Z. Wahrscheinlichkeitstheorie verw. Gebiete
34, 157–177 (1976)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Bounds on Conditional Probabilities with Applications in Multi-User Communication

R. Ahlswede*[1], P. Gács[2] and J. Körner[2]

[1] Dept. of Mathematics, The Ohio State University, 231 W 18th Avenue, Columbus, Ohio 43210, USA
[2] Mathematical Institute of the Hungarian Academy of Sciences, H-1053 Budapest, Reáltanoda u. 13–15, Hungary

We consider a sequence $\{Z_i\}_{i=1}^{\infty}$ of independent, identically distributed random variables where each Z_i is a pair (X_i, Y_i). For any pair of events $\{X^n \in \mathscr{A}\}$, $\{Y^n \in \mathscr{B}\}$ satisfying $\Pr(Y^n \in \mathscr{B} | X^n \in \mathscr{A}) \geqq 1-\varepsilon$ and for any non-negative real c we investigate how small $\Pr(Y^n \in \mathscr{B})$ can be in case $\Pr(X^n \in \mathscr{A})$ is larger than 2^{-nc}. We give the full answer to a generalized form of this question.

These estimates enable us to prove strong converses of the coding theorems for two recently emerged questions in Shannon's information theory, i.e. the source coding problem with side information and the coding problem for the degraded broadcast channel.

1. Statement of Problems and Results

The concept of a decoding set $\mathscr{B}$ corresponding to a sequence $\mathbf{x}$ of letters is basic in Shannon's information theory. Extending the classical problems to networks of information sources and noisy channels one is led in a natural way to the concept of a decoding set $\mathscr{B}$ corresponding to a set $\mathscr{A}$ of sequences of letters. Based on this tool the aim of our paper is to develop a technique for proving strong converses of coding theorems. The main result is Theorem 1. The results are applied to a source coding problem with side information and to the degraded broadcast channel.

This research is restricted to memoryless stationary sources and channels. All the random variables (r.v.) have finite range. Unless it is stated otherwise, exp's and log's are to the base 2. "ln" stands for the natural logarithm, $h(\varepsilon)$ denotes the entropy of the binary distribution $(\varepsilon, 1-\varepsilon)$. $\|Z\|$ denotes the cardinality of the range of the r.v. Z, $\|\mathscr{A}\|$ is the cardinality of the set $\mathscr{A}$. Throughout the paper the word measure stands for probability measures.

---

* Present address: Mathematisches Institut der Universität Bielefeld, K. Schumacher-Str. 6, D-4800 Bielefeld, Federal Republic of Germany.

Research of this author was supported by the National Science Foundation under Grant no GK-40492 and by the Deutsche Forschungsgemeinschaft.

We are given the finite sets $\mathscr{X}$, $\mathscr{Y}$ and the transition probabilities $W(y|x)$ for $x \in \mathscr{X}$, $y \in \mathscr{Y}$. For the n-th cartesian power of $\mathscr{X}$ and $\mathscr{Y}$ we define

$$W^n(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} W(y_i|x_i)$$

where $\mathbf{x} = x_1 x_2 \dots x_n$ and $\mathbf{y} = y_1 y_2 \dots y_n$.

Definition 1. The set $\mathscr{B} \subset \mathscr{Y}^n$ ε-decodes the sequence $\mathbf{x} \in \mathscr{X}^n$ if

$W^n(\mathscr{B}|\mathbf{x}) \geqq 1 - \varepsilon$.

We put $\Psi_\varepsilon(\mathscr{B}) \subset \mathscr{X}^n$ for the set of all the $\mathbf{x}$'s which are ε-decoded by $\mathscr{B}$.

We shall say that $\mathscr{A}$ is ε-decoded by $\mathscr{B}$ if $\mathscr{A} \subset \Psi_\varepsilon(\mathscr{B})$.

We are interested in the minimum "size" of a $\mathscr{B}$ which satisfies a prescribed lower bound on the "size" of $\Psi_\varepsilon(\mathscr{B})$. We measure the "size" of sets by probability measures of the product type.

Let us denote by Q a measure given on X and by R a measure on $\mathscr{Y}$. Q^n and R^n are the corresponding product measures on $\mathscr{X}^n$ and $\mathscr{Y}^n$. We suppose that Q and R never vanish.

Put

$$S_n(c, \varepsilon) = \frac{1}{n} \cdot \log \min_{\frac{1}{n} \log Q^n(\Psi_\varepsilon(\mathscr{B})) \geqq c} R^n(\mathscr{B}).$$

(Note that c and $S_n(c, \varepsilon)$ are non-positive quantities.)

We shall show that the limit of $S_n(c, \varepsilon)$ is independent of ε for any fixed value of c and give a computable formula for this limit.

To express this we have to introduce the concept of the relative entropy of a random variable Z having distribution P relative to an underlying measure Q. (See Kullback [6]. However, he uses a slightly different terminology.)

Definition 2. Given the r.v. Z with values in a finite set $\mathscr{Z}$, distribution P and measure Q on $\mathscr{Z}$, we define the relative entropy of Z as

$$H_Q(Z) \triangleq \sum_{z \in \mathscr{Z}} P(z) \cdot \log \frac{Q(z)}{P(z)}.$$

Given the r.v.'s U and Z with distribution P and values in the sets $\mathscr{U}$ and $\mathscr{Z}$ and the measure Q on $\mathscr{U} \times \mathscr{Z}$, the relative conditional entropy of Z given U is

$$H_Q(Z|U) \triangleq H_Q(Z, U) - H_Q(U) = \sum_{u \in \mathscr{U}} P(u) \cdot \sum_{z \in \mathscr{Z}} P(z|u) \cdot \log \frac{Q(z|u)}{P(z|u)}.$$

Remark that if $Q(z|u)$ does not depend on u, $H_Q(Z|U)$ depends only on the $\mathscr{Z}$-marginal of the measure Q. We are only interested in such situations and will simply define even for any distribution R on $\mathscr{Z}$:

$$H_R(Z|U) \triangleq \sum_{u \in \mathscr{U}} P(u) \cdot \sum_{z \in \mathscr{Z}} P(z|u) \cdot \log \frac{R(z)}{P(z|u)}.$$

Definition 3. Consider the sets $\mathscr{X}$, $\mathscr{Y}$ and a countable set $\mathscr{U}$. Let $\mathscr{P}(W)$ be the set of all the r.v.'s (U, X, Y) with values on $\mathscr{U} \times \mathscr{X} \times \mathscr{Y}$ such that U, X, Y form a

Markov chain, and $P(Y=y|X=x)=W(y|x)$. We define

$$T(c) \triangleq \inf_{\substack{H_Q(X|U)\geqq c \\ (U,X,Y)\in\mathscr{P}(W) \\ \|U\|<\infty}} H_R(Y|U).$$

We write $(X,Y)\in\mathscr{P}(W)$ if (X,Y) satisfy $P(Y=y|X=x)=W(y|x)$. (This is justified by the fact that in this case for any r.v. U taking a single value: $(U,X,Y)\in\mathscr{P}(W)$.)

Hence

$$T(c)\leqq \min_{\substack{H_Q(X)\geqq c \\ (X,Y)\in\mathscr{P}(W)}} H_R(Y).$$

We shall prove the following

Lemma 1A. *We can suppose that* $\|\mathscr{U}\|\leqq 3$, *that is*

$$T(c)= \min_{\substack{H_Q(X|U)\geqq c \\ (U,X,Y)\in\mathscr{P}(W),\ \|U\|\leqq 3}} H_R(Y|U).$$

Theorem 1.

$$\lim_{n\to\infty} S_n(c,\varepsilon)=T(c).$$

Remark that by this theorem $S_n(c,\varepsilon)$ is asymptotically independent of ε.

Though Theorem 1 is of no immediate use for the coding problems treated in later sections, it enlightens our topic from a probabilistic viewpoint. Our immediate purposes are served by a modified version of Theorem 1 where the sets underlying the minimization will be restricted to consist of "typical sequences".

Definition 4. For a sequence r_n of positive reals with $r_n\cdot n^{-1/2}\to\infty$, and $r_n\cdot n^{-1}\to 0$ $\mathbf{x}\in\mathscr{X}^n$ is a $(Q,\{r_n\})$-typical source sequence, if for every $x\in\mathscr{X}$

$$|\|\{i;x_i=x\}\|-nQ(x)|<r_n.$$

We denote by $\mathscr{T}_n(Q)$ the set of all the typical sequences of $\mathscr{X}^n$.

It is well-known that $Q^n(\mathscr{T}_n(Q))\to 1$.

