INFORMATION THEORY

CAPACITY OF THE ARBITRARILY VARYING CHANNEL UNDER LIST DECODING

V. M. Blinovsky[1], **P. Narayan, and M. S. Pinsker**[1]

UDC 621.391.15

We find necessary and sufficient conditions under which the capacity C_L of an arbitrarily varying channel (AVC), for deterministic codes with decoding into a list of size L and for the average error probability criterion, equals the capacity C_r of the AVC for random codes. For binary AVCs, we prove the existence of a finite $L^* < \infty$ such that $C_L = C_r$ for all $L > L^*$.

1. Introduction

An interesting property of an arbitrarily varying channel (AVC) is the following. The deterministic code capacity for the average probability of error criterion can equal zero while the random code capacity can be positive (see [1]). This fact suggests that one investigate necessary and sufficient conditions for the equality of these capacities under various restrictions on input signals. It also suggests that one investigate whether or not list decoding can change this situation.

List decoding for ordinary channels was first considered in the papers of Elias [2] and Wozencraft [3]. List decoding for such channels does not change their capacities although it can change other properties of the transmission.

In this paper, we study problems of AVC capacity under fixed size-L list decoding. In doing so, we restrict ourselves to the study of a discrete memoryless AVC with finite input and output alphabets and finite state space for the channel under the average probability of error criterion.

2. Basic notation and formulation of results

First we recall the definition of the discrete memoryless AVC and introduce some quantities that reflect its specific behavior.

Let $\mathcal{X}, \mathcal{Y}$, and $\mathcal{S}$ be finite sets representing the input, output, and state alphabet, respectively. The AVC is determined by a family of conditional distributions $w(y \,|\, x, s)$ on $\mathcal{Y} (y \in \mathcal{Y})$, defined by an input signal $x \in \mathcal{X}$ and a state $s \in \mathcal{S}$. The absence of memory means that the transition probability function $w^n(\mathbf{y} \,|\, \mathbf{x}, \mathbf{s})$, $\mathbf{x} = (x(1), \ldots, x(n)) \in \mathcal{X}^n$, $\mathbf{y} = (y(1), \ldots, y(n)) \in \mathcal{Y}^n$, $\mathbf{s} = (s(1), \ldots, s(n)) \in \mathcal{S}^n$, satisfies

$$w^n(\mathbf{y} \,|\, \mathbf{x}, \mathbf{s}) = \prod_{j=1}^{n} w(y(j) \,|\, x(j), s(j)). \tag{1}$$

We denote such a channel as $G = (w, \mathcal{X}, \mathcal{Y}, \mathcal{S})$. A deterministic code $\mathcal{K}^n$ of length n and cardinality M is the set $\mathcal{K}^n \triangleq \{(\mathbf{x}_i, A_i),\ i = 1, \ldots, M\}$, where $\mathbf{x}_i \in \mathcal{X}^n$, and $\{A_i,\ i = 1, \ldots, M\}$ is a partition of the space

[1] Partly supported by the Russian Fundamental Research Foundation (project No. 93-012-458.)

Translated from *Problemy Peredachi Informatsii*, Vol. 31, No. 2, pp. 3–19, April–June, 1995. Original article submitted May 24, 1993; revision submitted August 8, 1994.

$\mathcal{Y}^n$. A message m_i, $i \in \{1, \dots, M\}$, is encoded into the codeword $\mathbf{x}_i$ which is transmitted over the AVC. At the output of the channel in state $\mathbf{s}$, we observe the sequence $\mathbf{y} \in \mathcal{Y}^n$ with probability $w^n(\mathbf{y} \mid \mathbf{x}_i, \mathbf{s})$. The received sequence $\mathbf{y}$ is decoded into a message m_j, $j \in \{1, \dots, M\}$. The decoding rule is given by the function Φ: $\Phi(\mathbf{y}) = j$ if $\mathbf{y} \in A_j$. The average error probability $\overline{p}(\mathbf{s})$ of the code $\mathcal{K}^n$ when the AVC is in the state $\mathbf{s} \in \mathcal{S}^n$ equals

$$\overline{p}(\mathbf{s}) \triangleq M^{-1} \sum_{i=1}^{M} \sum_{\mathbf{y} \in \mathcal{Y}^n : \Phi(\mathbf{y}) \neq i} w^n(\mathbf{y} \mid \mathbf{x}_i, \mathbf{s}). \tag{2}$$

We are interested in the quantity

$$\overline{e}(R) = \limsup_{n \to \infty} \min_{\mathcal{K}^n : \log M \geq Rn} \max_{\mathbf{s} \in \mathcal{S}^n} \overline{p}(\mathbf{s}) \tag{3}$$

for $R > 0$. The capacity of the AVC with deterministic codes under the average error probability criterion is defined as

$$C \triangleq \sup_{\overline{e}(R)=0} R. \tag{4}$$

Ahlswede [1] has shown that the capacity C equals either 0 or the random code capacity C_r, where C_r is equal to

$$C_r = \max_{p(\cdot)} \min_{q(\cdot)} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} w_q(y \mid x) p(x) \log \frac{w_q(y \mid x)}{\sum\limits_{x' \in \mathcal{X}} w_q(y \mid x') p(x')}, \tag{5}$$

where $p(\cdot), q(\cdot)$ are probability distributions on $\mathcal{X}, \mathcal{S}$ respectively, and $w_q(y \mid x) = \sum\limits_{s \in \mathcal{S}} w(y \mid x, s)\, q(s)$.

In [4], a sufficient condition on the distributions $w(y \mid x, s)$ for which $C > 0$ was given. However, later it became clear that this condition is not necessary, and, in particular, in [2], an example was given where $C = C_r > 0$ but the condition in [4] is not satisfied. In [5], Ericson formulated a weaker condition that implies $C = 0$. Later Csiszár and Narayan [6] proved that the condition described in [5] is also necessary for $C = 0$.

In the present paper, we consider list decoding of fixed size L; the results of [1–7] then correspond to the special case $L = 1$. Suppose that the collection $\{A_i,\ i = \overline{1, M}\}$ satisfies $\bigcap\limits_{i \in J} A_i = \varnothing$ for all $J \subset \{1, \dots, M\}$, $|J| \geq L + 1$, $\bigcup\limits_{i=1}^{M} A_i = \mathcal{Y}^n$. The set of pairs $\{(\mathbf{x}_i, A_i),\ i = \overline{1, M}\} \triangleq \mathcal{K}^n$ is called the deterministic code decoded into a list of size L. The set $\mathcal{K}^n_{\mathbf{X}} = \{\mathbf{x}_i,\ i = \overline{1, M}\}$ is also called a code. A received sequence $\mathbf{y}$ is decoded into a list of $L' \leq L$ messages $m_{i_1}, \dots, m_{i_{L'}}$, $\{i_1, \dots, i_{L'}\} \subset \{1, \dots, M\}$. The decoding rule $\mathcal{Y}^n \to \{1, \dots, M\}^L$ is

$$\Phi_L(\mathbf{y}) = \{i : A_i \ni \mathbf{y},\ i = \overline{1, M}\}.$$

The average error probability of decoding into a list of size L for transmission over the AVC in the state $\mathbf{s} \in \mathcal{S}^n$ is defined as

$$\overline{p}_L(\mathbf{s}) = \overline{p}_L(\mathbf{s}, \mathcal{K}^n) = M^{-1} \sum_{i=1}^{M} \sum_{\mathbf{y} \in \mathcal{Y}^n : \Phi_L(\mathbf{y}) \not\ni i} w^n(\mathbf{y} \mid \mathbf{x}_i, \mathbf{s}). \tag{6}$$

Define the list-of-L size capacity C_L of the AVC for deterministic codes under the average error probability criterion as

$$C_L \triangleq \sup_{\overline{e}_L(R)=0} R, \tag{7}$$

where

$$\overline{e}_L(R) = \limsup_{n \to \infty} \min_{\mathcal{K}^n : \log M \geq Rn} \max_{\mathbf{s} \in \mathcal{S}^n} \overline{p}_L(\mathbf{s}).$$

In [8], transmission with list decoding was considered. In that paper, it was proved that for any $R < C_r$, $\varepsilon > 0$, there exists an encoding and decoding scheme with list size $L(\varepsilon, |\mathcal{X}|, |\mathcal{Y}|)$ such that for all sufficiently large n, $\overline{p}_L(\mathrm{s}) < \varepsilon$, $\mathrm{s} \in \mathcal{S}^n$. It was conjectured that $L(\varepsilon, |\mathcal{X}|, |\mathcal{Y}|)$ depends only on $|\mathcal{X}|, |\mathcal{Y}|$ (but not on ε).

In the present paper, we find a necessary and sufficient condition for $C_L = C_r$ to be valid. In the case $|\mathcal{X}| = |\mathcal{Y}| = |\mathcal{S}| = 2$ (the binary AVC), we prove the existence of a number $L(G) < \infty$ such that $C_L = C_r$ for $L > L(G)$. We also show that for any $N > 0$, there exists a binary AVC with $L(G) > N$.

Now, we give an exact formulation of the main results to be proved below. We commence with a theorem which extends the statement of Ahlswede [1] from $L = 1$ to the case $L \geq 1$.

Theorem 1. *C_L equals either C_r or 0.*

THE PROOF of Theorem 1 follows the deviation in [1]. When $C_r > 0$, we use a prefix code with list-of-L decoding at a nonzero rate.

A sufficient condition for $C_L = 0$ is formulated in the next lemma as an extension of the "symmetrizability" condition in [6].

