^n = p_1 \times \ldots \times p_n$, $p_t \in \mathcal{P}$, and for $\bar{p} = \frac{1}{n}\sum_{t=1}^{n} p_t$

$$p^n(\mathcal{Z}(\bar{p}, \theta_0, n)) > 1 - |\mathcal{Z}| \exp\{-f(\theta_0)n\}.$$

Now observe that

$$\mathcal{Z}(\bar{p}, \theta_0, n) = \left\{z^n : \left|n(z \mid z^n) - \sum_{t=1}^{n} p_t(z)\right| < \theta_0 n, z \in \mathcal{Z}\right\} = B_n, \quad \text{say.}$$

Denote by $Z^n = Z_1, \ldots, Z_n$ a RV with PD p^n.

For $\alpha > 0$

$$p^n(B_n^c) \leqslant \sum_{z\in\mathcal{Z}} E \exp\left\{\alpha\left(n(z \mid Z^n) - \sum_{t=1}^{n} p_t(z) - \theta_0 n\right)\right\}$$

$$= \sum_{z\in\mathcal{Z}} \exp\{-n(\bar{p}(z) + \theta_0)\alpha\} \prod_{t=1}^{n} E \exp\{\alpha\, \delta(z, Z_t)\},$$

where $\delta(\cdot, \cdot)$ is Kronecker's Delta.

Now $E \exp\{\alpha\, \delta(z, Z_t)\} = p_t(z)e^{\alpha} + 1 - p_t(z)$ and since log is a convex function we conclude that

$$p^n(B_n^c) \leqslant \sum_{z\in\mathcal{Z}} \exp\{-n[(\bar{p}(z) + \theta_0)\alpha - \log[\bar{p}(z)e^{\alpha} + 1 - \bar{p}(z)]]\}$$

and by the definition of $f(\theta)$ that

$$p^n(B_n^c) \leqslant \sum_{z\in\mathcal{Z}} \exp\{-f(\theta_0)n\}.$$

That $f(\theta) > 0$ for $\theta > 0$ can be seen by looking at the slopes for $\alpha = 0$.

Q.E.D.

The pair $(S(x^n, x'^n), S(x'^n, x^n))$, $x'^n, x^n \in \mathcal{X}^n$, is called an MSD for C^n, if $S(x^n, x^n) = \emptyset$ for $x'^n = x^n$ and otherwise

(a) $S(x^n, x'^n) \cap S(x'^n, x^n) = \emptyset$, $S(x^n, x'^n) \cup S(x'^n, x^n) = \mathcal{Y}^n$

(b) $\min\left(\min_{s^n} w(A|x^n|s^n), \min_{s^n} w(A^c|x'^n|s^n)\right)$

is maximal for $(A, A^c) = (S(x^n, x'^n), S(x'^n, x^n))$.

Lemma 8 will be applied as follows. If $d(x^n, x'^n) \geq \epsilon n$ then there exists a pair (x, x'), $x \neq x'$, such that $|\{t : (x_t, x'_t) = (x, x'), \ 1 \leq t \leq n\}| \geq |\mathcal{X}|^{-2}\epsilon n$.

By condition (S)

$$\theta^* = \frac{1}{2|\mathcal{Y}|} \min_{x \neq x'} \left(\min_{w(\cdot|x) \in \mathcal{R}(x),\, w(\cdot|x') \in \mathcal{R}(x')} \sum_y |w(y \mid x) - w(y \mid x')| \right) > 0,$$

and hence by the Lemma for an MSD $(S(x^n, x'^n), S(x'^n, x^n))$ and all $s^n \in \mathcal{S}^n$

$$w(S(x'^n, x^n)|x^n|s^n) \leq |\mathcal{Y}| \exp\{-f(\theta^*)|\mathcal{X}|^{-2}\epsilon n\} \tag{4.7}$$

$$w(S(x^n, x'^n)|x'^n|s^n) \leq |\mathcal{Y}| \exp\{-f(\theta^*)|\mathcal{X}|^{-2}\epsilon n\}. \tag{4.8}$$

We use the abbreviation $C^* = \frac{1}{2} f(\theta^*)|\mathcal{X}|^{-2}$.

The next Lemma 9, based on MLD, establishes inequalities like (4.7), (4.8) under condition (L) instead of (S). Since we already know that Theorem 3 follows from Theorem 2, we state the result for comparison only and without proof.

