nonanticipatory channels," *Inform. and Contr.*, vol. 26, pp. 381–391, 1971.
[26] A. I. Khinchine, *Mathematical Foundations of Information Theory*. New York: Dover, 1957.
[27] B. McMillan, "The basic theorems of information theory," *Ann. Math. Stat.*, vol. 24, pp. 196–219, 1953.
[28] J. Moser, E. Phillips, and S. Varadhan, *Ergodic Theory: A Seminar*. New York: Courant Inst. of Math. Sciences, 1975.
[29] J. Nedoma, "The capacity of a discrete channel," *Trans. First Prague Conf. Inform. Theory*, pp. 143–181, Prague, 1957.
[30] ——, "On nonergodic channels," *Trans. Second Prague Conf. Inform. Theory*, pp. 363–395, 1960.
[31] D. L. Neuhoff, R. M. Gray, and L. D. Davisson, "Fixed rate universal block source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 511–528, 1975.
[32] D. S. Ornstein, "An application of ergodic theory to probability theory," *Ann. Prob.*, vol. 1, pp. 43–58, 1973.
[33] ——, *Ergodic Theory, Randomness, and Dynamical Systems*. New Haven, CT: Yale Univ. Press, 1974.
[34] J. C. Oxtoby, "Ergodic Sets," *Bull. Amer. Math. Soc.*, vol. 58, pp. 116–136, 1952.
[35] K. R. Parthasarathy, "On the integral representation of the rate of transmission of a stationary channel," *Ill. J. Math.*, vol. 2, pp. 299–305, 1961.
[36] ——, "Effective entropy rate and transmission of information through channels with additive random noise," *Sankhya*, vol. A25, pp. 75–84, 1963.
[37] ——, *Probability Measures on Metric Spaces*. New York: Academic, 1968.
[38] E. Pfaffelhuber, "Channels with asymptotically decreasing memory and anticipation," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 379–385, 1971.
[39] M. Pinsker, *Information and Information Stability of Random Variables and Processes*. Moscow: Izd. Akad. Nauk. SSSR, 1960 (Translation: San Francisco, CA: Holden-Day, 1964).
[40] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.
[41] P. C. Shields, *The Theory of Bernoulli Shifts*. Chicago, IL: Univ. of Chicago Press, 1973.
[42] V. Strassen, "The existence of probability measures with given marginals," *Ann. Math. Stat.*, vol. 36, pp. 423–439, 1965.
[43] K. Takano, "On the basic theorems of information theory," *Ann. Inst. Stat. Math.*, vol. 9, pp. 53–77, 1957.
[44] I. P. Tsaregradsky, "On the capacity of a stationary channel with finite memory," *Teor. Veroyatnosti i Primenen*, vol. 3, pp. 84–96, 1958.
[45] K. Winkelbauer, "Channels with finite history," in *Trans. 2nd Prague Conf. Inform. Theory*, pp. 685–831, Prague, 1960.
[46] ——, "On discrete information sources," in *Trans. 3rd Prague Conf. Inform. Theory*, pp. 765–830, Prague, 1964.
[47] ——, "On the coding theorem for decomposable discrete information channels: I," *Kybernetika*, vol. 7, pp. 109–123, 1971.
[48] ——, "On the coding theorem for decomposable discrete information channels: II," *Kybernetika*, vol. 7, pp. 230–255, 1971.
[49] ——, "On the regularity condition for decomposable communication channels," *Kybernetika*, vol. 7, pp. 314–327, 1971.
[50] J. Wolfowitz, "Strong converse of the coding theorem for the general discrete finite-memory channel," *Inform. and Contr.*, vol. 3, pp. 89–93, 1960.
[51] J. Wolfowitz, *Coding Theorems of Information Theory*, 2nd ed. Berlin: Springer-Verlag, 1964.
[52] Shen S. Yi, "The fundamental problem of stationary channel," in *Trans. Third Prague Conf. Inform. Theory*, pp. 637–639, Prague, 1962.
[53] D. L. Neuhoff and P. C. Shields, "Channels with almost finite memory," to appear, *IEEE Trans. Inform. Theory*.