Lemma 1. *Suppose that there exist conditional probability distributions*

$$q(s \mid x_2, \ldots, x_{L+1}), \quad s \in \mathcal{S};\ x_2, \ldots, x_{L+1} \in \mathcal{X}$$

such that for any set $x_1, \ldots, x_{L+1} \in \mathcal{X}$, any permutation $\pi = (\pi_1, \ldots, \pi_{L+1})$ of the sequence $(1, \ldots, L+1)$, and any $y \in \mathcal{Y}$ the following equalities hold:

$$\sum_{s \in \mathcal{S}} w(y \mid x_1, s)\, q(s \mid x_2, \ldots, x_{L+1}) = \sum_{s \in \mathcal{S}} w(y \mid x_{\pi_1}, s)\, q(s \mid x_{\pi_2}, \ldots, x_{\pi_{L+1}}). \tag{8}$$

Then $C_L = 0$.

Remark. Clearly, condition (8) is equivalent to the equality

$$\min_{q(\cdot \mid \cdot)} \max_{\substack{y \in \mathcal{Y},\ \pi, \\ x_1, \ldots, x_{L+1} \in \mathcal{X}}} \sum_{s \in \mathcal{S}} [w(y \mid x_1, s)\, q(s \mid x_2, \ldots, x_{L+1}) - w(y \mid x_{\pi_1}, s)\, q(s \mid x_{\pi_2}, \ldots, x_{\pi_{L+1}})] = 0. \tag{9}$$

We say that the distribution $w(y \mid x, s)$ (or the AVC) is L-symmetrizable if relation (8) holds and is L-nonsymmetrizable otherwise. Obviously, if the AVC is L-symmetrizable, then it is L'-symmetrizable, where $L' \leq L$; in this case, it is sufficient to assume $q(s \mid x_2, \ldots, x_{L'+1}) = q(s \mid x_2, \ldots, x_{L'+1}, \ldots, x_{L+1})$, where $x_{L'+2}, \ldots, x_{L+1}$ is fixed.

A distribution $w(y \mid x, s)$ for which there exists $q(s)$ such that

$$\sum_{s \in \mathcal{S}} w(y \mid x, s)\, q(s) = w_q(y \mid x) = V(y)$$

does not depend on x is symmetrizable for any $L \geq 1$. To verify this, we can put $q(s \mid x_2, \ldots, x_{L+1}) = q(s)$, $s \in \mathcal{S}$. It follows from (5) that for such channels $C_r = 0$. The largest $L \triangleq L(G)$ for which the AVC is L-symmetrizable will be called the order of symmetrizability.

A necessary condition for $C_L = 0$ is given in the following lemma.

Lemma 2. *If $C_L = 0$, then (8) holds for a certain $q(s \mid x_2, \ldots, x_{L+1})$.*

As a corollary of Lemmas 1 and 2 we have

Theorem 2. *$C_L = 0$ iff the distribution $w(y \mid x, s)$ is L-symmetrizable.*

We get from Theorems 1 and 2

Corollary 1. *If the distribution $w(y \mid x, s)$ is L-symmetrizable then $C_L = 0$.*

Corollary 2. *If the distribution $w(y \mid x, s)$ is L-nonsymmetrizable then $C_L = C_r > 0$.*

Theorem 3. *For the binary AVC, ($|\mathcal{X}| = |\mathcal{Y}| = |\mathcal{S}| = 2$) with $C_r > 0$, we have $L(G) < \infty$. Moreover, for any $N > 0$, there exists a binary AVC with $L(G) > N$.*

3. Proof of Theorem 2

PROOF OF LEMMA 1. Let $\mathcal{K}_{\mathbf{X}}^n = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$ be a set of codewords. We show that if (8) holds, then there exists a vector $\mathbf{s} \in \mathcal{S}^n$, for which

$$\overline{p}_L(\mathbf{s}) \geq (L+1)^{-1}(1 + o(1)), \tag{10}$$

where $o(1) \to 0$ when $M \to \infty$. Lemma 1 follows from this inequality.

Let $\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_{L+1}} \in \mathcal{K}_{\mathbf{X}}^n$, $i_1, \ldots, i_{L+1} \in \{1, \ldots, M\}$. Consider an n-dimensional random variable

$$\mathbf{S}_{i_2 \ldots i_{L+1}} = (S_{i_2 \ldots i_{L+1}}(1), \ldots, S_{i_2 \ldots i_{L+1}}(n))$$

with independent components $S_{i_2 \ldots i_{L+1}}(j)$, $j = 1, \ldots, n$, and distributions

$$\Pr\left(S_{i_2 \ldots i_{L+1}}(j) = s\right) = q(s \,|\, x_{i_2}(j), \ldots, x_{i_{L+1}}(j))$$

that satisfy (8).

According to (1), we have

$$\begin{aligned} \mathbf{E}w^n(\mathbf{y} \,|\, \mathbf{x}_{i_1}, \mathbf{S}_{i_2 \ldots i_{L+1}}) &= \prod_{j=1}^{n} \mathbf{E}w(y(j) \,|\, x_{i_1}(j), S_{i_2 \ldots i_{L+1}}(j)) \\ &= \prod_{j=1}^{n} \sum_{s \in \mathcal{S}} w(y(j) \,|\, x_{i_1}(j), s) q(s \,|\, x_{i_2}(j), \ldots, x_{i_{L+1}}(j)). \end{aligned}$$

Let $\pi = (\pi_1, \ldots, \pi_{L+1})$ be an arbitrary permutation of the elements of the sequence $(i_1, \ldots, i_{L+1})$. It follows from (8) that

$$\mathbf{E}w^n(\mathbf{y} \,|\, \mathbf{x}_{\pi_1}, \mathbf{S}_{\pi_2 \ldots \pi_{L+1}}) = \mathbf{E}w^n(\mathbf{y} \,|\, \mathbf{x}_{i_1}, \mathbf{S}_{i_2 \ldots i_{L+1}}). \tag{11}$$

Let

$$\gamma_{i_1 \ldots i_{L+1}} \triangleq \sum_{\pi} \sum_{\mathbf{y}:\, \pi_1 \notin \Phi(\mathbf{y})} \mathbf{E}w^n(\mathbf{y} \,|\, \mathbf{x}_{\pi_1}, \mathbf{S}_{\pi_2 \ldots \pi_{L+1}}). \tag{12}$$

Since we are considering list-of-L decoding, it holds that

$$\sum_{\mathbf{y}} \mathbf{E}w^n(\mathbf{y} \,|\, \mathbf{x}_{\pi_1}, \mathbf{S}_{\pi_2 \ldots \pi_{L+1}}) = 1$$

implies for $\{i_1, \ldots, i_{L+1}\} \subset \{1, \ldots, M\}$ that

$$\sum_{j=1}^{L+1} \sum_{\pi} \sum_{\mathbf{y}:\, \pi_j \notin \Phi(\mathbf{y})} \mathbf{E}w^n(\mathbf{y} \,|\, \mathbf{x}_{\pi_1}, \mathbf{S}_{\pi_2 \ldots \pi_{L+1}}) \geq \sum_{\pi} \sum_{\mathbf{y}} \mathbf{E}w^n(\mathbf{y} \,|\, \mathbf{x}_{\pi_1}, \mathbf{S}_{\pi_2 \ldots \pi_{L+1}}) = (L+1)!. \tag{13}$$

Combining (11) with (13) yields the estimate

$$\gamma_{i_1 \ldots i_{L+1}} \geq L!, \quad i_j \neq i_k,\ j \neq k. \tag{14}$$

It follows from (11)–(13) that

$$\begin{aligned} &\sum_{\substack{i_1, \ldots, i_j, \ldots, i_k, \ldots, i_{L+1} = 1, \\ i_j \neq i_k,\ j \neq k}}^{M} \sum_{\mathbf{y}:\, i_1 \notin \Phi(\mathbf{y})} \mathbf{E}w^n(\mathbf{y} \,|\, \mathbf{x}_{i_1}, \mathbf{S}_{i_2 \ldots i_j \ldots i_k \ldots i_{L+1}}) \\ &\geq \left[(L+1)!\right]^{-1} \sum_{\substack{i_1, \ldots, i_j, \ldots, i_k, \ldots, i_{L+1} = 1, \\ i_j \neq i_k,\ j \neq k}}^{M} \gamma_{i_1 \ldots i_j \ldots i_k \ldots i_{L+1}}. \end{aligned} \tag{15}$$

Next, observe that

$$\begin{aligned} M^{-1} & \sum_{i_1=1}^{M} \sum_{\mathbf{y}:\, i_1 \notin \Phi(\mathbf{y})} \mathbf{E} w^n(\mathbf{y} \,|\, \mathbf{x}_{i_1}, \mathbf{S}_{i_2 \ldots i_{L+1}}) \\ &= M^{-1} \sum_{i_1=1}^{M} \sum_{\mathbf{y}:\, i_1 \notin \Phi(\mathbf{y})} \sum_{\mathbf{s} \in \mathcal{S}^n} w^n(\mathbf{y} \,|\, \mathbf{x}_{i_1}, \mathbf{s})\, q^n(\mathbf{s} \,|\, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_{L+1}}) \qquad (16) \\ &= \sum_{\mathbf{s} \in \mathcal{S}^n} \overline{p}_L(\mathbf{s})\, q^n(\mathbf{s} \,|\, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_{L+1}}). \end{aligned}$$

The last expression is the error probability $p_L(\mathbf{s})$ averaged over

$$q^n(\mathbf{s} \,|\, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_{L+1}}) = \prod_{k=1}^{n} q(s(k) \,|\, x_{i_2}(k), \ldots, x_{i_{L+1}}(k)).$$

In accordance with (14)–(16) we have

$$\begin{aligned} M^{-L} & \sum_{\substack{i_2, \ldots, i_j, \ldots, i_k, \ldots, i_{L+1} = 1, \\ i_j \neq i_k,\ j \neq k}}^{M} \sum_{\mathbf{s} \in \mathcal{S}^n} \overline{p}_L(\mathbf{s})\, q^n(\mathbf{s} \,|\, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_j}, \ldots, \mathbf{x}_{i_k}, \ldots, \mathbf{x}_{i_{L+1}}) \\ \geq\ & M^{-L-1}[(L+1)!]^{-1} \sum_{\substack{i_1, \ldots, i_j, \ldots, i_k, \ldots, i_{L+1} = 1, \\ i_j \neq i_k,\ j \neq k}}^{M} \gamma_{i_1 \ldots i_j \ldots i_k \ldots i_{L+1}} \\ \geq\ & M^{-L}(M-1)(M-2)\ldots(M-L)L![(L+1)!]^{-1} \geq (L+1)^{-1}(1+o(1)), \quad M \to \infty. \end{aligned}$$

From the last expression we deduce the existence of a collection $\{i_2, \ldots, i_{L+1}\}$ such that

$$\sum_{\mathbf{s} \in \mathcal{S}^n} \overline{p}_L(\mathbf{s})\, q^n(\mathbf{s} \,|\, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_{L+1}}) \geq (L+1)^{-1}(1+o(1)).$$

Hence, there exists a vector $\mathbf{s} \in \mathcal{S}^n$ for which (10) holds, thereby completing the proof of Lemma 1.