LEMMA 9. *Let the AVC satisfy the condition*

$$\max_{w \in \bar{\bar{C}}} \sum_y w(y \mid x) \log \frac{w^*(y \mid x')}{w^*(y \mid x)} < 0 \text{ for all } x, x' \in \mathcal{X}^* \text{ with } x \neq x'. \tag{L}$$

For MLD-sets $L(x'^n, x^n) = \left\{y^n : \log \frac{w^*(y^n \mid x'^n)}{w^*(y^n \mid x^n)} \geq 0\right\}$ *here* $\log \frac{0}{0} = 0$ *we have for all* $s^n \in \mathcal{S}^n$ $w(L(x'^n, x^n)|x^n|s^n) \leq \exp\{-c\epsilon n\}$, *if* $x'^n, x^n \in \mathcal{X}^{*n}$ *and* $d(x'^n, x^n) \geq \epsilon n$.

§ 3. THE CODING SCHEME

Let $M = \exp\{(\bar{\bar{C}} - \gamma - C^*\epsilon)n\}$ and define independent RV's $U_1, \ldots, U_M$ with Prob $(U_i = x^n) = p^*(x^n)$ and RV's $V_1, \ldots, V_M$ exactly as in the coding scheme for the DMC. Therefore also the bounds on $|\mathcal{M}|$ derived in § 2 of Section 3 apply here. To every V_i, $i \in \mathcal{M}$, we assign now the set

$$\bar{\bar{B}}(V_i) = \left\{y^n : \frac{1}{n} \log \frac{w^*(y^n \mid V_i)}{q^*(y^n)} \geq \bar{\bar{C}} - \gamma\right\}$$

The decoding sets $\{D_i^* : i \in \mathcal{M}\}$ are now defined by

$$D_i^* = D_i^*(U_1, \ldots, U_M) = \bar{\bar{B}}(V_i) - \bigcup_{j \in \mathcal{M} - \{i\}} (S(V_j, V_i) \cap \bar{\bar{B}}(V_j)), \tag{4.9}$$

where $S(V_j, V_i)$ is as defined in Lemma 8.

Notice that in the decoding rule described in Section 2 we take $B(V_i) \cap B(V_j)$ away from both, $B(V_i)$ and $B(V_j)$. Here we have to be more cautious.

We define the error probabilities

$$\lambda_i(s^n) = \lambda_i(U_1, \ldots, U_M, s^n) = w(D_i^{*c}|V_i|s^n) \text{ for } i \in \mathcal{M} \text{ and } s^n \in \mathcal{S}^n \tag{4.10}$$

and

$$\lambda(U_1, \ldots, U_M) = \max_{i \in \mathcal{M}, s^n \in \mathcal{S}^n} \lambda_i(s^n). \tag{4.11}$$

§ 4. The proof of Theorem 2 in case $|\mathcal{C}| < \infty$

The converse, that is, the inequality $C \leqslant \bar{\bar{C}}$ follows (also in case $|\mathcal{C}| = \infty$) from the fact that an (n, N, λ)-code for $\mathcal{C}^n$ is also an (n, N, λ)-code for $\bar{\bar{\mathcal{C}}}^n$ (see [11]) and hence the capacities cannot exceed the capacity of the DMC w^*. The crux of the matter is the opposite inequality, which we now derive from our coding scheme. This proof now follows the lines of the proof of Theorem 1. One just has to replace Lemmas 3, 4 used there by Lemmas 7, 8 and has to use the fact that $|\mathcal{C}^n|$ only grows exponentially in n.

We make the following *conventions*:

$$D_i^{*c} = \bar{\bar{B}}(V_i)^c = \mathcal{Y}^n \text{ if } V_i = x_\infty,$$
$$w(F|x_\infty|s^n) = 0 \text{ for every } F \subset \mathcal{Y}^n.$$

For fixed s^n we upper bound now $\lambda_1(s^n)$.

$$\lambda_1(s^n) = w(D_1^{*c}|V_1|s^n)$$
$$\leqslant \sum_{j=2}^{M} w(\bar{\bar{B}}(V_1) \cap \bar{\bar{B}}(V_j) \cap S(V_j, V_1)|V_1|s^n) + \exp[-\bar{\bar{C}}_1(\gamma)n]. \tag{4.12}$$