Put

$$\hat{S}_n(c,\varepsilon)\triangleq\frac{1}{n}\cdot\log_{\frac{1}{n}\log Q^n(\Psi_\varepsilon(\mathscr{B})\cap\mathscr{T}_n(Q))\geqq c}\min R^n(\mathscr{B}).$$

We shall prove that the limit of $\hat{S}_n(c,\varepsilon)$ is independent of both, ε and $\{r_n\}$. Define

$$\hat{T}(c)\triangleq \inf_{\substack{H_Q(X|U)\geqq c \\ (U,X,Y)\in\mathscr{P}(W,Q) \\ \|U\|<\infty}} H_R(Y|U)$$

where $\mathscr{P}(W,Q)$ consists of those triples $(U,X,Y)\in\mathscr{P}(W)$ where the distribution of X is Q. Similarly to Lemma 1A we shall show that

Lemma 1B. *We can suppose* $\|\mathscr{U}\|\leqq\|\mathscr{X}\|+2$ *and still have*

$$\hat{T}(c)= \min_{\substack{H_Q(X|U)\geqq c \\ (U,X,Y)\in\mathscr{P}(W,Q)}} H_R(Y|U).$$

After this we prove that

Theorem 2.

$$\lim_{n\to\infty} \hat{S}_n(c, \varepsilon) = \hat{T}(c).$$

Two problems involving communication networks will be treated below, one for source-coding and one for channel-coding. For the source-coding problem see [1], where a coding theorem and weak converse result is proved. The corresponding results for the channel coding problem are to be found in [1–4] and [9]. These results are "weak" converses in Wolfowitz' sense [10], meaning that they give precise asymptotic bounds on the exponent of the size of the respective coding functions for the case when the probabilities of decoding errors are tending to 0. A strong converse theorem states that allowing large probabilities for erroneous decoding does not effect the asymptotic bounds. In this paper we give strong converses for the above problems by a method which seems to apply to many coding problems.

It is the same technique which allows us to prove that the limit in Theorems 1 and 2 is independent of ε. The method is based on a combinatorial lemma of Margulis [8] which consists in a lower bound on the size of the Hamming *1*-neighbourhood of a set of binary sequences. The proof of a slightly generalized form of this lemma will be postponed to the last section of the present paper.

Let us formulate the coding problems.

Source Coding with Side Information

A sequence $\{(X_i, Y_i)\}_{i=1}^{\infty}$ of independent and identically distributed pairs of r.v.'s is called a discrete memoryless correlated stationary information source (DMCSS). Two independent encoders observe $X^n = X_1 X_2 \dots X_n$ and Y^n and produce the functions $f_n(X^n)$ and $g_n(Y^n)$. These are the codes. A decoder having access to both $f_n(X^n)$ and $g_n(Y^n)$ has to construct a function of the two with the property

$$\Pr(V_n(f_n(X^n), g_n(Y^n)) = Y^n) \geqq 1 - \varepsilon. \tag{1}$$

Thus the decoder reproduces only the Y^n-sequence.

A pair (R_1, R_2) of non-negative reals is called an *ε-achievable rate pair* if for any $\delta > 0$ and sufficiently large n there exist functions f_n, g_n and V_n satisfying (1) and the inequalities

$$\| f_n(X^n) \| \leqq \exp\{(R_1 + \delta) \cdot n\}; \qquad \| g_n(Y^n) \| \leqq \exp\{(R_2 + \delta) \cdot n\}. \tag{2}$$

A rate pair is *achievable* if it is ε-achievable for every $0 < \varepsilon \leqq 1$.

Let us denote by $\mathscr{R}(\varepsilon)$ the ensemble of all the ε-achievable rates, and by $\mathscr{R}$ that of all the achievable rates. Clearly $\mathscr{R} = \bigcap_{\varepsilon > 0} \mathscr{R}(\varepsilon)$.

In [1] the following theorem was proved:

$$\mathscr{R} = \{(R_1, R_2);\ R_1 \geqq I(X \wedge U),\ R_2 \geqq H(Y \mid U),\ \| U \| \leqq \| X \| + 2,$$
$$U, X, Y \text{ Markov chain}\}. \tag{3}$$

In this paper we prove the strong converse to this theorem, i.e.

Theorem 3.

$\mathscr{R}(\varepsilon)=\mathscr{R}$ for $0<\varepsilon\leqq 1$.

Degraded Broadcast Channel (DBC)

Broadcast channels were first considered by Cover [3]. His paper created immediate interest, because new information-theoretic techniques were needed in order to find characterizations of the capacity region. Those characterizations still do not exist for the general case, however, in an important special case, the broadcast channel with degraded components described below, the problem is completely solved. Those later channels were studied by Bergmans [2], who also described a coding scheme which he conjectured to be optimal. The conjecture was proved to be true by Wyner [9] in the special case of binary symmetric broadcast channels. His proof uses very special properties of binary symmetric channels and does not allow for extension to the general degraded case. Then Gallager [4] proved a coding theorem and weak converse for arbitrary degraded broadcast channels. However, he gives a slightly weaker characterization of the capacity region than the one conjectured by Bergmans. Finally this conjecture was also proved to be true in [1]. The result is stated in (6) and (7) below. We give now the necessary definitions.

Let us be given finite sets $\mathscr{X}, \mathscr{Y}, \mathscr{Z}$ and the stochastic matrices

$$\{W_1(y|x);\ x\in\mathscr{X},\ y\in\mathscr{Y}\},\qquad \{W_3(z|y);\ y\in\mathscr{Y},\ z\in\mathscr{Z}\}.$$

Put

$$W_2(z|x)\triangleq\sum_{y\in\mathscr{Y}} W_3(z|y)\cdot W_1(y|x),$$

and for each of the channels W_i denote by W_i^n its product extension to the corresponding sets $\mathscr{X}^n, \mathscr{Y}^n, \mathscr{Z}^n$.

Let M_1, M_2 and n be natural numbers. A set of triples $\{\mathbf{x}_{ij}, \mathscr{A}_i, \mathscr{B}_j,\ 1\leqq i\leqq M_1,\ 1\leqq j\leqq M_2\}$ is a code for the DBC if $\mathbf{x}_{ij}\in\mathscr{X}^n$, the $\mathscr{A}_i$'s are disjoint subsets of $\mathscr{Y}^n$ and the $\mathscr{B}_j$'s disjoint subsets of $\mathscr{Z}^n$. An error occurs if either a sequence $\mathbf{y}\notin\mathscr{A}_i$ or $\mathbf{z}\notin\mathscr{B}_j$ was received provided that the codeword $\mathbf{x}_{ij}$ had been sent. Thus the error probability of the code is the pair of reals $(\varepsilon_1, \varepsilon_2)$ where

$$\varepsilon_1=\max_{1\leqq j\leqq M_2}\ \max_{1\leqq i\leqq M_1} W_1^n(\bar{\mathscr{A}}_i|\mathbf{x}_{ij}) \tag{4}$$

and

$$\varepsilon_2=\max_{1\leqq j\leqq M_2}\ \max_{1\leqq i\leqq M_1} W_2^n(\bar{\mathscr{B}}_j|\mathbf{x}_{ij}).$$

(This is the so-called *maximal error*.)

We shall say that $\{(\mathbf{x}_{ij}, \mathscr{A}_i, \mathscr{B}_j)\}$ is an $(n, \varepsilon_1, \varepsilon_2)$-code if (4) holds.

A pair (R_1, R_2) of non-negative reals is called $(\varepsilon_1, \varepsilon_2)$-*achievable rate* for the DBC $\{W_1, W_3\}$ if for any $\delta>0$ and sufficiently large n there exists a code $\{(\mathbf{x}_{ij},$

$\mathscr{A}_i, \mathscr{B}_j)\}$ such that

a) $M_k \geqq \exp[n(R_k - \delta)]$ $k=1,2$

b) $\max_{1 \leqq j \leqq M_2} \max_{1 \leqq i \leqq M_1} W_1^n(\bar{\mathscr{A}}_i | \mathbf{x}_{ij}) \leqq \varepsilon_1$

c) $\max_{1 \leqq j \leqq M_2} \max_{1 \leqq i \leqq M_1} W_2^n(\bar{\mathscr{B}}_j | \mathbf{x}_{ij}) \leqq \varepsilon_2$. (5)

A rate pair is achievable if it is $(\varepsilon_1, \varepsilon_2)$-achievable for $0 < \varepsilon_k \leqq 1$; $k=1,2$. Denote the region of achievable rates by $\mathscr{C}$, and that of the $(\varepsilon_1, \varepsilon_2)$-achievable rates by $\mathscr{C}(\varepsilon_1, \varepsilon_2)$. Clearly, $\mathscr{C} = \bigcap_{\substack{0<\varepsilon_k \\ k=1,2}} \mathscr{C}(\varepsilon_1, \varepsilon_2)$.