Before the proof of the necessity of condition (8) we introduce some standard definitions.

Let $\mathcal{X}_1, \ldots, \mathcal{X}_m$ be finite sets and let $X_1, \ldots X_m$ be random variables with probability distribution

$$p(x_1, \ldots, x_m) = p_{X_1 \ldots X_m}(x_1, \ldots, x_m), \quad x_i \in \mathcal{X}_i,\ i = \overline{1,m}.$$

Suppose that

$$H(X_1, \ldots, X_m) = -\sum_{\substack{x_i \in \mathcal{X}_i, \\ i=\overline{1,m}}} p(x_1, \ldots, x_m) \log p(x_1, \ldots, x_m)$$

is the entropy of the distribution $p(x_1, \ldots, x_m)$ of the random variable $(X_1, \ldots, X_m)$,

$$I(X_1; \ldots; X_m) = \sum_{\substack{x_i \in \mathcal{X}_i, \\ i=\overline{1,m}}} p(x_1, \ldots, x_m) \log \frac{p(x_1, \ldots, x_m)}{p(x_1) \ldots p(x_m)},$$

and

$$I(X^k; X_{k+1}^m), \quad 1 \leq k < m,$$

is the amount of information between the random variables $X^k = (X_1, \ldots, X_k)$ and $X_{k+1}^m = (X_{k+1}, \ldots, X_m)$. Obviously, we have

$$\begin{aligned} I(X_1; \ldots; X_m) &= \sum_{i=1}^{m} H(X_i) - H(X_1, \ldots, X_m), \\ I(X^k; X_{k+1}^m) &= H(X^k) + H(X_{k+1}^m) - H(X^m). \end{aligned}$$

3. Proof of Theorem 2

Proof of Lemma 1. Let $\mathcal{K}_{\mathbf{X}}^n = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$ be a set of codewords. We show that if (8) holds, then there exists a vector $\mathbf{s} \in \mathcal{S}^n$, for which

$$\overline{p}_L(\mathbf{s}) \geq (L+1)^{-1}(1 + o(1)), \tag{10}$$

where $o(1) \to 0$ when $M \to \infty$. Lemma 1 follows from this inequality.

Let $\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_{L+1}} \in \mathcal{K}_{\mathbf{X}}^n$, $i_1, \ldots, i_{L+1} \in \{1, \ldots, M\}$. Consider an n-dimensional random variable

$$\mathbf{S}_{i_2 \ldots i_{L+1}} = (S_{i_2 \ldots i_{L+1}}(1), \ldots, S_{i_2 \ldots i_{L+1}}(n))$$

with independent components $S_{i_2 \ldots i_{L+1}}(j)$, $j = 1, \ldots, n$, and distributions

$$\Pr\left(S_{i_2 \ldots i_{L+1}}(j) = s\right) = q(s \mid x_{i_2}(j), \ldots, x_{i_{L+1}}(j))$$

that satisfy (8).

According to (1), we have

$$\begin{aligned} \mathbf{E}w^n(\mathbf{y} \mid \mathbf{x}_{i_1}, \mathbf{S}_{i_2 \ldots i_{L+1}}) &= \prod_{j=1}^{n} \mathbf{E}w(y(j) \mid x_{i_1}(j), S_{i_2 \ldots i_{L+1}}(j)) \\ &= \prod_{j=1}^{n} \sum_{s \in \mathcal{S}} w(y(j) \mid x_{i_1}(j), s) q(s \mid x_{i_2}(j), \ldots, x_{i_{L+1}}(j)). \end{aligned}$$

Let $\pi = (\pi_1, \ldots, \pi_{L+1})$ be an arbitrary permutation of the elements of the sequence $(i_1, \ldots, i_{L+1})$. It follows from (8) that

$$\mathbf{E}w^n(\mathbf{y} \mid \mathbf{x}_{\pi_1}, \mathbf{S}_{\pi_2 \ldots \pi_{L+1}}) = \mathbf{E}w^n(\mathbf{y} \mid \mathbf{x}_{i_1}, \mathbf{S}_{i_2 \ldots i_{L+1}}). \tag{11}$$

Let

$$\gamma_{i_1 \ldots i_{L+1}} \triangleq \sum_{\pi} \sum_{\mathbf{y}:\, \pi_1 \notin \Phi(\mathbf{y})} \mathbf{E}w^n(\mathbf{y} \mid \mathbf{x}_{\pi_1}, \mathbf{S}_{\pi_2 \ldots \pi_{L+1}}). \tag{12}$$

Since we are considering list-of-L decoding, it holds that

$$\sum_{\mathbf{y}} \mathbf{E}w^n(\mathbf{y} \mid \mathbf{x}_{\pi_1}, \mathbf{S}_{\pi_2 \ldots \pi_{L+1}}) = 1$$

implies for $\{i_1, \ldots, i_{L+1}\} \subset \{1, \ldots, M\}$ that

$$\sum_{j=1}^{L+1} \sum_{\pi} \sum_{\mathbf{y}:\, \pi_j \notin \Phi(\mathbf{y})} \mathbf{E}w^n(\mathbf{y} \mid \mathbf{x}_{\pi_1}, \mathbf{S}_{\pi_2 \ldots \pi_{L+1}}) \geq \sum_{\pi} \sum_{\mathbf{y}} \mathbf{E}w^n(\mathbf{y} \mid \mathbf{x}_{\pi_1}, \mathbf{S}_{\pi_2 \ldots \pi_{L+1}}) = (L+1)!. \tag{13}$$

Combining (11) with (13) yields the estimate

$$\gamma_{i_1 \ldots i_{L+1}} \geq L!, \quad i_j \neq i_k,\ j \neq k. \tag{14}$$

It follows from (11)–(13) that

$$\begin{aligned} &\sum_{\substack{i_1, \ldots, i_j, \ldots, i_k, \ldots, i_{L+1} = 1, \\ i_j \neq i_k,\ j \neq k}}^{M} \sum_{\mathbf{y}:\, i_1 \notin \Phi(\mathbf{y})} \mathbf{E}w^n(\mathbf{y} \mid \mathbf{x}_{i_1}, \mathbf{S}_{i_2 \ldots i_j \ldots i_k \ldots i_{L+1}}) \\ &\geq \left[(L+1)!\right]^{-1} \sum_{\substack{i_1, \ldots, i_j, \ldots, i_k, \ldots, i_{L+1} = 1, \\ i_j \neq i_k,\ j \neq k}}^{M} \gamma_{i_1 \ldots i_j \ldots i_k \ldots i_{L+1}}. \end{aligned} \tag{15}$$

Next, observe that

$$\begin{aligned}
&M^{-1}\sum_{i_1=1}^{M}\sum_{\mathbf{y}:\, i_1\notin\Phi(\mathbf{y})} \mathbf{E}w^n(\mathbf{y}\,|\,\mathbf{x}_{i_1},\mathbf{S}_{i_2\ldots i_{L+1}}) \\
&= M^{-1}\sum_{i_1=1}^{M}\sum_{\mathbf{y}:\, i_1\notin\Phi(\mathbf{y})}\sum_{\mathbf{s}\in\mathcal{S}^n} w^n(\mathbf{y}\,|\,\mathbf{x}_{i_1},\mathbf{s})\, q^n(\mathbf{s}\,|\,\mathbf{x}_{i_2},\ldots,\mathbf{x}_{i_{L+1}}) \qquad (16)\\
&= \sum_{\mathbf{s}\in\mathcal{S}^n} \overline{p}_L(\mathbf{s})\, q^n(\mathbf{s}\,|\,\mathbf{x}_{i_2},\ldots,\mathbf{x}_{i_{L+1}}).
\end{aligned}$$

The last expression is the error probability $p_L(\mathbf{s})$ averaged over

$$q^n(\mathbf{s}\,|\,\mathbf{x}_{i_2},\ldots,\mathbf{x}_{i_{L+1}}) = \prod_{k=1}^{n} q(s(k)\,|\,x_{i_2}(k),\ldots,x_{i_{L+1}}(k)).$$

In accordance with (14)–(16) we have

$$\begin{aligned}
&M^{-L}\sum_{\substack{i_2,\ldots,i_j,\ldots,i_k,\ldots,i_{L+1}=1,\\ i_j\neq i_k,\ j\neq k}}^{M}\ \sum_{\mathbf{s}\in\mathcal{S}^n}\overline{p}_L(\mathbf{s})\, q^n(\mathbf{s}\,|\,\mathbf{x}_{i_2},\ldots,\mathbf{x}_{i_j},\ldots,\mathbf{x}_{i_k},\ldots,\mathbf{x}_{i_{L+1}}) \\
&\geq M^{-L-1}[(L+1)!]^{-1}\sum_{\substack{i_1,\ldots,i_j,\ldots,i_k,\ldots,i_{L+1}=1,\\ i_j\neq i_k,\ j\neq k}}^{M}\gamma_{i_1\ldots i_j\ldots i_k\ldots i_{L+1}} \\
&\geq M^{-L}(M-1)(M-2)\ldots(M-L)L![(L+1)!]^{-1}\geq(L+1)^{-1}(1+o(1)),\quad M\to\infty.
\end{aligned}$$

From the last expression we deduce the existence of a collection $\{i_2,\ldots,i_{L+1}\}$ such that

$$\sum_{\mathbf{s}\in\mathcal{S}^n}\overline{p}_L(\mathbf{s})\, q^n(\mathbf{s}\,|\,\mathbf{x}_{i_2},\ldots,\mathbf{x}_{i_{L+1}})\geq(L+1)^{-1}(1+o(1)).$$

Hence, there exists a vector $\mathbf{s}\in\mathcal{S}^n$ for which (10) holds, thereby completing the proof of Lemma 1.