Here we have used Lemma 7 (a):

$$w(\bar{\bar{B}}(V_1)^c|V_1|s^n) \leqslant \exp[-\bar{\bar{C}}_1(\gamma)n].$$

Define RV's $\hat{V}_1, \ldots, \hat{V}_M$ as in (3.28) and RV's $T_2, \ldots, T_M$ by

$$T_i = w(\bar{\bar{B}}(\hat{V}_i) \cap \mathcal{S}(\hat{V}_i, \hat{V}_1)|\hat{V}_1|s^n) \tag{4.13}$$

Now notice that

$$\lambda_1(s^n) \leqslant \sum_{j=2}^{M} T_j + \exp[-\bar{\bar{C}}_1(\gamma)n], \hat{V}_1 = U_1 \tag{4.14}$$

and that

$$E(T_j|T_{j-1} = t_{j-1}, \ldots, T_2 = t_2) \leqslant Ew(\bar{\bar{B}}(U_j) \cap \mathcal{S}(U_j, U_1)|U_1|s^n).$$

It follows therefore from Lemma 7(b) that

$$E(T_j \mid T_{j-1} = t_{j-1}, \ldots, T_2 = t_2) \leqslant \exp\,[-(\bar{\bar{C}} - \gamma)n]. \tag{4.15}$$

Since $d(\hat{V}_j, \hat{V}_1) \geqslant \epsilon n$, (4.7) implies

$$T_j \leqslant \exp\,[-C^*\epsilon n] \text{ for } j = 1, 2, \ldots, M \text{ and } n \geqslant n_0(\theta^*, \epsilon). \tag{4.16}$$

With (4.15) and (4.16) we have the same properties of the T_j's as in Section 3 and therefore also

$$\text{Prob}\left(\sum_{j=2}^{M} T_j \geqslant \eta\right) \leqslant \exp\left\{-\frac{\eta}{2}\exp\,[C^*\epsilon n] + 1\right\} \tag{4.17}$$

This means that for $0 < \eta < 1$

$$\text{Prob}\,(\lambda_i(U_1, \ldots, U_M, s^n) \geqslant \eta + \exp\,[-\bar{\bar{C}}_1(\gamma)n]$$

$$\leqslant \exp\left\{-\frac{\eta}{2}\exp\,[C^*\epsilon n] + 1\right\} \text{ for all } s^n \in \mathcal{S}^n \text{ and all } i \in \mathcal{M}, \tag{4.18}$$

and therefore

$$\text{Prob}\,(\lambda(U_1, \ldots, U_M) \geqslant \eta + \exp\,[-\bar{\bar{C}}_1(\gamma)n] \tag{4.19}$$

$$\leqslant |\mathcal{X}|^n|\mathcal{S}|^n \exp\left\{-\frac{\eta}{2}\exp\,[C^*\epsilon n] + 1\right\} = o(n). \qquad \text{Q. E. D.}$$

§ 5. The case $|\mathcal{C}| = \infty$

We make use of an Approximation Lemma [18] in the formulation of [5].

Lemma 10. *Let $A \geqslant ab^2$, where $a = |\mathcal{X}|$, $b = |\mathcal{Y}|$. There exists a set $\mathcal{C}_A \subset \mathcal{C}$, $|\mathcal{C}_A| < (A+1)^{ab}$, such that for every $w \in \mathcal{C}$ there exists a $\tilde{w} \in \mathcal{C}_A$ with*

$$|w(y \mid x) - \tilde{w}(y \mid x)| \leqslant bA^{-1},\ x \in \mathcal{X},\ y \in \mathcal{Y} \tag{4.20}$$

$$w(y \mid x) \leqslant \exp\,\{2b^2A^{-1}\}\tilde{w}(y \mid x),\ x \in \mathcal{X},\ y \in \mathcal{Y}. \tag{4.21}$$