A Coding Theorem for the Discrete Memoryless Broadcast Channel

KATALIN MARTON

Abstract—**A coding theorem for the discrete memoryless broadcast channel is proved for the case where no common message is to be transmitted. The theorem is a generalization of the results of Cover and van der Meulen on this problem. The result is tight for broadcast channels having one deterministic component**

I. Introduction

A DISCRETE memoryless broadcast channel (DMBC) as defined by Cover [1] is determined by a pair of discrete memoryless channels with common input alphabet $\mathcal{Y}$. We denote by F and G the transition probability matrices of these channels:

$$F=\{F(x|y): y\in\mathcal{Y}, x\in\mathcal{X}\},\quad G=\{G(z|y): y\in\mathcal{Y}, z\in\mathcal{Z}\}.$$

Here $\mathcal{X}$ and $\mathcal{Z}$ are the output alphabets, with cardinalities $\|\mathcal{Y}\|$, $\|\mathcal{X}\|$, $\|\mathcal{Z}\|<\infty$. The DMBC corresponding to the matrices F, G will be denoted by (F,G).

We assume that the conditional probabilities of receiving the sequences $x^n\in\mathcal{X}^n$ and $z^n\in\mathcal{Z}^n$ at the outputs of the channels F and G, respectively, are given by

$$F^n(x^n|y^n)=\prod_{i=1}^{n}F(x_i|y_i),\quad G^n(z^n|y^n)=\prod_{i=1}^{n}G(z_i|y_i)$$

where $y^n=y_1y_2\cdots y_n$.

In connection with the DMBC (F,G), we consider the following coding problem. A sender has to transmit two independent messages over the channels F and G: one message for receiver I observing the output of the channel

Manuscript received November 28, 1977; revised October 26, 1978. This paper was presented at the IEEE Symposium on Information Theory, Ithaca, NY, 1977.

The author is with the Mathematical Institute of the Hungarian Academy of Sciences, H-1053 Budapest, Reáltanoda u. 13-15, Hungary.

0018-9448/79/0500-0306$00.75

F, and another for receiver II observing the output of the channel G. The messages take their values in the sets $\{1,2,\cdots,J\}$ and $\{1,2,\cdots,K\}$. The sender uses a block code of block length n. Given that the first message has value j and the second message has value k, the sender transmits a sequence $y_{jk}^n \in \mathcal{Y}^n$. The question is at which rate pairs $(n^{-1}\log J, n^{-1}\log K)$ can this be done so that both receivers can with high probability decode their respective messages correctly. The asymptotic values of these rate pairs consititute the capacity region of the broadcast channel.

No computable formulas are known for the capacity region of the DMBC, except for three special cases: 1) if one of the component channels is "more capable" in the sense of [2] than the other one; 2) if the DMBC (F,G) is the product of the DMBC's (F_1,G_1) and (F_2,G_2) (i.e., $\mathcal{Y}=\mathcal{Y}_1\times\mathcal{Y}_2$, $\mathcal{X}=\mathcal{X}_1\times\mathcal{X}_2$, $\mathcal{Z}=\mathcal{Z}_1\times\mathcal{Z}_2$, $F(x_1,x_2|y_1,y_2)=F_1(x_1|y_1)F_2(x_2|y_2)$, $G(z_1,z_2|y_1,y_2)=G_1(z_1|y_1)G(z_2|y_2)$, for $(y_1,y_2)\in\mathcal{Y}$, $(x_1,x_2)\in\mathcal{X}$, $(z_1,z_2)\in\mathcal{Z}$), G_1 is a degraded version of F_1, and F_2 is a degraded version of G_2; and 3) if the DMBC is deterministic, i.e., F and G are $(0,1)$-matrices. Case 1) has been solved recently in [3], generalizing earlier results of [13], [4], [5], and [2]. (See also [6] for a related problem.) Case 2) is settled in [15].

Although case 3) now seems almost trivial, it had been an open problem for a long time. (The Blackwell channel is a deterministic DMBC.) It was settled independently by Pinsker [7] and the author [9]. (See also [8] for a particular case; [9] contains only a heuristic proof.)

In the general case only an inner bound to the capacity region is known. This can be obtained from the results of van der Meulen [10] and Cover [11] (see [12]). The aim of the present paper is to prove a better inner bound (Theorem 2) to the capacity region of the DMBC. This bound is proved by a random coding method which is a combination of the coding techniques of Bergmans [13] and Cover and van der Meulen [10], [11], with the random coding technique used to prove source coding theorems in rate distortion theory.

Theorem 3 is an easy consequence of Theorem 2 and describes the capacity region of the deterministic DMBC. It is generalized by Theorem 4 to DMBC's with one deterministic component. Theorem 4 has also been proved independently by Gelfand and Pinsker [17]. The converse part of Theorem 4 is a special case of an outer bound for the general DMBC (Theorem 5), due to Körner and the author [16].