Before the proof of the necessity of condition (8) we introduce some standard definitions.

Let $\mathcal{X}_1,\ldots,\mathcal{X}_m$ be finite sets and let $X_1,\ldots X_m$ be random variables with probability distribution

$$p(x_1,\ldots,x_m) = p_{X_1\ldots X_m}(x_1,\ldots,x_m),\quad x_i\in\mathcal{X}_i,\ i=\overline{1,m}.$$

Suppose that

$$H(X_1,\ldots,X_m) = -\sum_{\substack{x_i\in\mathcal{X}_i,\\ i=\overline{1,m}}} p(x_1,\ldots,x_m)\log p(x_1,\ldots,x_m)$$

is the entropy of the distribution $p(x_1,\ldots,x_m)$ of the random variable $(X_1,\ldots,X_m)$,

$$I(X_1;\ldots;X_m) = \sum_{\substack{x_i\in\mathcal{X}_i,\\ i=\overline{1,m}}} p(x_1,\ldots,x_m)\log\frac{p(x_1,\ldots,x_m)}{p(x_1)\ldots p(x_m)},$$

and

$$I(X^k;X_{k+1}^m),\quad 1\leq k<m,$$

is the amount of information between the random variables $X^k=(X_1,\ldots,X_k)$ and $X_{k+1}^m=(X_{k+1},\ldots,X_m)$. Obviously, we have

$$\begin{aligned}
I(X_1;\ldots;X_m) &= \sum_{i=1}^{m}H(X_i)-H(X_1,\ldots,X_m),\\
I(X^k;X_{k+1}^m) &= H(X^k)+H(X_{k+1}^m)-H(X^m).
\end{aligned}$$

Furthermore, for $m \geq 2$ we have

$$\begin{aligned} p_{X_1 \mid X_2} &= \frac{p_{X_1 X_2}}{p_{X_2}}; \\ H(X_1 \mid X_2) &= H(X_1, X_2) - H(X_2); \\ I(X_1; X_2 \mid X_3) &= H(X_1 \mid X_3) + H(X_2 \mid X_3) - H(X_1, X_2 \mid X_3). \end{aligned}$$

Let $p(\cdot)$ and $q(\cdot)$ be distributions on $\mathcal{X}$. Then

$$D(p \,\|\, q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \tag{17}$$

is the (Kullback–Leibler) divergence between the distributions $p(\cdot)$ and $q(\cdot)$, and

$$\|p - q\| = \sum_{x \in \mathcal{X}} |p(x) - q(x)| \tag{18}$$

is the variational distance between $p(\cdot)$ and $q(\cdot)$. It holds (see [9]) that

$$a \, \|p - q\|^2 \leq D(p \,\|\, q), \quad a = \text{const} > 0. \tag{19}$$

We also use the concept of the type of a sequence $\mathbf{x} = (x(1), \dots, x(n)) \in \mathcal{X}^n$. The type of a sequence $\mathbf{x}$ is the distribution $p_{\mathbf{x}}$ on $\mathcal{X}$ given by the formula $p_{\mathbf{x}}(x) = n_x / n$, where n_x is the number of elements in the sequence $\mathbf{x}$ that are equal to x.

Similarly we define the type $p_{\mathbf{xys}}$ of the sequence $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ as a distribution on the product $\mathcal{X} \times \mathcal{Y} \times \mathcal{S}$. This distribution is defined by the formula $p_{\mathbf{xys}}(x, y, s) = n_{xys}/n$, where n_{xys} is the number of triples $(x(i), y(i), s(i))$, $i = \overline{1, n}$, that are equal to (x, y, s).

With the types $p_{\mathbf{x}}$, $p_{\mathbf{xys}}$, we associate random variables X, (X, Y, S) with distributions $p_X = p_{\mathbf{x}}$, $p_{XYS} = p_{\mathbf{xys}}$. We use the following quantities:

$$\begin{aligned} &\tau_X = \{\mathbf{x} \in \mathcal{X}^n : p_{\mathbf{x}} = p_X\}, \\ &\tau_{XY} = \{(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : p_{\mathbf{xy}} = p_{XY}\}, \\ &\tau_{XYS} = \{(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{S}^n : p_{\mathbf{xys}} = p_{XYS}\}, \\ &\tau_{Y \mid X}(\mathbf{x}) = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}) \in \tau_{XY}\}, \\ &\tau_{Y \mid XS}(\mathbf{x}, \mathbf{s}) = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \tau_{XYS}\}. \end{aligned}$$

The following relations are valid (see [10]):

$$(n+1)^{-|\mathcal{X}|} 2^{nH(X)} \leq |\tau_X| \leq 2^{nH(X)}, \quad \text{if } \tau_X \neq \varnothing, \tag{20}$$

$$(n+1)^{-|\mathcal{X}|\,|\mathcal{Y}|} 2^{H(Y \mid X)} \leq |\tau_{Y \mid X}(\mathbf{x})| \leq 2^{nH(Y \mid X)}, \quad \text{if } \tau_{Y \mid X}(\mathbf{x}) \neq \varnothing, \tag{21}$$

$$\sum_{\mathbf{y} \in \tau_{Y \mid XS}(\mathbf{x}, \mathbf{s})} w^n(\mathbf{y} \mid \mathbf{x}, \mathbf{s}) \leq 2^{-nD(p_{XYS} \,\|\, p_{XS} \times w)}, \tag{22}$$

where

$$p_{XS} \times w(x, y, s) = p_{XS}(x, s) w(y \mid x, s).$$

Lemma 2 obviously follows from the next lemma.

Lemma 3. *Let $L > L(G)$. Then there exists $\gamma > 0$ such that for all $R \in (0, \gamma)$,*

$$\nu_L(R) > 0,$$

where

$$\nu_L(R) = \liminf_{n \to \infty} \min_{\mathcal{K}^n : \log M \geq Rn} \max_{\mathbf{s} \in \mathcal{S}^n} (-n^{-1} \log p_L(\mathbf{s})).$$

PROOF. Consider a set of codewords $\mathcal{K}_{\mathbf{x}}^n = \{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ of the same type $p_X(x) = p_{\mathbf{x}}(x)$ that coincides with the type of the fixed word $\mathbf{x} \in \mathcal{X}^n$. Denote by $A_{\mathbf{x}}^n$ the ensemble of all such codes for which codewords are chosen independently and with equal probabilities. Next, we construct a decoding algorithm that we use with such codes. The algorithm consists of two steps.

1. Compose the list of vectors

$$\Gamma = \{\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_N}\} \subset \mathcal{K}_{\mathbf{x}}^n$$

such that for every $\mathbf{x}_i \in \Gamma$, there exists a vector $\mathbf{s}_i \in \mathcal{S}^n$ for which

$$D(p_{\mathbf{x}_i \mathbf{s}_i \mathbf{y}} \,\|\, p_{\mathbf{x}_i} \times q_{\mathbf{s}_i} \times w) \leq \delta_1, \quad \delta_1 > 0, \tag{23}$$

where $p_{\mathbf{x}_i} \times q_{\mathbf{s}_i} \times w(x, s, y) = p_{\mathbf{x}_i}(x) q_{\mathbf{s}_i}(s) w(y \mid x, s)$.

2. Put $\Phi(\mathbf{y}) = i$ if $\mathbf{x}_i \in \Gamma$ and if, for some $\mathbf{s}_i$ satisfying (23) and for any set of vectors $\{\mathbf{x}_{j_1}, \dots, \mathbf{x}_{j_L}\} \subset \Gamma$ not containing $\mathbf{x}_i$, the following relation holds:

$$I((X_i, Y); (X_{j_1}, \dots, X_{j_L}) \,|\, S_i) \leq \delta_1, \tag{24}$$

where $X_i, X_{j_1}, \dots, X_{j_L}, S_i, Y$ are random variables on $\mathcal{X}^{L+1} \times \mathcal{S} \times \mathcal{Y}$ with joint distribution equal to the joint type of $p_{\mathbf{x}_i, \mathbf{x}_{j_1}, \dots, \mathbf{x}_{j_L}, \mathbf{s}_i, \mathbf{y}}$. We set $\Phi(\mathbf{y}) = 1$ if no vector satisfies these conditions. We say that the code $\mathcal{K}_{\mathbf{x}}^n$ can be decoded into a list of size L if the function $\Phi(\cdot)$ takes no more than L values for any $\mathbf{y} \in \mathcal{Y}^n$.

The natural character of the above algorithm is revealed by the following lemma. Put $M = 2^{nR}$.

Lemma 4. *Let $L > L(G)$, and $A_{\mathbf{x}}^n$ be the ensemble of codes $\mathcal{K}_{\mathbf{x}}^n$ with $p_{\mathbf{x}}(x) = p_X(x) > 0$, $x \in \mathcal{X}$. Then for sufficiently small $R, \delta_1 > 0$, the probability that the code from $A_{\mathbf{x}}^n$ is list-of-L decodable tends to* 1 *as* $n \to \infty$.