Let $\mathcal{C}_A$ be the index set of $\mathcal{C}_A$ and denote the approximating matrix of $w(\cdot\,|\,\cdot\,|s)$ by $w(\cdot\,|\,\cdot\,|\tilde{s})$. By (4.21) we have for $s^n = (s_1, \ldots, s_n) \in \mathcal{S}^n$ and $\tilde{s}^n = (\tilde{s}_1, \ldots, \tilde{s}_n) \in \mathcal{S}_A^n$

$$w(y^n|x^n|s^n) \leqslant w(y^n|x^n|\tilde{s}^n)\exp\,\{2b^2nA^{-1}\} \text{ for all } x^n \in \mathcal{X}^n,\ y^n \in \mathcal{Y}^n \tag{4.22}$$

and with the choice $A(n) = n^2$ for $n \geqslant n_0(b)$

$$w(y^n|x^n|s^n) \leqslant w(y^n|x^n|\tilde{s}^n)2 \tag{4.23}$$

and

$$|\mathcal{S}_{n^2}| \leqslant (n^2+1)^{|\mathcal{X}|\,|\mathcal{Y}|} = \exp\{0(\log n)\}. \tag{4.24}$$

We now apply Lemma 10 to $\bar{\bar{C}}$ and call the approximating set of matrices $\bar{\bar{C}}_{n^2}$ and its index set $\bar{\bar{\mathcal{S}}}_{n^2}$.

We now replace in the arguments of the previous paragraph C by $\bar{\bar{C}}_{n^2}$. Thus we get instead of (4.19) for the approximating channel

$$\text{Prob}\,(\lambda_{n^2}(U_1, \ldots, U_M) \geqslant \eta + \exp(-\bar{\bar{C}}_1(\gamma)n)$$
$$\leqslant |\mathcal{X}|^n \exp\{0(n \log n)\} \exp\left\{-\frac{\eta}{2}\exp(C^*\epsilon n) + 1\right\} = o(n). \tag{4.25}$$

It follows from (4.23) that for any $s^n \in \bar{\bar{\mathcal{S}}}^n$, $w(D_i^{*C}|V_i|s^n) \leqslant 2 \cdot w(D_i^{*C}|x^n|\tilde{s}^n)$, $\tilde{s}^n \in \bar{\bar{\mathcal{S}}}^n_{n^2}$, and hence the result.

APPENDIX

5. An Alternative Model for our Coding Method

We now show how the coding method for the DMC can be described in terms of typical sequences and generated sequences (see [14]). The reasons for going through this are:

(a) The method becomes very intuitive for those familiar with [14].

(b) It is of interest to know that saddle point arguments can be avoided.

(c) This approach may be of use elsewhere.

§ 1 The alternative model

Let us consider a DMC with alphabets $\mathcal{X}$, $\mathcal{Y}$ and transmission matrix w. $X^n = X_1, \ldots, X_n$ shall be a sequence of i.i.d. RV's with values in $\mathcal{X}$ and $Y^n = Y_1, \ldots, Y_n$ denotes the corresponding channel output variables. $\mathcal{T}_\delta(X^n)$ is the set of typical sequences for the source X^n and $\mathcal{G}_\delta(Y^n|x^n)$, $x^n \in \mathcal{X}^n$, denotes the set of elements in $\mathcal{Y}^n$ which are generated by x^n. Definitions for those concepts are given in the next paragraph.

Let $U_1, \ldots, U_M$ be i.i.d. RV's with distribution

$$\text{Prob}\,(U_i = x^n) = |\mathcal{T}_\delta(X^n)|^{-1} \quad \text{for all } x^n \in \mathcal{T}_\delta(X^n). \tag{5.1}$$

Those RV's are now used to define a new sequence $V_1, \ldots, V_M$ of RV's as follows:

$$V_i = \begin{cases} U_i & \text{if } d(U_i, U_j) \geqslant \epsilon n \quad \text{for all } j \neq i \\ x_\infty & \text{otherwise.} \end{cases} \tag{5.2}$$

$\mathcal{M} = \{i : V_i \neq x_\infty,\ 1 \leqslant i \leqslant M\}$ is a random set of indices.

We choose $\{V_i : i \in \mathcal{M}\}$ as set of code words and define the decoding sets $\{D_i : i \in \mathcal{M}\}$ by the rule

$$D_i = D_i(U_1, \ldots, U_M) = \mathcal{G}_\delta(Y^n \mid U_i) - \bigcup_{j \in \mathcal{M} - \{i\}} \mathcal{G}_\delta(Y^n \mid V_j). \tag{5.3}$$

Finally, $\lambda_i = \lambda_i(U_1, \ldots, U_M)$ is the error probability for code word V_i.