In [1] it is proved that (R_1, R_2) is achievable iff there exist r.v.'s U, X, Y, Z forming a Markov chain in this order with given conditional probabilities

$$\Pr(Y=y|X=x) = W_1(y|x), \qquad \Pr(Z=z|Y=y) = W_2(z|y), \tag{6}$$

satisfying $\|U\| \leqq \min\{\|X\|, \|Y\|, \|Z\|\}$ and such that

$$R_1 \leqq I(X \wedge Y|U); \qquad R_2 \leqq I(U \wedge Z). \tag{7}$$

Here again we prove the corresponding strong converse. This will be

Theorem 4. *If (R_1, R_2) is $(\varepsilon_1, \varepsilon_2)$-achievable for a fixed pair $0 < \varepsilon_k \leqq 1$; $k=1,2$; then it is achievable, i.e.*

$$\mathscr{C} = \mathscr{C}(\varepsilon_1, \varepsilon_2) \quad \textit{for any } 0 < \varepsilon_k \leqq 1; \quad k=1,2. \tag{8}$$

2. Proof of Theorem 1. Weak Version

In this Section we shall prove Theorem 1 for "small" ε's. We recall the following well-known property of relative entropies:

Fact 1 ([6]). Given a finite set $\mathscr{X}$, the product measure Q^n on $\mathscr{X}^n$, a sequence Z_n of i.i.d.r.v.'s with values in $\mathscr{X}$ and distribution P, and any sequence δ_n bounded away from 1 and satisfying $n^{-1} \cdot \log \delta_n \to 0$ we have

$$\inf_{\mathscr{C}: \Pr(Z^n \in \mathscr{C}) \geqq 1-\delta_n} \frac{1}{n} \cdot \log Q^n(\mathscr{C}) \to H_Q(Z)$$

A) Consider any triple of r.v.'s $(U, X, Y) \in \mathscr{P}(W)$. We shall construct a sequence $\{\mathscr{B}_n\}$ of subsets of $\mathscr{Y}^n$ and a sequence $\varepsilon_n \to 0$ such that

$$n^{-1} \cdot \log R^n(\mathscr{B}_n) \to H_R(Y|U)$$
$$\liminf_{n \to \infty} n^{-1} \cdot \log Q^n(\Psi_{\varepsilon_n}(\mathscr{B}_n)) \geqq H_Q(X|U). \tag{9}$$

We shall first show that for r.v.'s $(X, Y) \in \mathscr{P}(W)$ we can construct a sequence $\{\mathscr{B}_n\}$ of subsets of $\mathscr{Y}^n$ and a sequence $\varepsilon_n \to 0$ with

$$n^{-1} \cdot \log R^n(\mathscr{B}_n) \to H_R(Y); \qquad \liminf_{n \to \infty} n^{-1} \cdot \log Q^n(\Psi_{\varepsilon_n}(\mathscr{B}_n)) \geqq H_Q(X). \tag{10}$$

By Fact 1 for $\delta_n = n^{-1}$ there exists a sequence of sets $\mathscr{B}_n \subset \mathscr{Y}^n$ such that

$$n^{-1} \cdot \log R^n(\mathscr{B}_n) \to H_R(Y); \qquad \Pr(Y^n \in \mathscr{B}_n) > 1 - n^{-1}. \tag{11}$$

Since $\Pr(Y^n \in \mathscr{B}_n) = \sum_{\mathbf{x} \in \mathscr{X}^n} \Pr(X^n = \mathbf{x}) \cdot W^n(\mathscr{B}_n | \mathbf{x})$, we conclude from (11) by a "reverse" Markov inequality (see [7]) that

$$\Pr(X^n \in \Psi_{n^{-1/2}}(\mathscr{B}_n)) > 1 - n^{-1/2}.$$

Hence by Fact 1

$$\liminf_{n \to \infty} \frac{1}{n} \cdot \log Q^n(\Psi_{n^{-1/2}}(\mathscr{B}_n)) \geqq H_Q(X).$$

Putting now $\varepsilon_n = n^{-1/2}$ the last inequality and (11) establish (10).

Considering the given U, X, Y write

$$H_R(Y|U) = \sum_{u \in \mathscr{U}} \Pr(U = u) \cdot H_R(Y | U = u),$$

$$H_Q(X|U) = \sum_{u \in \mathscr{U}} \Pr(U = u) \cdot H_Q(X | U = u).$$

For any integer n and pr.d. P on $\mathscr{U}$ there exist integers $J_n(u)$ such that

$$\sum_{u \in \mathscr{U}} J_n(u) = n; \qquad |J_n(u) - P(u) \cdot n| < 1. \tag{12}$$

Clearly $J_n(u) \to \infty$ for every $u \in \mathscr{U}$.

Applying (10) to a pair of r.v.'s (X_u, Y_u) having joint pr.d. $\Pr(X_u = x, Y_u = y) \triangleq \Pr(X = x, Y = y | U = u)$, we construct a sequence $\mathscr{B}_n(u)$ of subsets of $\mathscr{Y}^{J_n(u)}$ with

$$[J_n(u)]^{-1} \cdot \log R^{J_n(u)}(\mathscr{B}_n(u)) \to H_R(Y | U = u)$$

$$\liminf_{n \to \infty} [J_n(u)]^{-1} \cdot \log Q^{J_n(u)}(\Psi_{[J_n(u)]^{-1/2}}(\mathscr{B}_n(u)) \geqq H_Q(X | U = u). \tag{13}$$

For any fixed n we consider

$$\mathscr{B}_n \triangleq \prod_{u \in \mathscr{U}} \mathscr{B}_n(u) \subset \mathscr{Y}^n$$

the cartesian product of the $\mathscr{B}_n(u)$'s.

(12) and (13) imply that this set satisfies (9) for

$$\varepsilon_n = 1 - \prod_{u \in \mathscr{U}} (1 - [J_n(u)]^{-1/2}).$$

B) The proof of inequality

$$\liminf_{\substack{n \to \infty \\ \varepsilon_n \to 0}} S_n(c, \varepsilon_n) \geqq T(c) \tag{14}$$

goes by several lemmas.

The Proof of Lemma 1A is the very same as that of Lemma 3 in [1]. Denote by $\mathbf{D}(\mathscr{X})$ the set of all pr.d.'s on $\mathscr{X}$. For $\mathbf{p}\in\mathbf{D}(\mathscr{X})$ we consider the functions

$$\varphi_0(\mathbf{p})\triangleq\sum_{\mathbf{x}\in X} p(x)\cdot\log\frac{Q(x)}{p(x)}$$

and

$$\varphi_1(\mathbf{p})\triangleq\sum_{y\in\mathscr{Y}}\Big(\sum_{x\in\mathscr{X}} p(x)\cdot W(y|x)\Big)\cdot\log\frac{R(y)}{\sum\limits_{x\in\mathscr{X}} p(x)\cdot W(y|x)}.$$

We denote the conditional distribution of $\tilde{X}$ on $\{\tilde{U}=u\}$ by $\mathbf{p}_u$.

Notice that the conditional entropies are the convex linear combinations

$$H_Q(X|U)=\sum_{u\in\mathscr{U}}\Pr(U=u)\cdot\varphi_0(\mathbf{p}_u)$$

$$H_R(Y|U)=\sum_{u\in\mathscr{U}}\Pr(U=u)\cdot\varphi_1(\mathbf{p}_u). \tag{15}$$

Hence the vector $(H_Q(X|U), H_R(Y|U))$ is an element of $\mathscr{C}$, the convex hull of the image of $\mathbf{D}(\mathscr{X})$ under (φ_0,φ_1). Since $\mathbf{D}(\mathscr{X})$ is compact, and the functions φ_0 and φ_1 are continuous, $\mathscr{C}$ is a compact subset of $\mathbb{E}^2$. Thus by Carathéodory's theorem every element of $\mathscr{C}$ is a convex linear combination of at most 3 extremal points. Clearly, the extremal points are contained in the image of $\mathbf{D}(\mathscr{X})$. Hence there exist elements $\mathbf{p}_i$ of $\mathbf{D}(\mathscr{X})$ and nonnegative reals α_i $(1\leqq i\leqq 3)$ summing up to 1 with

$$H_Q(X|U)=\sum_{i=1}^{3}\alpha_i\,\varphi_0(\mathbf{p}_i),$$

$$H_R(Y|U)=\sum_{i=1}^{3}\alpha_i\,\varphi_1(\mathbf{p}_i).$$

Choosing a $(U,X,Y)\in\mathscr{P}(W)$ with

$$\Pr(U=i)=\alpha_i,\qquad \Pr(X=x|U=i)=\mathbf{p}_i(x)$$

we get the statement of the Lemma.

Lemma 2. *$T(c)$ is convex $(\cup)$ and monotonically increasing in c.*

Proof. Let us be given the triples $(U_i,X_i,Y_i)\in\mathscr{P}(W)$ for $i=1,2$. We introduce a new r.v. T ranging over the set $\{1,2\}$ and independent of the U_i's, X_i's and Y_i's.

$$(H_Q(X|U_T,T), H_R(Y|U_T,T))=\sum_{i=1,2}\Pr(T=i)\cdot(H_Q(X_i|U_i), H_R(Y_i|U_i)).$$

Varying the distribution of T we thus get every point of the segment of the straight line connecting the points $(H_Q(X_i|U_i), H_R(Y_i|U_i))$; $i=1,2$. Hence the convexity of $\mathscr{R}$ follows because $((T,U_T),X_T,Y_T)\in\mathscr{P}(W)$.