To prove this lemma, we use the next lemma.

Lemma 5. *The probability that for all collections of $L+1$ vectors $\{\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{L+1}}\} \subset \mathcal{K}_{\mathbf{x}}^n$ the inequality*

$$I(X_{i_1}; X_{i_2}; \dots; X_{i_{L+1}}) < 3L\delta_2, \quad R < \delta_2, \ \delta_2 > 0, \tag{25}$$

is valid tends to 1 *as* $n \to \infty$.

PROOF. We construct the code $\mathcal{K}_{\mathbf{x}}^n$ by choosing codewords independently from the set τ_X of vectors of type $p_{\mathbf{x}}$ with a uniform distribution on it.

Next we estimate the probability ρ of picking a collection $\{\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{L+1}}\} \subset \mathcal{K}_{\mathbf{x}}^n$ of type $p_{\mathbf{z}_{11}, \dots, \mathbf{z}_{1L}}$ for which we have

$$I(Z_{11}; \dots; Z_{1(L+1)}) \geq 3L\delta_2.$$

It is easy to see that for any $\delta_3 > 0$ and for all sufficiently large n, the quantity ρ can be estimated using (20), (21) as follows:

$$\begin{aligned} \rho &= \left|\tau_{Z_{11}}\right| \left|\tau_{Z_{12} \mid Z_{11}}(\mathbf{z}_{11})\right| \cdots \left|\tau_{Z_{1(L+1)} \mid Z_{11} \dots Z_{1L}}(\mathbf{z}_{11}, \dots, \mathbf{z}_{1L})\right| / \left|\tau_{Z_{11}}\right|^{L+1} \\ &\leq 2^{-n[I(Z_{11}; \dots; Z_{1(L+1)}) - \delta_3]}. \end{aligned} \tag{26}$$

We used here (20), (21).

Let λ_1 be the number of subsets of $\mathcal{K}_{\mathbf{x}}^n$ of cardinality $L+1$ and of type $p_{\mathbf{z}_{11} \dots \mathbf{z}_{1(L+1)}} = p_{Z_{11} \dots Z_{1(L+1)}}$. We estimate their average number $\overline{\lambda}_1$. Since $M = 2^{nR}$, the number of all the sets of $L+1$ vectors equals C_M^{L+1}, and

$$\overline{\lambda}_1 = \rho C_M^{L+1} \leq 2^{n[R(L+1) - I(Z_{11}; \dots; Z_{1(L+1)}) + \delta_3]}.$$

A similar relation holds for another distribution $p_{Z_{21} \dots Z_{2(L+1)}}$, and so on. We examine all the distributions for which $I(Z_{j1}; \dots; Z_{j(L+1)}) \geq 3L\delta_2$, $j = 1, \dots m$ (the set $\{Z_{j1}, \dots, Z_{j(L+1)}\}$ which repeats the distribution already used is not considered in later steps). We obtain

$$\overline{\lambda} = \sum_j \overline{\lambda}_j \leq \sum_j 2^{n[R(L+1) - I(Z_{j1}; \dots; Z_{j(L+1)}) + \delta_3]} \leq m 2^{n[R(L+1) - 3L\delta_2 + \delta_3]}. \tag{27}$$

Here $\overline{\lambda}_j$ is the mathematical expectation of the quantity λ_j for the jth distribution, $\overline{\lambda}$ is the mathematical expectation of the number of collections $\{\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_{L+1}}\} \subset \mathcal{K}_{\mathbf{x}}^n$ for which $I(X_{i_1}; \ldots; X_{i_{L+1}}) \geq 3L\delta_2$, m is the number of elements in the sum (27),

$$m \leq cn^{\alpha},$$

where c and α are constants.

According to Chebyshev's inequality, $\Pr(\lambda \geq \overline{\lambda}n) \leq n^{-1}$. Hence, if $\overline{\lambda}n < 1$, then with probability not less than $1 - n^{-1}$ for any collection $\{\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_{L+1}}\} \subset \mathcal{K}_{\mathbf{x}}^n$, inequality (25) holds. According to (27), for inequality $\overline{\lambda}n < 1$ to be valid, it is sufficient that $R(L+1) - 3L\delta_2 + \delta_3 < 0$.

Since $\delta_3 > 0$ is arbitrary, the latter inequality is true if $R < \delta_2$, thus establishing Lemma 5.

Proof of Lemma 4. Consider a code $\mathcal{K}_{\mathbf{x}}^n$ that satisfies (25). Now it suffices to show that for sufficiently small $\delta_1, \delta_2 > 0$ the decoding result is a set that contains no more than L codewords.

Let us assume the contrary, i.e., that for any $\delta_1, \delta_2 > 0$ there exists a vector $\mathbf{y}$, a collection of $L+1$ different codewords $\{\mathbf{x}_{k_1}, \ldots, \mathbf{x}_{k_{L+1}}\} \subset \mathcal{K}_{\mathbf{x}}^n$, and corresponding vectors $\{\mathbf{s}_{k_1}, \ldots, \mathbf{s}_{k_{L+1}}\}$ such that the conditions 1 and 2 of the decoding algorithm are fulfilled. We show that in this case (9) holds.

For simplicity of calculations, we assume that $k_i = i$. According to the definitions of divergence and mutual information, conditions 1 and 2 of the decoding algorithm imply

$$\begin{aligned}
2\delta_1 &\geq D(p_{X_i S_i Y} \,\|\, p_{X_i} \times q_{S_i} \times w) + I\big((X_i, Y); (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{L+1}) \,|\, S_i\big) \\
&= \sum_{\substack{x_1, \ldots, x_{L+1} \in \mathcal{X}, \\ s \in \mathcal{S},\ y \in \mathcal{Y}}} p_{X_1 \ldots X_{L+1} S_i Y}(x_1, \ldots, x_{L+1}, s, y) \\
&\times \log \frac{p_{X_1 \ldots X_{L+1} S_i Y}(x_1, \ldots, x_{L+1}, s, y)}{p_{X_1 \ldots X_{i-1} X_{i+1} \ldots X_{L+1} S_i}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{L+1}, s) p_{X_i}(x_i) w(y \,|\, x_i, s)}.
\end{aligned}$$

The last expression is the divergence of two distributions defined on the space $\mathcal{X}^{L+1} \times \mathcal{Y} \times \mathcal{S}$. Restricting these distributions to the space $\mathcal{X}^{L+1} \times \mathcal{Y}$ and taking into account that this does not increase divergence, we obtain

$$D(p_{X_1 \ldots X_{L+1} Y} \,\|\, p_{X_1 \ldots X_{i-1} X_{i+1} \ldots X_{L+1}} \times p_{X_i} \times w_i) \leq 2\delta_1. \tag{28}$$

Here

$$w_i(y \,|\, x_1, \ldots, x_{L+1}) = \sum_{s \in \mathcal{S}} w(y \,|\, x_i, s)\, q_i(s \,|\, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{L+1}),$$

with $q_i(s \,|\, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{L+1})$ being the conditional distribution defined by the random variable $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{L+1}, S)$.

Comparing (19) and (28), we obtain the following inequality:

$$\|p_{X_1 \ldots X_{L+1} Y} - p_{X_1 \ldots X_{i-1} X_{i+1} \ldots X_{L+1}} \times p_{X_i} \times w_i\| \leq \sqrt{2\delta_1 / a}. \tag{29}$$

On the other hand, (25) and (19) together imply

$$\begin{aligned}
&\|p_{X_1} \times \ldots \times p_{X_{L+1}} \times w_i - p_{X_1 \ldots X_{i-1} X_{i+1} \ldots X_{L+1}} \times p_{X_i} \times w_i\| \\
&\quad = \|p_{X_1} \times \ldots \times p_{X_{L+1}} - p_{X_1 \ldots X_{i-1} X_{i+1} \ldots X_{L+1}} \times p_{X_i}\| \\
&\quad \leq \|p_{X_1} \times \ldots \times p_{X_{L+1}} - p_{X_1 \ldots X_{L+1}}\| \leq \sqrt{3L\delta_2 / a}.
\end{aligned} \tag{30}$$

Under the assumption that $\delta_1 < \delta_2$, we derive from (29) and (30) the following inequality:

$$\|p_{X_1 \ldots X_{L+1} Y} - p_{X_1} \times \ldots \times p_{X_{L+1}} \times w_i\| \leq 2\sqrt{3L\delta_2 / a} \tag{31}$$

and hence,

$$\max_{\substack{x_1, \ldots, x_{L+1}; \\ y;\ i,j}} \big|w_i(y \,|\, x_1, \ldots, x_{L+1}) - w_j(y \,|\, x_1, \ldots, x_{L+1})\big| \leq \frac{\delta_4}{\min_x p_X^{L+1}(x)}, \tag{32}$$

where $\delta_4 = 4\sqrt{3L\delta_2/a}$; $i, j = \overline{1, L+1}$.

We introduce the following conditional probability:

$$q(s \mid x_2, \ldots, x_{L+1}) = \left[(L+1)!\right]^{-1} \sum_{i=1}^{L+1} \sum_{\pi^*} q_i(s \mid x_{\pi_2}, \ldots, x_{\pi_{L+1}}), \tag{33}$$

where the inner sum is taken over all permutations $\pi^* = (\pi_2, \ldots, \pi_{L+1})$ of the sequence $(2, \ldots, L+1)$.