THEOREM 1′. *For the coding method described in* (5.1)–(5.3) *with* M *specified by* $M = \exp\{(I(X \wedge Y) - \tau)n\}$, $\tau > 0$, *and* ϵ *sufficiently small, the following estimate holds*:

$$\text{Prob}\ (|\mathcal{M}| \geqslant \tfrac{1}{2}M,\ \max_{i \in \mathcal{M}} \lambda_i(U_1, \ldots, U_m) \leqslant \lambda) \tag{5.4}$$

$$\geqslant 1 - \exp\{-\exp[\rho(\tau, \lambda)n]\} \quad \textit{for } n \geqslant n_0(\lambda, \tau),$$

where $\rho(\tau, \lambda) > 0$ *for any* λ, $0 < \lambda < 1$.

§ 2 AUXILIARY RESULTS FOR THE PROOF OF THEOREM 1′

We give here the notions of typical sequences and generated sequences and those of their properties needed in the sequel. Lemmas which can be found in [14] or can be proved with easy modifications of proofs given there, will be stated without proof. Thus only Intersection Lemmas 1, 2 require a proof. If possible we have kept the terminology of [6].

Let $Z^n = Z_1, \ldots, Z_n$ be a sequence of i.i.d. RV's with values in $\mathcal{Z}$ and distribution p. For $z^n \in \mathcal{Z}^n$ and $z \in Z$ denote by $n(z \mid z^n)$ the number of components in which z^n has z.

z^n is (Z^n, δ)-typical (or (p, δ)-typical) if for $\delta \geqslant 0$

$$|np(z) - n(z \mid z^n)| \leqslant \delta[np(z)(1 - p(z))]^{1/2} \text{ for all } z \in \mathcal{Z}. \tag{5.5}$$

Denote the set of those sequences by $\mathcal{T}_\delta(Z^n)$.

All the $0(\sqrt{n})$ occurring below are bounded by $(C_1 + C_2\delta)\sqrt{n}$. For fixed alphabets C_1 and C_2 are universal constants (see chapter 2 of [14]).

LEMMA T. (a) *For every* n $\text{Prob}\ (Z^n \in \mathcal{T}_\delta(Z^n)) = 1 - 0\left(\frac{1}{\delta^2}\right)$, *that is, the probability goes to* 1 *uniformly in* n, *if* $\delta \to \infty$.

(b) $\text{Prob}\ (Z^n = z^n) = \exp\{-H(Z)n + 0(\sqrt{n})\}$ for $z^n \in \mathcal{T}_\delta(Z^n)$.

(c) $|\mathcal{T}_\delta(Z^n)| = \exp\{H(Z)n + 0(\sqrt{n})\}$.

Let $\left(Z_t, Y_t\right)_{t=1}^{\infty}$ be a discrete memoryless correlated source (DMCS), that is a sequence of i.i.d. pairs of RV's. Abbreviate $\text{Prob}\ (Y = y \mid Z = z)$ as $p(y \mid z)$. We also use notations such as $Y(z^n)$ for a RV with distribution $\text{Prob}\ (Y(z^n) = y^n) = \text{Prob}\ (Y^n = y^n \mid Z^n = z^n)$. For $\mathcal{Z} \in \mathcal{Z}^n$, $y^n \in \mathcal{Y}^n$, $z \in \mathcal{Z}$, $y \in \mathcal{Y}$ denote by $n(y, z \mid y^n, z^n)$ the number of positions in which

y^n has y and z^n has z.

A sequence $y^n \in \mathcal{Y}^n$ is $(Y^n \mid z^n, \delta)$-generated (or $(Y(z^n), \delta)$-typical), if

$$|n(y, z \mid y^n, z^n) - n(z \mid z^n)p(y \mid z)| \leqslant \delta[n(z \mid z^n)p(y \mid z)(1 - p(y \mid z))]^{1/2} \quad (5.6)$$

for all $y \in \mathcal{Y}$, $z \in \mathcal{Z}$.

Denote the set of those sequences by $\mathcal{G}_\delta(Y^n \mid z^n)$ or $\mathcal{T}_\delta(Y(z^n))$.

LEMMA G_1.

(a) *For every n*

$$\text{Prob } (Y^n \in \mathcal{G}_\delta(Y^n(z^n)) \mid Z^n = z^n) = \text{Prob } (Y(z^n) \in \mathcal{T}_\delta(Y(z^n)) = 1 - 0\left(\frac{1}{\delta^2}\right).$$

(b) $\text{Prob } (Y(z^n) = y^n) = \exp\{-H(Y(z^n)) + 0(\sqrt{n})\}$.