The monotonicity is obvious.

Lemma 3. *Let us consider arbitrary sets $\mathscr{X}$ and $\mathscr{Y}$ satisfying $\mathscr{X}=\mathscr{X}_1\times\mathscr{X}_2$, $\mathscr{Y}=\mathscr{Y}_1\times\mathscr{Y}_2$. Set $W(y_1,y_2|x_1,x_2)=\prod\limits_{i=1,2}W_i(y_i|x_i)$. We suppose further that the*

measures Q on $\mathscr{X}$ and R on $\mathscr{Y}$ are the products of the $Q^{(i)}$'s on the $\mathscr{X}_i$'s, and of the $R^{(i)}$'s on the sets $\mathscr{Y}_i$ for $i=1,2$. For W, Q, R and the $(W_i, Q^{(i)}, R^{(i)})$'s we define the functions $T(c)$ and $T_i(c)$ $(i=1,2)$ as in Definition 3. Then the following identity holds

$$T(c)=\inf_{c_1+c_2\geqq c}[T_1(c_1)+T_2(c_2)].$$

Proof. It is easy to see that $T(c_1+c_2)\leqq T_1(c_1)+T_2(c_2)$. Actually, consider two triples $(U_i, X_i, Y_i)\in\mathscr{P}(W_i)$. We choose a (U, X, Y) such that its distribution is the product of those of the (U_i, X_i, Y_i)'s for $i=1,2$. This triple will establish the statement, since relative entropies are additive for independent r.v.'s.

Now we prove that for any c there exist c_1, c_2 with $T(c)=T_1(c_1)+T_2(c_2)$; $c_1+c_2=c$. We write

$$\begin{aligned} H_R(Y_1Y_2|U)&=H_{R^{(1)}}(Y_1|U)+H_{R^{(2)}}(Y_2|UY_1)\\ &\geqq H_{R^{(1)}}(Y_1|U)+H_{R^{(2)}}(Y_2|UY_1X_1) \end{aligned} \tag{16}$$

where the last inequality follows from the identity

$$H_{R^{(2)}}(Y_2|UY_1)-H_{R^{(2)}}(Y_2|UY_1X_1)=I(X_1\wedge Y_2|UY_1)$$

by the non-negativity of conditional mutual information. By the same identity,

$$H_{R^{(2)}}(Y_2|UY_1X_1)=H_{R^{(2)}}(Y_2|UX_1)-I(Y_2\wedge Y_1|UX_1). \tag{17}$$

Since Y_1 is independent of the remaining variables given the value of X_1, the conditional mutual information in (17) is 0. From (16) and (17) we thus get that

$$H_R(Y_1Y_2|U)\geqq H_{R^{(1)}}(Y_1|U)+H_{R^{(2)}}(Y_2|UX_1).$$

Since $(U, X_1, Y_1)\in\mathscr{P}(W_1)$ and $((U, X_1), X_2, Y_2)\in\mathscr{P}(W_2)$, we conclude that

$$H_R(Y_1Y_2|U)\geqq T_1(H_{Q^{(1)}}(X_1|U))+T_2(H_{Q^{(2)}}(X_2|UX_1)). \tag{18}$$

For the given c consider any $\varepsilon>0$ and a triple $(U, (X_1, X_2), (Y_1, Y_2))$ achieving

$$H_Q(X_1X_2|U)\geqq c;\qquad H_R(Y_1Y_2|U)\leqq T(c)+\varepsilon.$$

Applying (18) to this triple we get

$$T(c)\geqq T_1(H_{Q^{(1)}}(X_1|U))+T_2(H_{Q^{(2)}}(X_2|UX_1))-\varepsilon.$$

Our statement follows now because

$$H_{Q^{(1)}}(X_1|U)+H_{Q^{(2)}}(X_2|UX_1)=H_Q(X_1X_2|U)\geqq c.$$

We extend now the function $T(c)$ to product spaces.

Definition. For the given sets $\mathscr{U}$, $\mathscr{X}$, $\mathscr{Y}$, measures Q, R and transition matrix W put

$$T_n(c)\triangleq\inf_{\substack{\frac{1}{n}H_{Q^n}(X^n|U)\geqq c\\ (U,X^n,Y^n)\in\mathscr{P}(W^n)}}\frac{1}{n}H_{R^n}(Y^n|U)$$

Corollary.

$$T_n(c)=T_1(c)$$

Proof. The inequality

$$T_n(c) \leqq T_1(c)$$

is a trivial consequence of Lemma 3, also the equality

$$T_n(c) = \inf_{\frac{1}{n}\sum_i c^{(i)} \geqq c} \frac{1}{n} \cdot \sum_{i=1}^{n} T_1(c^{(i)})$$

The convexity and the monotonicity of $T_1(c) = T(c)$, as expressed in Lemma 2, yield

$$\frac{1}{n} \sum_{i=1}^{n} T_1(c^{(i)}) \geqq T_1\left(\frac{1}{n} \sum_{i=1}^{n} c^{(i)}\right) \geqq T_1(c)$$

and therefore $T_n(c) \geqq T_1(c)$.

We go over to the proof of inequality (14).

Let us be given a set $\mathscr{B} \subset \mathscr{Y}^n$. Put $\mathscr{A} = \Psi_\varepsilon(\mathscr{B})$. If $\mathscr{A}$ is not the empty set, we shall construct a r.v. X^n, with distribution concentrated on $\mathscr{A}$ and give an estimate of the probabilities of $\mathscr{A}$ and $\mathscr{B}$ through relative entropies.

We define

$$\Pr(X^n = \mathbf{x}) \triangleq \begin{cases} \dfrac{Q^n(\mathbf{x})}{Q^n(\mathscr{A})} & \text{if } \mathbf{x} \in \mathscr{A} \\ 0 & \text{otherwise} \end{cases}$$

and an Y^n with $(X^n, Y^n) \in \mathscr{P}(W^n)$.

Then

$$H_{Q^n}(X^n) = \log Q^n(\mathscr{A}).$$

Let $\chi_{\mathscr{B}}$ denote the characteristic function of the set $\mathscr{B}$. Clearly,

$$\begin{aligned} H_{R^n}(Y^n) &= H_{R^n}(\chi_{\mathscr{B}}(Y^n)) + H_{R^n}(Y^n | \chi_{\mathscr{B}}(Y^n)) \\ &\leqq H_{R^n}(\chi_{\mathscr{B}}(Y^n)) = \Pr(Y^n \in \mathscr{B}) \cdot \log \frac{R^n(\mathscr{B})}{\Pr(Y^n \in \mathscr{B})} \\ &\quad + \Pr(Y^n \in \bar{\mathscr{B}}) \cdot \log \frac{R^n(\bar{\mathscr{B}})}{\Pr(Y^n \in \bar{\mathscr{B}})} \\ &= H(\chi_{\mathscr{B}}(Y^n)) + \Pr(Y^n \in \mathscr{B}) \cdot \log R^n(\mathscr{B}) + \Pr(Y^n \in \bar{\mathscr{B}}) \cdot \log R^n(\bar{\mathscr{B}}) \\ &\leqq 1 + \Pr(Y^n \in \mathscr{B}) \cdot \log R^n(\mathscr{B}). \end{aligned} \tag{19}$$

Notice that

$$\Pr(Y^n \in \mathscr{B}) = \sum_{\mathbf{x} \in \mathscr{A}} \Pr(X^n = \mathbf{x}) \cdot W^n(\mathscr{B} | \mathbf{x}) \geqq 1 - \varepsilon. \tag{20}$$

Comparing the last inequality with (19) we get

$$\frac{1}{n}\cdot\log R^n(\mathscr{B})\geqq(1-\varepsilon)^{-1}\cdot\left[\frac{1}{n}\cdot H_{R^n}(Y^n)-\frac{1}{n}\right]$$

$$\geqq(1-\varepsilon)^{-1}\cdot\left[T\left(\frac{1}{n}H_{Q^n}(X^n)\right)-\frac{1}{n}\right]$$

and substituting $\frac{1}{n}H_{Q^n}(X^n)=\frac{1}{n}\cdot\log Q^n(\mathscr{A})$ this becomes

$$\frac{1}{n}\cdot\log R^n(\mathscr{B})\geqq(1-\varepsilon)^{-1}\cdot\left[T\left(\frac{1}{n}\cdot\log Q^n(\Psi_\varepsilon(\mathscr{B}))\right)-\frac{1}{n}\right].$$

By the definition of $S_n(c,\varepsilon)$ this means that

$$S_n(c,\varepsilon)\geqq(1-\varepsilon)^{-1}\cdot\left[T(c)-\frac{1}{n}\right]. \tag{21}$$

This establishes the relation (14).