Now we shall prove the inequality

$$\left| \sum_s (w(y \mid x_1, s)\, q(s \mid x_2, \ldots, x_{L+1}) - w(y \mid x_{\pi_1}, s)\, q(s \mid x_{\pi_2}, \ldots, x_{\pi_{L+1}})) \right| \le \frac{\delta_4}{\min\limits_x p_X^{L+1}(x)}. \tag{34}$$

We remark that, according to (33), $\sum\limits_{s \in \mathcal{S}} w(y \mid x_1, s)\, q(s \mid x_2, \ldots, x_{L+1})$ can be represented as the sum of $(L+1)!$ terms

$$\left[(L+1)!\right]^{-1} w_i(y \mid x_{\pi'_1}, \ldots, x_{\pi'_{L+1}}), \quad i = 1, \ldots, L+1,$$

such that every permutation $\pi' = (\pi'_1, \ldots, \pi'_{L+1})$ of $(1, \ldots, L+1)$ occurs. Likewise, the expression

$$\sum_{s \in \mathcal{S}} w(y \mid x_{\pi_1}, s)\, q(s \mid x_{\pi_2}, \ldots, x_{\pi_{L+1}})$$

can be represented as the sum of terms $\left[(L+1)!\right]^{-1} w_j(y \mid x_{\pi'_1}, \ldots, x_{\pi'_{L+1}})$. Then it follows that the sum in the left-hand side of (34) can be represented as the sum of $(L+1)!$ differences

$$\left[(L+1)!\right]^{-1} \left(w_i(y \mid x_{\pi'_1}, \ldots, x_{\pi'_{L+1}}) - w_j(y \mid x_{\pi'_1}, \ldots, x_{\pi'_{L+1}})\right).$$

Thus, the left-hand side of (34) is not greater than the sum of the absolute values of these differences. From (32) and (33) we obtain that the sum of the absolute values of these differences is not greater than

$$\frac{(L+1)!\, \delta_4}{(L+1)! \min\limits_x p_X^{L+1}(x)} = \frac{\delta_4}{\min\limits_x p_X^{L+1}(x)}.$$

The estimate (34) is established.

Since $\delta_4 > 0$ can be chosen arbitrarily small, relation (9) follows from (34) and, thus, relation (8) is also valid. This contradicts the condition $L > L(G)$ in the assumption of Lemma 4. This completes the proof of Lemma 4.

Now we complete the proof of Lemma 3. It remains to show that in the ensemble of codes considered, for sufficiently small R there exist codes such that the above decoding algorithm leads to an error probability which decreases exponentially uniformly over the states $\mathbf{s} \in \mathcal{S}^n$.

Here we use two auxiliary lemmas. Suppose again that $A^n_{\mathbf{x}}$ is the ensemble of codes $\mathcal{K}^n_{\mathbf{x}} = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$ of cardinality $M = 2^{nR}$ and codeword type $p_{\mathbf{x}}(x) = p_X(x)$, $x \in \mathcal{X}$.

Let $|a|^+$ denote a if $a > 0$, and 0 if $a \le 0$.

Lemma 6. *For all $\delta_5 > 0$, the inequality*

$$\left|\left\{i : \mathbf{x}_i \in \tau_{X \mid S}(\mathbf{s})\right\}\right| \le 2^{n\left(|R - I(X;S)|^+ + \delta_5\right)} \tag{35}$$

holds for all $\mathbf{s} \in \mathcal{S}^n$ with probability tending to 1 as $n \to \infty$.

This lemma is a simple corollary of the Chernoff estimate (see, e.g., [6, Eq. (A.8)]).

Lemma 7. *With probability tending to 1 as $n \to \infty$ for all $\mathbf{s} \in \mathcal{S}^n$ and all types $p_{\mathbf{x}\mathbf{z}_1 \ldots \mathbf{z}_L \mathbf{s}} = p_{XZ_1 \ldots Z_L S}$, $\mathbf{x}, \mathbf{z}_1, \ldots, \mathbf{z}_L \in \tau_X$, such that*

$$I(X; (Z_1, \ldots, Z_L, S)) \ge R(L+1), \tag{36}$$

the following inequalities hold:

$$\left|\{i : (\mathbf{x}_i, \mathbf{x}_{j_1}, \ldots, \mathbf{x}_{j_L}, \mathbf{s}) \in \tau_{XZ_1\ldots Z_LS};\ j_l \neq j_k \textit{ for } l \neq k;\ j_1, \ldots, j_L \neq i\}\right| \leq 2^{nR/3}. \tag{37}$$

PROOF. The proof is based on the following auxiliary result (see [4; 6, Lemma A.1]).

Suppose that $\zeta_1, \ldots, \zeta_M$ are random variables taking values in $\mathcal{X}^n$ and $0 \leq f_i(u_1, \ldots, u_i) \leq 1$, $i = 1, \ldots, M$, are functions of arguments $u_1, \ldots, u_i \in \mathcal{X}^n$. If the conditions

$$\mathbf{E}\{(f_i(\zeta_1, \ldots, \zeta_i) \mid \zeta_1, \ldots, \zeta_{i-1})\} \leq a \quad \text{a.s.}, \quad i = 1, \ldots, M, \tag{38}$$

hold, then

$$\Pr\left\{\sum_{i=1}^{M} f_i(\zeta_1, \ldots, \zeta_i) \geq Mt\right\} \leq 2^{-M(t - a\log_2 e)}. \tag{39}$$

Suppose now that the ζ_i are independent random variables uniformly distributed on τ_X. Let $\mathbf{x}_i$ be understood as a value of the random variable ζ_i. Further, let $f_i(\zeta_1, \ldots, \zeta_i) = 1$ if among the set of vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}\}$ there exists a collection $\{\mathbf{x}_{j_1}, \ldots, \mathbf{x}_{j_L}\}$ such that $p_{\mathbf{x}_i\mathbf{x}_{j_1}\ldots\mathbf{x}_{j_L}\mathbf{s}} = p_{XZ_1\ldots Z_LS}$, and $f_i(\zeta_1, \ldots, \zeta_i) = 0$ otherwise, i.e., $f_i(\zeta_1, \ldots, \zeta_i)$ is the characteristic function of the following event: In the set of vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}\}$ there exists a subset $\{\mathbf{x}_{j_1}, \ldots, \mathbf{x}_{j_L}\}$ such that

$$p_{\mathbf{x}_i\mathbf{x}_{j_1}\ldots\mathbf{x}_{j_L}\mathbf{s}} = p_{XZ_1\ldots Z_LS}.$$

In accordance with (20), (21), for n sufficiently large, we have

$$\begin{aligned}
&\mathbf{E}\{(f_i(\zeta_1, \ldots, \zeta_i) \mid \zeta_1, \ldots, \zeta_{i-1})\} \\
&= \Pr\Bigg\{\zeta_i \in \bigcup_{\substack{j_1, \ldots, j_L < i, \\ (\mathbf{x}_{j_1}, \ldots, \mathbf{x}_{j_L}) \in \tau_{Z_1\ldots Z_L \mid S}(\mathbf{s})}} \tau_{X \mid Z_1\ldots Z_LS}(\mathbf{x}_{j_1}, \ldots, \mathbf{x}_{j_L}, \mathbf{s}) \mid \zeta_1, \ldots, \zeta_{i-1})\Bigg\} \\
&\leq \frac{2^{nRL}2^{nH(X \mid Z_1, \ldots, Z_L, S) + Rn/4}}{2^{nH(X)}} = 2^{n\left[RL - I(X;(Z_1, \ldots, Z_L, S)) + R/4\right]} \triangleq b.
\end{aligned}$$

A comparison of these expressions with (36) shows that $b \leq 2^{-3nR/4}$. Using (39), we obtain

$$\begin{aligned}
&\Pr\left\{\left|\{i : (\mathbf{x}_i, \mathbf{x}_{j_1}, \ldots, \mathbf{x}_{j_L}, \mathbf{s}) \in \tau_{XZ_1\ldots Z_LS},\ j_1, \ldots, j_L < i\}\right| \geq \frac{1}{n^2}2^{nR/3}\right\} \\
&\leq\ 2^{-\frac{1}{n^2}2^{nR/3} + 2^{nR/4}\log_2 e} \leq 2^{-2^{nR/4}}.
\end{aligned} \tag{40}$$

Under the assumption that $j_1, \ldots, j_L < i$, this implies Lemma 7. To eliminate this latter restriction, we use the following procedure. We form n^2 sequences by rewriting n^2 times the sequence of M messages $m_1, \ldots, m_M$ and numbering messages in all the sequences independently and with uniform distribution. Then the probability $\widehat{p}$ that in all these sequences the number i_k, $k = 1, \ldots, n^2$, of the message m_i does not exceed the numbers $j_{1_k}, \ldots, j_{L_k}$ of the messages $m_{j_1}, \ldots, m_{j_L}$ can be estimated by the following expression:

$$\widehat{p} = (L/(L+1))^{n^2} \leq e^{-n^2/(L+1)}.$$

The probability p^* that this happens with at least one message m_i and at least one of C_{M-1}^L collections $\{m_{j_1}, \ldots, m_{j_L}\}$, is estimated by the following formula:

$$p^* \leq M(M-1)^L e^{-n^2/(L+1)} \leq e^{-n^2/(2(L+1))}.$$

Let us fix the set of n^2 numberings $\{1, \ldots, n^2\}$ such that for arbitrary m_i and an arbitrary sequence of messages $m_{j_1}, \ldots, m_{j_L}$, there exists a numbering k of this set such that $i_k > j_{1_k}, \ldots, j_{L_k}$.