(c) $|\mathcal{G}_\delta(Y^n \mid z^n)| = \exp\{H(Y(z^n)) + 0(\sqrt{n})\}$.

(d) $H(Y(z^n)) = \sum_{t=1}^{n} H(Y \mid Z = z_t) = nH(Y \mid Z) + 0(\sqrt{n})$, if $z^n \in \mathcal{T}_\delta(Z^n)$.

LEMMA G_2. *Let* $\left(Z_t, Y_t\right)_{t=1}^{\infty}$ *be a* DMCS. *For every* $\delta \geqslant 0$ *there is a* $\delta_1 = \delta_1(|\mathcal{Y}|, |\mathcal{Z}|, \delta)$ *such that for all n*

(a) $\mathcal{G}_\delta(Y^n \mid z^n) \subset \mathcal{T}_{\delta_1}(Y^n)$ for $z^n \in \mathcal{T}_\delta(Z^n)$

(b) $\bigcup_{z^n \in \mathcal{T}_\delta(Z^n)} \mathcal{G}_{\delta_1}(Y^n \mid z^n) \supset \mathcal{T}_\delta(Y^n)$

(c) If $y^n \in \mathcal{T}_{\delta_1}(Y^n)$, then y^n is contained in

$\exp\{H(Z \mid Y)n + 0(\sqrt{n})\}$ sets $\mathcal{G}_\delta(Y^n \mid z^n)$ with $z^n \in \mathcal{T}_\delta(Z^n)$.

LEMMA I_1. *Let w be a* $|\mathcal{Z}| \times |\mathcal{Y}|$*-transmission matrix of a DMC with distinct rows. Let* $\left(Z_t, Y_t\right)_{t=1}^{\infty}$ *be a DMCS with* $\text{Prob } (Y_t = y \mid Z_t = z) = w(y \mid z)$, $y \in \mathcal{Y}$, $z \in \mathcal{Z}$.

For $z^n, z'^n \in \mathcal{T}_\delta(Z^n)$ *with* $d(z^n, z'^n) \geqslant \epsilon n$, $\epsilon > 0$,

$$|\mathcal{G}_\delta(Y^n \mid z^n) \cap \mathcal{G}_\delta(Y^n \mid z'^n)| \, |\mathcal{G}_\delta(Y^n \mid z^n)|^{-1} \leqslant \exp[-g(\epsilon)n], \quad (5.7)$$

for n large enough, where $g(\epsilon)$ *is positive and independent of n.*

Proof. It follows from Lemma 8 in Section 4 or already the special case thereof treated in chapter 5.3 of [15] that there exists an $A_n \subset \mathcal{Y}^n$:

$$\begin{aligned} &\text{Prob } (Y^n \in A_n \mid Z^n = z^n) \geqslant 1 - |\mathcal{Y}| \exp[-f(\theta_0)n\epsilon] \\ &\text{Prob } (Y^n \in A_n^c \mid Z^n = z'^n) \geqslant 1 - |\mathcal{Y}| \exp[-f(\theta_0)n\epsilon] \end{aligned} \quad (5.8)$$

where $\theta_0 = \dfrac{1}{2|\mathcal{Y}|} \min_{z \neq z'} \sum_y |w(y \mid z) - w(y \mid z')| > 0$.

By (b), (c) of Lemma G_1 we have for y^n in both generated sets

$$\begin{aligned} &\text{Prob } (Y^n = y^n \mid Z^n = z^n) \geqslant |\mathcal{G}_\delta(Y^n \mid z^n)|^{-1} \exp\{0(\sqrt{n})\} \\ &\text{Prob } (Y^n = y^n \mid Z^n = z'^n) \geqslant |\mathcal{G}_\delta(Y^n \mid z'^n)|^{-1} \exp\{0(\sqrt{n})\}. \end{aligned} \quad (5.9)$$

This, (5.8) and (c), (d) of Lemma G_1 imply

$$|\mathcal{G}_\delta(Y^n \mid z^n) \cap \mathcal{G}_\delta(Y^n \mid z'^n)|\,|\mathcal{G}_\delta(Y^n \mid z^n)|^{-1} \leqslant \exp\,(-f(\theta_0)n\epsilon + 0(\sqrt{n})).$$

The result follows by choosing $g(\epsilon) = \frac{1}{2}f(\theta_0)\epsilon$ for $n \geqslant n_0(\epsilon, \delta)$ suitable.