3. The Strong Version of Theorem 1: Blowing up a Decoding Set

Let us introduce in the set $\mathscr{Y}^n$ the Hamming-distance

$$d(\mathbf{y}',\mathbf{y}'')=\|\{i\colon 1\leqq i\leqq n,\ y_i'\neq y_i''\}\|.$$

We define the k-Hamming-neighbourhood $\Gamma^k\mathscr{B}$ of a set $\mathscr{B}\subset\mathscr{Y}^n$ as

$$\Gamma^k\mathscr{B}\triangleq\{\mathbf{y};\mathbf{y}\in\mathscr{Y}^n,\ \exists\,\mathbf{y}'\in\mathscr{B}\colon d(\mathbf{y},\mathbf{y}')\leqq k\}.$$

Notice that $\Gamma^1\Gamma^k\mathscr{B}=\Gamma^{k+1}\mathscr{B}$.

We write Γ instead of Γ^1.

$$\partial\mathscr{B}\triangleq\mathscr{B}\cap\Gamma\bar{\mathscr{B}}.$$

We put

$$\varphi(t)=(2\pi)^{-1/2}\cdot e^{-\frac{t^2}{2}};\qquad \Phi(t)=\int_{-\infty}^{t}\varphi(u)\,du\quad\text{and}\quad f(s)\triangleq\varphi(\Phi^{-1}(s)) \tag{22}$$

where Φ^{-1} is the inverse function of Φ. By Margulis's theorem (see our Theorem 5 in Section 6) for any set $\mathscr{B}\subset\mathscr{Y}^n$ and $\mathbf{x}\in\mathscr{X}^n$

$$W^n(\partial\mathscr{B}|\mathbf{x})\geqq a\cdot n^{-1/2}\cdot f(W^n(\mathscr{B}|\mathbf{x})),$$

where the constant a depends only on W.

As an application, we obtain

Lemma 4. *Given the sets $\mathscr{X}^n$, $\mathscr{Y}^n$, the transition probability matrix W^n from $\mathscr{X}^n$ to $\mathscr{Y}^n$, there is a constant a depending only on W such that for any $\mathscr{B}\subset\mathscr{Y}^n$ and*

$\mathbf{x}\in\mathscr{X}^n$

$$W^n(\Gamma^k\mathscr{B}|\mathbf{x})\geqq\Phi[\Phi^{-1}(W^n(\mathscr{B}|\mathbf{x}))+n^{-1/2}\cdot(k-1)\cdot a].$$

Proof. Estimating $W^n(\Gamma^k\mathscr{B}|\mathbf{x})$ we shall use the relations

$$\Gamma\mathscr{B}-\mathscr{B}\supset\partial(\Gamma\mathscr{B}),\qquad \Gamma\mathscr{B}-\mathscr{B}=\partial\bar{\mathscr{B}}. \tag{23}$$

Let us denote for a moment

$$t_k\triangleq\Phi^{-1}(W^n(\Gamma^k\mathscr{B}|\mathbf{x})).$$

By Margulis' theorem, and (23)

$$\Phi(t_{k+1})-\Phi(t_k)\geqq n^{-1/2}\cdot a\cdot\max\{\varphi(t_k),\varphi(t_{k+1})\}.$$

Now, φ is monotone on both $(-\infty,0)$ and $(0,\infty)$. So, unless $t_k<0<t_{k+1}$,

$$\max_{t_k\leqq u\leqq t_{k+1}}\varphi(u)=\max\{\varphi(t_k),\varphi(t_{k+1})\},$$

and hence by Lagrange's theorem $t_{k+1}-t_k\geqq n^{-1/2}\cdot a$. Q.e.d.

For the applications of Lemma 4 we note that as $t\to-\infty$,

$$\Phi(t)=1-\Phi(-t)\sim\frac{1}{|t|}\cdot\varphi(t) \tag{24}$$

(see [11]).

Hence it follows easily that as $s\to 0$,

$$-\Phi^{-1}(s)=\Phi^{-1}(1-s)\sim\sqrt{-2\cdot\log s}. \tag{25}$$

Let us prove one more – rather trivial –

Lemma 5. *Given a set $\mathscr{Y}$, a measure Q on $\mathscr{Y}$ which never vanishes and a sequence k_n of positive integers with $n^{-1}\cdot k_n\to 0$, then*

$$\lim_{n\to\infty}\frac{1}{n}\cdot\log\sup_{\mathbf{y}\in\mathscr{Y}^n}[Q^n(\Gamma^{k_n}\{\mathbf{y}\})\cdot[Q^n(\mathbf{y})]^{-1}]=0.$$

Proof. Let us denote by m_Q the minimum of Q on $\mathscr{Y}$. For any $\mathbf{y}'\in\Gamma^k\{\mathbf{y}\}$,

$$Q^n(\mathbf{y}')\leqq Q^n(\mathbf{y})\cdot(m_Q)^{-k_n}.$$

Hence

$$Q^n(\Gamma^{k_n}\{\mathbf{y}\})\cdot[Q^n(\mathbf{y})]^{-1}\leqq(m_Q)^{-k_n}\cdot\sum_{i=0}^{k_n}\left[\binom{n}{i}\cdot\|\mathscr{Y}\|^i\right]$$

$$\leqq(m_Q)^{-k_n}\cdot(k_n+1)\cdot\binom{n}{k_n}\cdot\|\mathscr{Y}\|^{k_n}$$

because $k_n\leqq\frac{n}{2}$. The rest is trivial by Stirling's formula.

Now we turn to the strong version of Theorem 1.

Choose any sequence of integers $\{k_n\}_{n=1}^{\infty}$ with

$$k_n\cdot n^{-1}\to 0,\qquad k_n\cdot n^{-1/2}\to\infty. \tag{26}$$

For an arbitrary ε put

$$\varepsilon_n = 1 - \Phi[\Phi^{-1}(1-\varepsilon) + n^{-1/2} \cdot a \cdot (k_n - 1)].$$

Then $\varepsilon_n \to 0$ as $n \to \infty$, and any set $\mathscr{B} \subset \mathscr{Y}^n$ satisfies the relation

$$\Psi_\varepsilon(\mathscr{B}) \subset \Psi_{\varepsilon_n}(\Gamma^{k_n}\mathscr{B}). \tag{27}$$

On the other hand $\Gamma^{k_n}\mathscr{B}$ is "not much larger" than $\mathscr{B}$. In fact,

$$R^n(\Gamma^{k_n}\mathscr{B}) \leqq \sum_{\mathbf{y}\in\mathscr{B}} R^n(\Gamma^{k_n}\{\mathbf{y}\}) \leqq R^n(\mathscr{B}); \qquad \sup_{\mathbf{y}\in\mathscr{Y}^n} \left\{\frac{R^n(\Gamma^{k_n}\{y\})}{R^n(y)}\right\}.$$

Hence by the estimate of Lemma 5,

$$\lim_{n\to\infty} \sup_{\mathscr{B}\subset\mathscr{Y}^n} \frac{1}{n} \cdot \log \frac{R^n(\Gamma^{k_n}\mathscr{B})}{R^n(\mathscr{B})} = 0. \tag{28}$$

From (27) and (28) it follows that

$$|S_n(c, \varepsilon) - S_n(c, \varepsilon_n)| \to 0.$$

This establishes Theorem 1.

4. Source Coding with Side Information. Proof of Theorems 2 and 3

We start with the

Proof of Lemma 1B. This is an obvious analogon of Lemma 1A. Remark that adding to the conditions we had in Lemma 1A that for every $\mathbf{x} \in \mathscr{X}$ and the function $\varphi_x(\mathbf{p}) \triangleq p(x)$ we must have

$$\sum_{u\in\mathscr{U}} \Pr(U = u) \cdot \varphi_x(\mathbf{p}_u) = Q(x)$$

and observing that one of these conditions can be omitted since it follows from the others (Q is a pr.d.!), we get the statement of Lemma 1B by the very same arguments, which led to Lemma 1A.