Applying the union bound for the probability of the sum of events and using inequalities (40), we obtain

$$\begin{aligned}
&\Pr\left\{\left|\{i : (\mathbf{x}_i, \mathbf{x}_{j_1}, \dots, \mathbf{x}_{j_L}, \mathbf{s}) \in \tau_{XZ_1\dots Z_L S};\ j_1, \dots, j_L \neq i,\ j_l \neq j_m,\ l \neq m\}\right| \geq 2^{nR/3}\right\} \\
&= \Pr\left\{\left|\left\{\bigcup_{\substack{\text{over all } n^2 \text{ numberings} \\ k=1,\dots,n^2}} i_k : (\mathbf{x}_i, \mathbf{x}_{j_1}, \dots, \mathbf{x}_{j_L}, \mathbf{s}) \in \tau_{XZ_1\dots Z_L S},\ j_{1_k}, \dots, j_{L_k} < i_k\right\}\right| \geq 2^{nR/3}\right\} \\
&\leq \sum_{k=1}^{n^2} \Pr\left\{\left|\{i_k : (\mathbf{x}_i, \mathbf{x}_{j_1}, \dots, \mathbf{x}_{j_L}, \mathbf{s}) \in \tau_{XZ_1\dots Z_L S};\ j_{1_k}, \dots, j_{L_k} < i_k\}\right| \geq \frac{1}{n^2} 2^{nR/3}\right\} \\
&\leq n^2 2^{-2^{nR/4}}.
\end{aligned}$$

The last inequalities imply Lemma 7.

Now we consider a code $\mathcal{K}_{\mathbf{x}}^n$ that satisfies the conditions of Lemmas 4, 5, 6, and 7. We shall prove that for such a code and for some $R > 0$, the average probability of the list-of-L decoding error tends to 0 exponentially and uniformly over the choice of the state $\mathbf{s} \in \mathcal{S}^n$. This completes the proof of Lemma 3.

According to (35), for $R > 2\varepsilon > 0$, we have

$$\begin{aligned}
2^{-nR}\left|\left\{i : \mathbf{x}_i \in \bigcup_{I(X;S)\geq\varepsilon} \tau_{X\,|\,S}(\mathbf{s})\right\}\right| &\leq 2^{-nR} \sum_{I(X;\S)\geq\varepsilon} \left|\{i : \mathbf{x}_i \in \tau_{X\,|\,S}(\mathbf{s})\}\right| \\
&\leq mn^{-n\varepsilon} \leq 2^{-n\varepsilon/2},
\end{aligned}$$

where m is the number of types p_{XS} with $I(X;S) \geq \varepsilon$. Thus, the contribution to the error probability of the list-of-L decoding of the codewords $\mathbf{x}_i \in \tau_{X\,|\,S}(\mathbf{s})$ for which $I(X;S) \geq \varepsilon$ is exponentially small when $n \to \infty$, and it suffices to consider only those $\mathbf{x}_i$ for which

$$I(X;S) < \varepsilon. \tag{41}$$

The error will occur if the set of codewords $\mathbf{x}_{j_1}, \dots, \mathbf{x}_{j_k}$, $k \leq L$, that forms the decoding result does not contain the transmitted word $\mathbf{x}_i$. This event occurs if condition 1 or 2 of the decoding algorithm does not hold for the transmitted codeword $\mathbf{x}_i$.

First, we estimate the probability that condition 1, i.e., inequality (23), does not hold. For the transmitted word $\mathbf{x}_i$ and the state $\mathbf{s}$, this probability can be written as follows:

$$p^{(1)}(i, \mathbf{s}) \triangleq \sum_{P_{XSY}} \sum_{\mathbf{y}\in\tau_{Y\,|\,XS}(\mathbf{x}_i,\mathbf{s})} w^n(\mathbf{y}\,|\,\mathbf{x}_i, \mathbf{s}),$$

where the outer sum is taken over all the types with

$$D(p_{XSY}\,\|\,p_X \times q_S \times w) \geq \delta_1$$

and $m \leq n^\alpha$, $\alpha > 0$, is the number of such types.

For $\delta_1 > 2\varepsilon$, (41) implies that

$$D(p_{XSY}\,\|\,p_{XS} \times w) = D(p_{XSY}\,\|\,p_X \times q_S \times w) - I(X;S) \geq \varepsilon,$$

and according to (22) and (23), the value $p^{(1)}(i, \mathbf{s})$ can be estimated as follows:

$$p^{(1)}(i, \mathbf{s}) \leq m 2^{-n\varepsilon} \leq 2^{-n\varepsilon/2},$$

i.e., the probability that condition 1 does not hold for the transmitted codeword $\mathbf{x}_i$ exponentially decreases in n uniformly over the choice of the state $\mathbf{s} \in \mathcal{S}^n$.

Now let us estimate the contribution to the average error probability of the situation where condition 2 of the decoding algorithm does not hold for the transmitted codeword $\mathbf{x}_i$ and the state $\mathbf{s}$. This contribution can be estimated as follows:

$$p^{(2)}(i,\mathbf{s}) = \sum_{p_{XZ_1\dots Z_LSY}} e_{XZ_1\dots Z_LSY}(i,\mathbf{s}),$$

where the sum is taken over all the types $p_{XZ_1\dots Z_LSY}$ for which

$$I((X,Y);(Z_1,\dots,Z_L)\,|\,S) \geq \delta_1 \tag{42}$$

and

$$e_{XZ_1\dots Z_LSY}(i,\mathbf{s}) = \sum_{j_1,\dots,j_L:(\mathbf{x}_i,\mathbf{x}_{j_1},\dots,\mathbf{x}_{j_L},\mathbf{s})\in\tau_{XZ_1\dots Z_LS}} \quad \sum_{\mathbf{y}\in\tau_{Y\,|\,XZ_1\dots Z_LS}(\mathbf{x}_i,\mathbf{x}_{j_1},\dots,\mathbf{x}_{j_L},\mathbf{s})} w^n(\mathbf{y}\,|\,\mathbf{x}_i,\mathbf{s}).$$

Since $w^n(\mathbf{y}\,|\,\mathbf{x}_i,\mathbf{s})$ is a constant for $\mathbf{y}\in\tau_{Y\,|\,XS}(\mathbf{x}_i,\mathbf{s})$ and this constant is at most $\left|\tau_{Y\,|\,XS}(\mathbf{x}_i,\mathbf{s})\right|^{-1}$, the inner sum is bounded from above by the quantity

$$\left|\tau_{Y\,|\,XZ_1\dots Z_LS}(\mathbf{x}_i,\mathbf{x}_{j_1},\dots,\mathbf{x}_{j_L},\mathbf{s})\right|\left|\tau_{Y\,|\,XS}(\mathbf{x}_i,\mathbf{s})\right|^{-1},$$

which in turn, according to (20) and (21), is not greater than $2^{-n[I(Y;(Z_1,\dots,Z_L)\,|\,X,S)-\varkappa]}$ for any $\varkappa > 0$ and sufficiently large n. From this we obtain

$$e_{XZ_1\dots Z_LSY}(i,\mathbf{s}) \leq 2^{n[RL-I(Y;(Z_1,\dots,Z_L)\,|\,X,S)+\varkappa]}. \tag{43}$$

Thus,

$$M^{-1}\sum_{i=1}^{M} p^{(2)}(i,\mathbf{s}) = M^{-1}\sum_{i=1}^{M}\sum_{p_{XZ_1\dots Z_LSY}} e_{XZ_1\dots Z_LSY}(i,\mathbf{s}), \tag{44}$$

where the inner sum in the last expression is taken over all the types that satisfy inequality (42). According to (37), the sum of the terms in (44) with

$$I(X;(Z_1,\dots,Z_L,S)) \geq R(L+1)$$

does not exceed $2^{nR/2}$ and hence the contribution of these terms to the expression (44) is at most $2^{-nR/2}$. Now let us estimate the contribution of the terms with

$$I(X;(Z_1,\dots,Z_L,S)) < R(L+1). \tag{45}$$

From (43) and (44), it follows that this contribution can be bounded from above by the quantity

$$M^{-1}\sum_{i=1}^{M}\sum_{p_{XZ_1\dots Z_LS}} 2^{n[RL-I(Y;(Z_1,\dots,Z_L)\,|\,X,S)+\varkappa]}, \tag{46}$$

where the inner sum is taken over the distributions that satisfy (42) and (45). Next, we have

$$\begin{aligned} I(Y;(Z_1,\dots,Z_L)\,|\,X,S) &= I((X,Y);(Z_1,\dots,Z_L)\,|\,S) - I(X;(Z_1,\dots,Z_L)\,|\,S) \\ &\geq \delta_1 - I(X;(Z_1,\dots,Z_L)\,|\,S) \geq \delta_1 - I(X;(Z_1,\dots,Z_L,S)) \\ &> \delta_1 - R(L+1). \end{aligned}$$

Therefore, for $R < \delta_1/4(L+1)$ and n sufficiently large, expression (44) can be bounded from above by the quantity

$$2^{n[\delta_1/4-\delta_1+\delta_1/4+\varkappa]} + 2^{-nR/2} = 2^{-n[\delta_1/2-\varkappa]} + 2^{-nR/2}.$$

Thus we have shown that for $R,\delta_1 > 0$ sufficiently small, there exists a list-of-L decoding code whose average error probability $\overline{p}_L(\mathbf{s})$ tends to zero exponentially as $n \to \infty$ and uniformly over the choice of the state $\mathbf{s}\in\mathcal{S}^n$.

Lemma 3 is proved, and so are Lemma 2 and Theorem 2.