LEMMA I_2. *Let* $\left(Z_t, Y_t\right)_{t=1}^{\infty}$ *be defined as in the previous Lemma and let* U *be a RV which takes values in* $\mathcal{T}_\delta(Z^n)$ *with equal probabilities. Then for* $z^n \in \mathcal{T}_\delta(Z^n)$

$$E(|\mathcal{G}_\delta(Y^n \mid z^n) \cap \mathcal{G}_\delta(Y^n \mid U))|\,|\mathcal{G}_\delta(Y^n \mid z^n)|^{-1}) = \exp\{-I(Y \wedge Z)n + 0(\sqrt{n})\}.$$

Proof. It follows from Lemma G_2 that

$$\sum_{z'^n \in \mathcal{T}_\delta(Z^n)} |\mathcal{G}_\delta(Y^n \mid z^n) \cap \mathcal{G}_\delta(Y^n \mid z'^n)| = |\mathcal{G}_\delta(Y^n \mid z^n)| \exp\{H(Z \mid Y)n + 0(\sqrt{n})\}$$

and therefore

$$\sum_{z'^n \in \mathcal{T}_\delta(Z^n)} |\mathcal{G}_\delta(Y^n \mid z^n) \cap \mathcal{G}_\delta(Y^n \mid z'^n)|\,|\mathcal{G}_\delta(Y^n \mid z^n)|^{-1}|\mathcal{T}_\delta(Z^n)|^{-1} = \exp\{(H(Z \mid Y) - H(Z))n + 0(\sqrt{n})\}.$$

Those *two Intersection Lemmas* are basic for the analysis of our coding technique for the DMC.

§ 3. PROOF OF THEOREM 1′

The estimate on the rate Prob $(|\mathcal{M}| \leqslant \frac{1}{2}M) \leqslant 2 \cdot \exp\,(-\frac{1}{8}M)$ can be derived by exactly the same arguments as those used in Part I, if one replaces (3.6)–(3.8) there by

$$I(X \wedge Y) - \gamma + h(\epsilon) + \epsilon \log |\mathcal{X}| < H(X) - \epsilon \tag{5.10}$$

$$\left|\bigcup_{j=1}^{M} S_\epsilon(u_j)\right| \leqslant \exp\{(H(X) - \epsilon)n\} \tag{5.11}$$

$$\left|\bigcup_{j=1}^{M} S_\epsilon(u_j)\right|\,|\mathcal{T}_\delta(X^n)|^{-1} \leqslant \exp\{-g(\epsilon)n\} \tag{5.12}$$

We derive now the super exponential bound. With the *conventions*:

$$D_i^c = \mathcal{G}_\delta(X^n \mid V_i)^c = {}^{c}\mathcal{Y}^n \quad \text{if } V_i = x_\infty, \tag{5.13}$$
$$w(F \mid x_\infty) = 0 \quad \text{for every } F \subset {}^{c}\mathcal{Y}^n$$

we can write

$$\lambda_1(U_1, \ldots, U_M) = w(D_1^c \mid V_1) \leqslant \sum_{j=2}^{M} w(\mathcal{G}_\delta(X^n \mid V_1) \cap \mathcal{G}_\delta(X^n \mid V_j) \mid V_1) + w(\mathcal{G}_\delta(X^n \mid V_1)^c \mid V_1). \tag{5.14}$$

This and Lemma G_1 imply

$$\lambda_1(U_1, \ldots, U_M) \leqslant \sum_{j=2}^{M} w(\mathcal{G}_\delta(X^n \mid V_1) \cap \mathcal{G}_\delta(X^n \mid V_j) \mid V_1) + 0\left(\frac{1}{\delta^2}\right) \quad (5.15)$$

Define now RV's $\hat{V}_1, \ldots, \hat{V}_M$ by

$$\hat{V}_i = \begin{cases} U_i & \text{if } U_i \notin \bigcup_{j<i} S_\epsilon(U_j) \\ x_\infty & \text{otherwise} \end{cases} \quad (5.16)$$

and finally RV's $T_2, \ldots, T_M$ by

$$T_i = w(\mathcal{G}_\delta(X^n \mid \hat{V}_1) \cap \mathcal{G}_\delta(X^n \mid \hat{V}_j) \mid \hat{V}_1). \quad (5.17)$$