Next we pass to the

Proof of Theorem 2. The inequality

$$\limsup_{n\to\infty} \hat{S}_n(c, \varepsilon) \leqq \hat{T}(c)$$

easily follows from the proof of Theorem 1. Now we prove that

$$\liminf_{n\to\infty} \hat{S}_n(c, \varepsilon) \geqq \hat{T}(c). \tag{29}$$

As in the deduction following (26) one proves that

$$|\hat{S}_n(c, \varepsilon) - \hat{S}_n(c, \varepsilon_n)| \to 0,$$

for a suitable sequence $\varepsilon_n \to 0$. Thus it is enough to show that

$$\liminf_{\substack{\varepsilon_n \to 0 \\ n \to \infty}} \hat{S}_n(c, \varepsilon_n) \geqq \hat{T}(c). \tag{30}$$

Let us consider a set $\mathscr{B} \subset \mathscr{Y}^n$ with

$$n^{-1} \cdot \log Q^n(\Psi_{\varepsilon_n}(\mathscr{B}) \cap \mathscr{T}_n(Q)) \geqq c.$$

We define on the set $\mathscr{A} \triangleq \Psi_{\varepsilon_n}(\mathscr{B}) \cap \mathscr{T}_n(Q)$ a random variable X^n with distribution

$$\Pr(X^n = \mathbf{x}) = \begin{cases} Q^n(\mathbf{x}) \cdot [Q^n(\mathscr{A})]^{-1} & \text{if } \mathbf{x} \in \mathscr{A}, \\ 0 & \text{otherwise} \end{cases}$$

Y^n is defined by the relation $(X^n, Y^n) \in \mathscr{P}(W^n)$. Now we have $H_{Q^n}(X^n) = \log Q^n(\mathscr{A})$, and, as in (19) and (20) we get

$$H_{R^n}(Y^n) \leqq 1 + (1 - \varepsilon_n) \cdot \log R^n(\mathscr{B}).$$

Paralleling the treatment of Lemmas 2 and 3 we introduce the random variables $\tilde{U}, \tilde{X}, \tilde{Y}$ as follows. Let I be uniformly distributed on $\{1, 2, \ldots, n\}$ and independent of X^n, Y^n. Then put

$$\tilde{U} = (I, X^{I-1}); \qquad \tilde{X} = X_I$$

$$\tilde{Y} = Y_I$$

with the convention that X^0 is a constant. Notice that $(\tilde{U}, \tilde{X}, \tilde{Y}) \in \mathscr{P}(W)$; $H_Q(\tilde{X} | \tilde{U}) = \frac{1}{n} \cdot H_{Q^n}(X^n)$; and, as in Lemma 3,

$$H_R(\tilde{Y} | \tilde{U}) \leqq \frac{1}{n} \cdot H_{R^n}(Y^n).$$

We are done if we show that, roughly speaking, the distribution of $\tilde{X}$ is "close" to Q. Let us introduce for a moment the function

$$t(c, Q, \tilde{Q}) \triangleq \inf_{\substack{H_Q(X|U) \geqq c \\ (U, X, Y) \in \mathscr{P}(W, \tilde{Q})}} H_R(Y | U).$$

Then $t(c, Q, Q) = \hat{T}(c)$. Obviously t is continuous in $\tilde{Q}$ at any nonvanishing $\tilde{Q}$.

Denote by $\tilde{Q}_{\mathscr{B}}$ the distribution of $\tilde{X}$. We have shown that

$$n^{-1} \cdot H_{R^n}(Y^n) \geqq t(n^{-1} \cdot \log Q^n(\mathscr{A}), Q, \tilde{Q}_{\mathscr{B}}).$$

It remains to show that $\tilde{Q}_{\mathscr{B}}$ tends to Q (uniformly in $\mathscr{B}$ as n tends to ∞).

Let us introduce an arbitrary nonvanishing measure μ on $\mathscr{X}$. An elementary computation shows that for any $\mathbf{x} \in \mathscr{T}_n(Q)$

$$|n^{-1} \cdot \log \mu^n(\mathbf{x}) - [H_\mu(Q) - H(Q)]| \to 0$$

uniformly in $\mathbf{x}$. Then

$$|n^{-1} \cdot H_{\mu^n}(X^n) - [H_\mu(Q) - H(Q) + n^{-1} \cdot H(X^n)] \to 0.$$

Since $n^{-1} \cdot H_{\mu^n}(X^n) = H_\mu(\tilde{X} \mid \tilde{U})$; $n^{-1} \cdot H(X^n) = H(\tilde{X} \mid \tilde{U})$, we have

$$|[H_\mu(\tilde{X} \mid \tilde{U}) - H(\tilde{X} \mid \tilde{U})] - [H_\mu(Q) - H(Q)]| \to 0$$

i.e.

$$\sum_{x \in \mathscr{X}} (\tilde{Q}_{\mathscr{B}}(x) - Q(x)) \cdot \log \mu(x) \to 0$$

uniformly in $\mathscr{B}$. Since this holds for an arbitrary nonvanishing measure μ, it implies that

$$\lim_{n \to \infty} \sup_{\mathscr{B} \subset \mathscr{Y}^n} |Q_{\mathscr{B}}(x) - Q(x)| = 0$$

for every $x \in \mathscr{X}$.

This completes the proof of (30).

Theorem 3 follows now easily.

Let us fix an arbitrary $0 < \varepsilon \leqq 1$. Consider a code $f_n(X^n)$, $g_n(Y^n)$, and a decoder $V_n(f_n, g_n)$ which together are ε-reproducing the DMCSS $\{(X_i, Y_i)\}_{i=1}^{\infty}$, i.e. satisfy condition (1).

For a given value u of f_n we denote

$$\mathscr{B}_U \triangleq \{\mathbf{y}; \mathbf{y} = V_n(u, g_n(\mathbf{y}))\}.$$

This means that $\mathscr{B}_u$ is the set of those $\mathbf{y}$'s which are correctly decoded given a value f_n of the code of X^n. With this notation (1) becomes

$$\sum_{\mathbf{x} \in \mathscr{X}^n} \Pr(X^n = \mathbf{x}) \cdot W^n(\mathscr{B}_{f_n(\mathbf{x})} \mid \mathbf{x}) \geqq 1 - \varepsilon.$$

Applying a reverse Markov inequality this yields

$$\Pr(W^n(\mathscr{B}_{f_n(X^n)} \mid X^n) \geqq 1 - \sqrt{\varepsilon}) \geqq 1 - \sqrt{\varepsilon}. \tag{31}$$

Putting

$$\mathscr{A}_n \triangleq \{\mathbf{x}; W^n(\mathscr{B}_{f_n(\mathbf{x})} \mid \mathbf{x}) \geqq 1 - \sqrt{\varepsilon}\} \quad \text{and denoting} \quad \Pr(X_1 = x) = Q(x)$$

we get from (31) that

$$Q^n(\mathscr{A}_n \cap \mathscr{T}_n(Q)) > 1 - 2\sqrt{\varepsilon} \tag{31a}$$

for all sufficiently large n.

We observe that by definition

$$\mathscr{A}_n = \bigcup_u [\Psi_{\sqrt{\varepsilon}}(\mathscr{B}_u) \cap f^{-1}(u)] \quad \text{(a disjoint union)}$$

and thus (31a) implies that there exists a value u^* of f_n such that

$$Q^n(\Psi_{\sqrt{\varepsilon}}(\mathscr{B}_{u^*}) \cap \mathscr{T}_n(Q)) > (1 - 2\sqrt{\varepsilon}) \cdot \|f_n\|^{-1}. \tag{32}$$

On the other hand we also have the obvious estimate

$$\|g_n\| \geqq \|\mathscr{B}_{u^*}\|. \tag{33}$$

Now we shall apply Theorem 2 to this situation in the following set-up:

For R choose the uniform distribution on $\mathscr{Y}$, and for Q the distribution of X_1 With this choice (32) and (33) imply (by the definition of $\hat{S}_n$) that

$$n^{-1}\cdot\log\|g_n\|\geqq\hat{S}_n\left(\frac{1}{n}\cdot[\log(1-2\sqrt{\varepsilon})-\log\|f_n\|],\sqrt{\varepsilon}\right)+\log\|\mathscr{Y}\|,$$

and hence by Theorem 2

$$n^{-1}\cdot\log\|g_n\|\geqq\hat{T}(n^{-1}\cdot[\log(1-2\sqrt{\varepsilon})-\log\|f_n\|])+\log\|\mathscr{Y}\|+\alpha_n \tag{34}$$

where α_n tends to 0.

Consider now an element $(R_1, R_2)\in\mathscr{R}(\varepsilon)$. By the definition of the rate regions there exists a sequence $\{(f_n, g_n, V_n)\}_{n=1}^{\infty}$ of ε-reproductions of the given DMCSS such that $n^{-1}\cdot\log\|f_n\|\to R_1$ and $n^{-1}\cdot\log\|g_n\|\to R_2$. Hence substituting the limits in (34) the continuity of $\hat{T}(c)$ in c implies that

$$R_2\geqq\hat{T}(-R_1)+\log\|\mathscr{Y}\|. \tag{35}$$

Now we observe that since $(U, X, Y)\in\mathscr{P}(W, Q)$, we have $H_Q(X|U)=-I(X\wedge U)$ and since R is the uniform distribution on $\mathscr{Y}$, we also have

$$H_R(Y|U)=H(Y|U)-\log\|\mathscr{Y}\|.$$

By these remarks the triple (U, X, Y) yielding $\hat{T}(-R_1)$ satisfies

$$I(X\wedge U)\leqq R_1 \quad\text{and}\quad \hat{T}(-R_1)=H(Y|U)-\log\|\mathscr{Y}\|. \tag{36}$$

Comparing (35) and (36) we get that $R_2\geqq H(Y|U)$. This and (35) when compared with (3) mean that

$$(R_1, R_2)\in\mathscr{R}$$

what we wanted to prove.