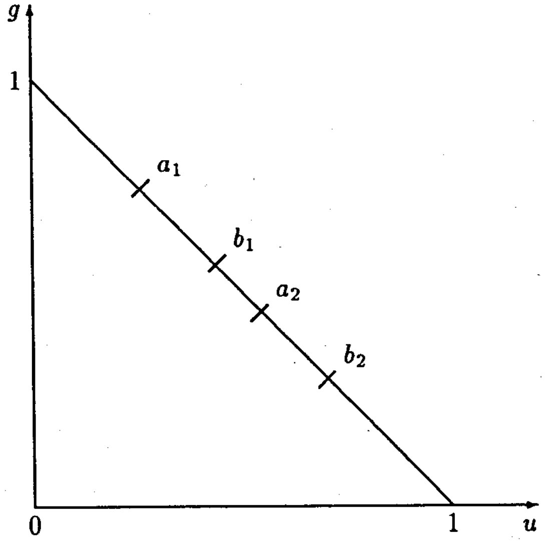

Fig. 1

4. Proof of Theorem 3

For a binary AVC, relations (8) can be reduced to the system of L equations

$$\begin{aligned}
\sum_{s=0,1} w(y\,|\,1,s)\,q(s\,|\,0,\dots,0) &= \sum_{s=0,1} w(y\,|\,0,s)\,q(s\,|\,1,\dots,0)\\
&= \dots = \sum_{s=0,1} w(y\,|\,0,s)\,q(s\,|\,0,\dots,1), \qquad (i_1)\\
\sum_{s=0,1} w(y\,|\,1,s)\,q(s\,|\,1,\dots,0) &= \dots = \sum_{s=0,1} w(y\,|\,1,s)\,q(s\,|\,0,\dots,1)\\
= \sum_{s=0,1} w(y\,|\,0,s)\,q(s\,|\,1,1,0,\dots,0) &= \dots = \sum_{s=0,1} w(y\,|\,0,s)\,q(s\,|\,0,\dots,0,1,1), \qquad (i_2)\\
&\vdots\\
\sum_{s=0,1} w(y\,|\,1,s)\,q(s\,|\,1,1,\dots,1,0) &= \dots = \sum_{s=0,1} w(y\,|\,1,s)\,q(s\,|\,0,1,\dots,1)\\
&= \sum_{s=0,1} w(y\,|\,0,s)\,q(s\,|\,1,\dots,1). \qquad (i_L)
\end{aligned}$$

We need to show that for fixed $w(y\,|\,x,s)$, $y,x,s \in \{0,1\}$, these equations are inconsistent starting from some L.

Toward this end, let us consider Fig. 1. On the plane $\mathbb{R}^2$, there are two pairs of points $(\mathbf{a}_1,\mathbf{b}_1)$ and $(\mathbf{a}_2,\mathbf{b}_2)$, where

$$\mathbf{a}_1 = \big(w(0\,|\,1,0), w(1\,|\,1,0)\big), \qquad \mathbf{a}_2 = \big(w(0\,|\,1,1), w(1\,|\,1,1)\big),$$
$$\mathbf{b}_1 = \big(w(0\,|\,0,0), w(1\,|\,0,0)\big), \qquad \mathbf{b}_2 = \big(w(0\,|\,0,1), w(1\,|\,0,1)\big),$$

with nonnegative coordinates that lie on the straight line $u + v = 1$. Consider the segments $A = [\mathbf{a}_1, \mathbf{a}_2]$, $B = [\mathbf{b}_1, \mathbf{b}_2]$. Evidently, the sums

$$\sum_{s=0,1} w(y\,|\,x,s)\,q(s\,|\,x_2,\dots,x_{L+1}), \quad y = 0,1,$$

are the coordinates of some point $\mathbf{a}$ on the segment A if $x = 1$, and of some point $\mathbf{b}$ on segment B if $x = 0$. The quantities $q(s\,|\,x_2, \dots, x_{L+1})$, $s = 0, 1$, are the barycentric coordinates of the points $\mathbf{a}$ and $\mathbf{b}$, respectively.

Each of the conditions (i_1)–(i_L) implies the existence of a common point on the segments A and B, i.e., $A \cap B \neq \varnothing$. If at least one of the segments A or B degenerates into the point and $A \cap B \neq \varnothing$, then clearly the expression for C_r implies $C_r = 0$.

Thus, it remains to consider the situation where segments A and B do not degenerate into points, i.e., $\mathbf{a}_1 \neq \mathbf{a}_2$ and $\mathbf{b}_1 \neq \mathbf{b}_2$. In this case, it follows from conditions (i_1) that the probability $q(s\,|\,1, \dots, 0)$ does not change under any permutation of the indices 0 and 1. Similarly, the conditions (i_k), $k = \overline{1, L}$, imply that the probabilities $q(s\,|\,1, \dots, 1, 0, \dots, 0)$ do not change after any permutation of indices 0 and 1.

Let

$$\alpha_k = q(0\,|\,\underbrace{1, \dots, 1}_{k-1}, 0, \dots, 0); \quad k = \overline{1, L+1}.$$

The condition (i_k), $k = \overline{1, L}$, yields the equation

$$\mathbf{a}_1 \alpha_k + \mathbf{a}_2 (1 - \alpha_k) = \mathbf{b}_1 \alpha_{k+1} + \mathbf{b}_2 (1 - \alpha_{k+1}),$$

which we rewrite in the form

$$(\mathbf{a}_1 - \mathbf{a}_2)\alpha_k + \mathbf{a}_2 = (\mathbf{b}_1 - \mathbf{b}_2)\alpha_{k+1} + \mathbf{b}_2. \tag{47}$$

Next, we introduce some notation. Suppose that $\mathbf{e} = (u, v)$; $e_1 = u$ and $e_2 = v$ are the first and second coordinates of the vector $\mathbf{e} \in \mathbb{R}^2$. Let

$$\ell \triangleq (\mathbf{a}_1 - \mathbf{a}_2)_1 / (\mathbf{b}_1 - \mathbf{b}_2)_1; \qquad m \triangleq (\mathbf{a}_2 - \mathbf{b}_2)_1 / (\mathbf{b}_1 - \mathbf{b}_2)_1.$$

Clearly $(\mathbf{a}_1 - \mathbf{a}_2)_1 (\mathbf{b}_1 - \mathbf{b}_2)_1 \neq 0$, $\ell \neq 0$, since the points $\mathbf{a}_1, \mathbf{a}_2, \mathbf{b}_1, \mathbf{b}_2$ lie on the line $u + v = 1$. In addition, we can suppose that $|\ell| \geq 1$; otherwise we exchange segments A and B.

From the recursion equations (47) we obtain

$$\alpha_{k+1} = \ell \alpha_k + m; \quad k = \overline{1, L}. \tag{48}$$

Hence

$$\alpha_{L+1} = \alpha_1 \ell^L + m \sum_{k=1}^{L} \ell^{k-1}. \tag{49}$$

Here, we distinguish among three cases.

1. $|\ell| > 1$; in this case expression (49) can be reduced to the form

$$\alpha_{L+1} = \ell^L \big(\alpha_1 + m/(\ell - 1)\big) - m/(\ell - 1). \tag{50}$$

If $-m(\ell - 1) \geq 0$, then $\alpha_1 = -m/(\ell - 1)$ is a fixed point of transformation (48) and for

$$q(s) = \begin{cases} \alpha_1, & s = 0, \\ 1 - \alpha_1, & s = 1, \end{cases}$$

we have

$$w_q(y\,|\,0) \triangleq \sum_{s=0,1} w(y\,|\,0, s)\, q(s) = \sum_{s=0,1} w(y\,|\,1, s)\, q(s) \triangleq w_q(y\,|\,1).$$

This means that $C_r = 0$.

If $m(\ell - 1) > 0$, as seen from (50), for any $\alpha_1 \in [0, 1]$, there exists an interval $\delta(\alpha_1)$ which contains α_1 such that starting from a sufficiently large L, we have $|\alpha^*_{L+1}| > 1$ for $\alpha^*_1 \in \delta(\alpha_1)$, i.e., α^*_{L+1} cannot be a probability.

From such a covering of the segment $[0,1]$, according to the Heine–Borel lemma, we can choose a finite number of intervals that cover the segment $[0,1]$. We take the maximum number $L_{\max}$ among the quantities L that correspond to these intervals and find that for any $\alpha_1 \in [0,1]$, the quantity L cannot exceed $L_{\max}$.

2. $\ell = 1$. In this case, expression (48) leads to the following equation:

$$\alpha_{L+1} = \alpha_1 + mL.$$

If $m = 0$, then $A = B$ and every $\alpha_1 \in [0,1]$ is a fixed point of transformation (47). This means that $C_r = 0$. If $m \neq 0$, then for all $\alpha_1 \in [0,1]$ starting from some L we have $|\alpha_{L+1}| > 1$, i.e., α_{L+1} cannot be a probability.

3. $\ell = -1$. In this case, $\alpha_1 = m/2$ is a fixed point of transformation (48). This means that, in this case, $C_r = 0$.

Remark. It is not difficult to see that if $l = 1$ and m is small, then $L(G)$ can be arbitrarily large.

REFERENCES

1. R. Ahlswede, "Elimination of correlation in random codes for arbitrary varying channels," *Z. Wahrscheinlichkeitsrechnung verw. Geb.*, **44**, 159–175 (1978).
2. P. Elias, "List decoding for noisy channels," *IRE WESCON Convention Record*, Pt. 2 (1957), pp. 94–104.
3. J. Wozencraft, "List decoding," *Quarterly Progress Report, Research Laboratory of Electronics*, MIT, **48**, 90–95 (1958).
4. R. Dobrushin and S. Stambler, "Coding theorems for classes of arbitrarily varying discrete memoryless channels," *Probl. Peredachi Inf.*, **11**, No. 2, 3–28 (1975).
5. T. Ericson, "Exponential error bounds for random codes in arbitrary varying channels," *IEEE Trans. Inf. Theory*, **31**, 42–48 (1985).
6. I. Csiszár and P. Narayan, "The capacity of the arbitrary varying channel revisited: positivity, constraints," *IEEE Trans. Inf. Theory*, **34**, 181–193 (1988).
7. D. Blackwell, L. Breiman, and A. Thomasian, "The capacities of certain channel classes under random coding," *Ann. Math. Stat.*, **31**, 558–567 (1960).
8. R. Ahlswede and N. Cai, "Two proofs of Pinsker's conjecture concerning AV channels," *IEEE Trans. Inf. Theory*, **37**, 1647–1649 (1991).
9. M. Pinsker, *Information and Information Stability of Random Variables and Processes* [in Russian], *Probl. Peredachi Inf.*, No. 7, Moscow (1960).
10. I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York (1981).