Notice that

$$\lambda_1 \leqslant \sum_{j=2}^{M} T_j + 0\left(\frac{1}{\delta^2}\right), \quad \hat{V}_1 = U_1 \quad (5.18)$$

$$\begin{aligned} &E(T_j \mid = T_{j-1}, \ldots, T_2 = t_2) \\ &\leqslant Ew(\mathcal{G}_\delta(X^n \mid U_1) \cap \mathcal{G}_\delta(X^n \mid U_j) \mid U_1) \\ &\leqslant \exp\{-I(X \wedge Y)n + 0(\sqrt{n})\}, \quad (5.19) \\ &\quad \text{by Lemma } I_2, \end{aligned}$$

and that

$$0 \leqslant T_j \leqslant \exp(-g(\epsilon)n), \quad \text{by Lemma } I_1. \quad (5.20)$$

The inequalities (5.19), (5.20) are the analogues to inequalities (3.32) and (3.33) and the proof can be completed as before.

References

[1] C. E. Shannon (1948), "The mathematical theory of communication", *Bell System Techn., J.*, **27**, 379–423, 623–656.

[2] ———— (1957), "Certain results in coding theory for noisy channels", *Information and Control*, **1**, 6–25.

[3] ———— (1956), "The zero-error capacity of a noisy channel", *IRE Trans. Inf. Th.*, **3**, 3–15.

[4] R. Ahlswede (1973), "Channels with arbitrarily varying channel probability functions in the presence of noiseless feedback", *Z. Wahrscheinlichkeitstheorie u. verw. Geb.*, **25** (3), 239–252.

[5] ———— (1978), "Elimination of correlation in random codes for arbitrarily varying channels", *Z. Wahrscheinlichkeitstheorie u. verw. Geb.*, **44**, 159–175.

[6] ————, "Coloring hypergraphs: a new approach to multi-user source coding", *J. of Combinatorics, Information Theory & System Sciences*, Part I, Vol. 4, No. 1, 76–115.

[7] ———— (1970), "A note on the existence of the weak capacity for channels with arbitrarily varying channel probability functions and its relation to Shannon's zero-error capacity", *Ann. Math. Stat.*, **41** (3), 1027–1033.

[8] D. Blackwell, L. Breiman and A. J. Thomasian (1960), "The capacities of certain channel classes under random coding", *Ann. Math. Stat.*, **31**, 558–561.

[9] J. Kiefer and J. Wolfowitz (1962), "Channels with arbitrarily varying channel probability functions", *Information and Control*, **5**, 44–54.

[10] R. L. Dobrushin and S. Z. Stambler (1975), "Coding theorems for classes of arbitrarily varying discrete memoryless channels", *Problemy Peredachi Informatsii*, **11** (2), 3–22.

[11] R. Ahlswede and J. Wolfowitz (1970), "The capacity of a channel with arbitrarily varying channel probability functions and binary output alphabet, *Z. Wahrscheinlichkeitstheorie u. verw. Geb.*, **15** (3), 186–194.

[12] L. Lovász, "On the Shannon capacity of a graph". (to appear).

[13] I. Csiszar, and J. Körner, *Information Theory* (forthcoming book).

[14] J. Wolfowitz (1978), *Coding Theorems of Information Theory*, third edition, Springer Verlag, Berlin–Heidelberg-New York.

[15] R. G. Gallager (1968), *Information Theory and Reliable Communication*, John Wiley and Sons, New York.

[16] S. Z. Stambler (1975), "Shannon's theorems for a complete class of discrete channels whose state is known at the output", *Probl. Peredachi Inform*, **11** (2), 3–22.

[17] R. Ahlswede (1973), "Channel capacities for list codes", *J. Appl. Probability*, **10**, 824–836.

[18] D. Blackwell, L. Breiman and A. J. Thomasian (1959), "The capacity of a class of channels", *Ann. Math. Stat.*, **30**, 1229–1241.

[19] V. D. Goppa (1975), "Universal decoding for symmetric channels", *Probl. Peredachi Informatsii*, **11** (1), 15–22.

[Received : Jan. 1979]