We remark that in proving the strong converse we have not made any use of the weak converse theorem.

5. The Degraded Broadcast Channel. Proof of Theorem 4

The main idea of the proof is that the error probability of every code of the DBC can be decreased substantially by "blowing up" its decoding sets. The original code becomes a list code with so small a list size (non-exponential) that Fano's lemma can still be applied and give the strong converse.

For an arbitrary $0<\tilde{\varepsilon}_t\leqq 1$, $t=1,2$ and a natural number n let us be given an $(n,\tilde{\varepsilon}_1,\tilde{\varepsilon}_2)$-code $\{\mathbf{x}_{ij}, \tilde{\mathscr{A}}_i, \tilde{\mathscr{B}}_j;\ 1\leqq i\leqq M_1,\ 1\leqq j\leqq M_2\}$ for the DBC described in the introduction.

Consider a sequence k_n of integers with $k_n\cdot n^{-1/2}\to\infty$ and $k_n\cdot n^{-1}\to 0$. For any $1\leqq i\leqq M_1$ and $1\leqq j\leqq M_2$ we define the sets

$$\mathscr{A}_i\triangleq\Gamma^{k_n}\tilde{\mathscr{A}}_i;\qquad \mathscr{B}_j\triangleq\Gamma^{k_n}\cdot\tilde{\mathscr{B}}_j.$$

Since our original code was an $(n,\tilde{\varepsilon}_1,\tilde{\varepsilon}_2)$-code, we had $W_1^n(\tilde{\mathscr{A}}_i|\mathbf{x}_{ij})\geqq 1-\tilde{\varepsilon}_1$, $W_2^n(\tilde{\mathscr{B}}_j|\mathbf{x}_{ij})\geqq 1-\tilde{\varepsilon}_2$ for every i,j by definition. Applying Lemma 4 we thus obtain

that $W_1^n(\mathscr{A}_i|\mathbf{x}_{ij})\geqq 1-\varepsilon_n$, $W_2^n(\mathscr{B}_j|\mathbf{x}_{ij})\geqq 1-\varepsilon_n$ for every i, j, where $\varepsilon_n\to 0$ if $n\to\infty$. Notice that this is not any more an "ordinary" code, since the new decoding sets $\mathscr{A}_i$ and $\mathscr{B}_j$ are not disjoint. However every $\mathbf{y}\in\mathscr{Y}^n$ is contained in a small number of $\mathscr{A}_i$'s and the same holds true for the elements of $\mathscr{Z}^n$ and the decoding sets $\mathscr{B}_j$. In fact, denoting

$$\mathscr{N}_1(\mathbf{y})\triangleq\{i;\,\mathbf{y}\in\mathscr{A}_i\}\quad\text{and}\quad\mathscr{N}_2(\mathbf{z})\triangleq\{j,\,\mathbf{z}\in\mathscr{B}_j\}$$

we clearly have $\|\mathscr{N}_1(\mathbf{y})\|\leqq\|\Gamma^{k_n}\{\mathbf{y}\}\|$, since

$$\mathbf{y}\in\mathscr{A}_i=\Gamma^{k_n}\tilde{\mathscr{A}}_i\quad\text{iff}\ \tilde{\mathscr{A}}_i\cap\Gamma^{k_n}\{y\}\neq\emptyset$$

and the $\tilde{\mathscr{A}}_i$'s are disjoint. The same holds for any $\mathbf{z}\in\mathscr{Z}^n$ and the $\mathscr{B}_j^*$'s.

Hence by Lemma 5

$$\|\mathscr{N}_1(\mathbf{y})\|\leqq 2^{n\delta_n}\quad\text{and}\quad\|\mathscr{N}_2(\mathbf{z})\|\leqq 2^{n\delta_n}\tag{37}$$

for every $\mathbf{y}\in\mathscr{Y}^n$ and $\mathbf{z}\in\mathscr{Z}^n$, where $\delta_n\to 0$.

Our remaining argument is just Fano's lemma as in [1]. We introduce a r.v. U ranging over $1\leqq j\leqq M_2$ and taking its values with equal probability. We define

$$\mathscr{T}(j)\triangleq\{\mathbf{x}_{ij};\,1\leqq i\leqq M_1\}$$

and introduce a r.v. X^n which conditional on any fixed value j of U has uniform distribution on $\mathscr{T}(j)$. Y^n and Z^n are defined to be the output r.v.'s on $\mathscr{Y}^n$ and $\mathscr{Z}^n$ if the input is X^n. Clearly,

$$\log M_2=H(U)=I(U\wedge Z^n)+H(U|Z^n).\tag{38}$$

We define for $\mathbf{z}\in\mathscr{Z}^n$

$$\varepsilon(\mathbf{z})\triangleq\Pr(U\notin\mathscr{N}_2(\mathbf{z})\,|\,Z^n=\mathbf{z}).$$

By Fano's inequality and (37) we have

$$H(U|Z^n=\mathbf{z})\leqq h(\varepsilon(\mathbf{z}))+\varepsilon(\mathbf{z})\cdot\log M_2+(1-\varepsilon(\mathbf{z}))\cdot n\delta_n$$

and hence – introducing the notation $\zeta_n\triangleq\sum_{z\in\mathscr{Z}^n}\Pr(Z^n=\mathbf{z})\cdot\varepsilon(\mathbf{z})$ – passing to the expected values on both sides and using the concavity of the entropy h, we obtain

$$H(U|Z^n)\leqq h(\zeta_n)+\zeta_n\cdot\log M_2+(1-\zeta_n)\cdot n\delta_n$$

Substituting this into (38) and observing that $0\leqq\zeta_n\leqq\varepsilon_n$ thus ζ_n also tends to 0, we get that

$$\begin{aligned}n^{-1}\cdot\log M_2&\leqq n^{-1}I(U\wedge Z^n)+h(\zeta_n)+\zeta_n\cdot\log M_2+(1-\zeta_n)\cdot n\delta_n\\&=n^{-1}\cdot I(U\wedge Z^n)+o(1).\end{aligned}\tag{39}$$

Furthermore,

$$I(X^n\wedge Y^n|U)=H(X^n|U)-H(X^n|U,Y^n)=\log M_1-H(X^n|U,Y^n).\tag{40}$$

By Fano's inequality and (37) analogously to the foregoing we obtain

$$H(X^n|U,Y^n)\leqq h(\zeta_n)+\zeta_n\cdot\log M_1+(1-\zeta_n)\cdot n\delta_n$$

and substituting this into (40) we obtain that

$$\begin{aligned} n^{-1}\cdot\log M_1 &\leqq n^{-1}\cdot I(X^n\wedge Y^n|U)+n^{-1}\cdot h(\zeta_n)\\ &\quad +n^{-1}\cdot\zeta_n\cdot\log M_1+(1-\zeta_n)\cdot n^{-1}\cdot n\delta_n\\ &=n^{-1}\cdot I(X^n\wedge Y^n|U)+o(1). \end{aligned} \tag{41}$$

By the weak converse to the DBC coding theorem [1] for $R_1=n^{-1}\cdot I(X^n\wedge Y^n|U)$; $R_2=n^{-1}\cdot I(U\wedge Z^n)$ we have $(R_1,R_2)\in\mathscr{C}$. Hence observing that any element of $\mathscr{C}(\varepsilon_1,\varepsilon_2)$ can be obtained as limit of code rates $(n^{-1}\cdot\log M_1, n^{-1}\cdot\log M_2)$, the relations (39), (41) and the closedness of $\mathscr{C}$, yield that

$$\mathscr{C}(\varepsilon_1,\varepsilon_2)\subset\mathscr{C}$$

which proves Theorem 4.

6. On a Theorem of Margulis

Given the sets $\mathscr{X}$, $\mathscr{Y}$ and a transition probability matrix W from $\mathscr{X}$ to $\mathscr{Y}$ we denote by m_w the smallest non-zero element of W. In this section we use the natural logarithm ln.

We prove that

Theorem 5. *There is a constant a depending only on W such that for any $\mathscr{B}\subset\mathscr{Y}^n$ and $\mathbf{x}\in\mathscr{X}^n$*

$$W^n(\partial\mathscr{B}|\mathbf{x})\geqq a\cdot n^{-1/2}\cdot f(W^n(\mathscr{B}|\mathbf{x})). \tag{42}$$

Proof. We put $a\triangleq\frac{1}{3}\cdot m_w\cdot(-\ln m_w)^{-1/2}$.

The proof goes by induction based on two simple combinatorial observations. For $\mathscr{B}\subset\mathscr{Y}^n$ we define the following subsets of $\mathscr{Y}^{n-1}$:

$$\mathscr{B}_y\triangleq\{\mathbf{v}\in\mathscr{Y}^{n-1};\ \mathbf{v}\,y\in\mathscr{B}\}.$$

Notice that $\mathscr{B}$ is the disjoint union of the sets $\mathscr{B}_y y$ and

$$W_n(\mathscr{B}|\mathbf{x})=\sum_y W(y|x_n)\cdot W^{n-1}(\mathscr{B}_y|\mathbf{x}^{n-1})$$

where $\mathbf{x}=x_1x_2\ldots x_n$, $\mathbf{x}^{n-1}=x_1x_2\ldots x_{n-1}$, $\mathbf{x}=\mathbf{x}^{n-1}x_n$.

We use the inequalities

(i) $W^n(\partial\mathscr{B}|\mathbf{x})\geqq\sum_y W(y|x_n)\cdot W^{n-1}(\partial\mathscr{B}_y|\mathbf{x}^{n-1})$,

(ii) $W^n(\partial\mathscr{B}|\mathbf{x})\geqq m_w\cdot d$

where

$$d=\max_{y\in\mathscr{S}_{x_n}} W^{n-1}(\mathscr{B}_y|\mathbf{x}^{n-1})-\min_{y\in\mathscr{S}_{x_n}} W^{n-1}(\mathscr{B}_y|\mathbf{x}^{n-1})$$

and $\mathscr{S}_x=\{y;\ W(y|x)>0\}$.

(i) follows from the fact that

$$\partial\mathscr{B} \supseteq \bigcup_y [(\partial\mathscr{B}_y) \times \{y\}].$$

To prove (ii) observe that for any $y_0, y_1 \in \mathscr{Y}$

$$\partial\mathscr{B} \supseteq (\mathscr{B}_{y_0} - \mathscr{B}_{y_1}) \times \{y_0\}.$$

We start with some analytic properties of f:

$$f' = \sqrt{-2\ln(\sqrt{2\pi}f)}; \quad f'' = -\frac{1}{\sqrt{2\pi}} f \tag{43}$$

$f(s)$ is defined on $0 \leqq s \leqq 1$, it is concave and symmetric around $\frac{1}{2}$. (Notice that in 0 $f(s)$ is asymptotically equal to $s \cdot \sqrt{-2 \cdot \ln \cdot s}$, though this will not be used in the sequel.)

Denote $\bar{s} \triangleq \min\{s, 1-s\}$. Then obviously

$$f(s) \geqq 2 \cdot (2\pi)^{-1/2} \cdot \bar{s} \geqq \bar{s}^2. \tag{44}$$

(It suffices to check this at $s = \frac{1}{2}$.)

Hence using (43)

$$|f'(s)| \leqq 2\sqrt{-\ln \bar{s}}. \tag{45}$$

Starting the induction proof one easily sees that (42) holds for $n = 1$. Suppose that it is true for $n-1$.

Now we consider two cases. Introducing the notation $c \triangleq (3 \cdot \sqrt{-\ln m_w})^{-1}$. Suppose first

$$d \geqq c \cdot n^{-1/2} \cdot f(W^n(\mathscr{B}|\mathbf{x})).$$

Then (42) follows from (ii). Now suppose

$$d < c \cdot n^{-1/2} \cdot f(W^n(\mathscr{B}|\mathbf{x})). \tag{46}$$

By (i) and the induction hypothesis we have

$$\begin{aligned} W^n(\partial\mathscr{B}|\mathbf{x}) &\geqq \sum_y W(y|x_n) \cdot W^{n-1}(\partial\mathscr{B}_y|\mathbf{x}^{n-1}) \\ &\geqq \sum_y W(y|\mathbf{x}_n) \cdot m_w \cdot c \cdot (n-1)^{-1/2} \cdot f(W^{n-1}(\mathscr{B}_y|\mathbf{x}^{n-1})). \end{aligned} \tag{47}$$

Denote $s \triangleq W^n(\mathscr{B}|\mathbf{x})$, $s_y \triangleq W^{n-1}(\mathscr{B}_y|\mathbf{x}^{n-1})$, and consider the interval

$$\Delta \triangleq [\min_y s_y, \max_y s_y].$$

By Taylor's formula

$$f(s_y) = f(s) + (s_y - s) \cdot f'(s) + \tfrac{1}{2} \cdot (s_y - s)^2 \cdot f''(\sigma_y) \quad \text{where } \sigma_y \in \Delta.$$

Hence

$$\sum_y W(y|x_n) \cdot f(s_y) \geqq f(s) - \tfrac{1}{2} \cdot d^2 \cdot \max_{\sigma \in \Delta} |f''(\sigma)|.$$

This, (46) and (43) imply in (47) that

$$W^n(\partial\mathscr{B}|\mathbf{x}) \geqq m_w \cdot c \cdot (n-1)^{-1/2} \cdot \left[f(s) - (2n)^{-1} \cdot c^2 \cdot f^2(s) \cdot \max_{\sigma \in \Delta} \left| \frac{1}{f(\sigma)} \right| \right].$$

We denote by s_0 the point of Δ, where $f(\sigma)$ takes its minimum. By a simple rearrangement one gets:

$$W^n(\partial\mathscr{B}|\mathbf{x}) \geqq m_w \cdot c \cdot n^{-1} \cdot f(s) \cdot \left[\sqrt{\frac{n}{n-1}} - \frac{c^2 \cdot f(s)}{2 \cdot f(s_0) \cdot \sqrt{n(n-1)}} \right].$$

It is enough to show that the term in brackets is not smaller than 1. This is equivalent to

$$f(s_0) \cdot (f(s))^{-1} \geqq c^2 \cdot (\sqrt{n} + \sqrt{n-1}) \cdot (2 \cdot \sqrt{n})^{-1}.$$

Therefore we are ready if we show that

$$f(s_0) \cdot (f(s))^{-1} \geqq c^2. \tag{48}$$

Using Lagrange's formula we have

$$f(s_0) \geqq f(s) - d \cdot |f'(\sigma)|$$

for some $\sigma \in \Delta$. Applying (45) and (46) this becomes

$$f(s_0) \geqq f(s) \left(1 - 2c \cdot \sqrt{\frac{-\ln \bar{s}}{n}} \right).$$

Since our distribution is a finite one, we know that

$$\bar{s} \geqq m_w^n$$

hence writing out c,

$$\frac{f(s_0)}{f(s)} \geqq 1 - 2 \cdot m_w \cdot (3\sqrt{-\ln m_w})^{-1} \cdot \sqrt{-\ln m_w} = 1 - \tfrac{2}{3} m_w \geqq \tfrac{1}{3}$$

while clearly $c^2 < \frac{1}{3}$, which proves (48).

Note. The estimate given by this form of Margulis's theorem is exact up to a multiplicative constant, as it can be verified either directly, or by this same method, for "spheres" in $\{0, 1\}^n$.

Recently, Katona [5] showed by combinatorial methods the exact result that – roughly speaking – among all the subsets of $\{0, 1\}^n$ with given cardinality the "spheres" have smallest "surface". (The surface of a set $\mathscr{B}$ is $\partial\mathscr{B}$).

Acknowledgement. During this research the authors had many useful conversations with I. Csiszár. Thanks are also due to G. Tusnády for his remarks concerning Theorem 5.

References

1. Ahlswede, R., Körner, J.: Source Coding with Side Information and a Converse for Degraded Broadcast Channels. IEEE Trans. Information Theory. Vol. IT-21, 629–637 (1975)
2. Bergmans, P. P.: Random Coding Theorem for Broadcast Channels with Degraded Components. IEEE-IT **19**, 197–207 (1973)

3. Cover, T.: Broadcast Channels. IEEE-IT **18**, 2–14 (1972)
4. Gallager, R. G.: Coding for Degraded Broadcast Channels. (To appear)
5. Katona, G. O. H.: The Hamming-sphere has minimum boundary. (To appear in Studia Sci. Math. Hungar.)
6. Kullback, S.: Information Theory and Statistics. New York: Wiley 1959
7. Loève, M.: Probability Theory. pp. 157 and 28–42. New York: Van Nostrand 1955
8. Margulis, G. A.: Veroyatnostniye characteristiki grafov s bolshoy svyaznostyu. [In Russian] Problemy Peredači. Informačii, **X**, 101–108 (1974)
9. Wyner, A. D.: A theorem on the entropy of certain binary sequences and applications. Part II. IEEE-IT **19**, 769–777 (1973)
10. Wolfowitz, J.: Coding Theorems of Information Theory. 2nd edition. Berlin-Heidelberg-New York: Springer 1964
11. Rényi, A.: Wahrscheinlichkeitsrechnung. (Exercise 24 on p. 137.) Berlin: VEB Deutscher Verlag der Wissenschaften, 1962

Received February 5, 1975; In revised form October 30, 1975