Theorems 3 and 4 show that Theorem 2 is more than a formal generalization of Theorem 1. As a matter of fact, we shall show in Appendix II that Theorem 1 is not tight for the Blackwell channel.

II. Definitions and Results

Definition: For $n=1,2,\cdots$ a set of *codewords* of length n

$$\{y_{jk}: 1\leqslant j\leqslant J,\ 1\leqslant k\leqslant K\}\subseteq\mathcal{Y}^n$$

is an (n,ϵ)-code $(\epsilon>0)$ for the DMBC (F,G) if there exist two disjoint families of *decoding sets*

$$\{\mathcal{A}_j: 1\leqslant j\leqslant J\},\quad \{\mathcal{B}_k: 1\leqslant k\leqslant K\}$$

$(\mathcal{A}_j\subseteq\mathcal{X}^n,\ \mathcal{B}_k\subseteq\mathcal{Z}^n,\ \mathcal{A}_j\cap\mathcal{A}_{j'}=\mathcal{B}_k\cap\mathcal{B}_{k'}=\varnothing$, for all $j\neq j'$, $k\neq k')$ such that

$$\frac{1}{JK}\sum_{j,k}\left[2-F^n(\mathcal{A}_j|y_{jk})-G^n(\mathcal{B}_k|y_{jk})\right]<\epsilon. \tag{1}$$

The pair of numbers $(n^{-1}\log J, n^{-1}\log K)$ is the *rate pair* of the code. A pair of numbers is called *achievable* if, for any fixed $\epsilon>0$, it can be approximated by rate pairs of (n,ϵ)-codes. The capacity region of the channel is the set of all achievable pairs.

We shall use the following notation. All random variables (r.v.'s) in the paper are supposed to have finite ranges. The symbols W, U, and V will always denote r.v.'s and Y, X, and Z will denote r.v.'s with ranges $\mathcal{Y}$, $\mathcal{X}$, and $\mathcal{Z}$, respectively. We write $(Y,X,Z)\in\mathcal{P}(F,G)$ if the conditional distributions of X and Z given Y are defined by the matrices F and G, respectively. $(U,Y,X,Z)\in\mathcal{P}(F,G)$ will mean that 1) $(Y,X,Z)\in\mathcal{P}(F,G)$, and 2) both triples (U,Y,X) and (U,Y,Z) are Markov chains.

We recall the following inner bound for the capacity region of the DMBC.

Theorem 1 (Cover–van der Meulen–Hajek–Pursley): Let $\hat{\mathcal{R}}$ denote the convex closure of the set

$$\begin{aligned}\{(R_x,R_z):\ & R_x,R_z\geqslant 0,\ R_x\leqslant I(WU\wedge X),\ R_z\leqslant I(WV\wedge Z),\ R_x+R_z\leqslant \min\{I(W\wedge X),I(W\wedge Z)\}+I(U\wedge X|W)+I(V\wedge Z|W)\ \text{for some}\ ((U,V,W),Y,X,Z)\in\mathcal{P}(F,G)\ \text{such that}\ U,\ V,\ \text{and}\ W\ \text{are independent}\}.\end{aligned}$$

Then any rate pair $(R_x,R_z)\in\hat{\mathcal{R}}$ is achievable for the DMBC (F,G).

Our goal is to prove the following generalization of this theorem.

Theorem 2: Let

$$\begin{aligned}\mathcal{R}=\{(R_x,R_z):\ & R_x,R_z\geqslant 0,\ R_x\leqslant I(WU\wedge X),\\ & R_z\leqslant I(WV\wedge Z),\ R_x+R_z\\ & \leqslant \min\{I(W\wedge X),I(W\wedge Z)\}+I(U\wedge X|W)\\ & +I(V\wedge Z|W)-I(U\wedge V|W)\\ & \text{for some}\ ((U,V,W),Y,X,Z)\in\mathcal{P}(F,G)\}.\end{aligned}$$

Then any rate pair $(R_x,R_z)\in\mathcal{R}$ is achiev1bl5 for t8e DMBC (F,G).

Remarks: 1) It is easy to see that $\mathcal{R}$ is convex. 2) In the definition of $\mathcal{R}$, no condition on the independence of the r.v.'s W, U, V is imposed. The set $\hat{\mathcal{R}}$ consists of those rate pairs in $\mathcal{R}$ that correspond to independent r.v.'s U, V, W. 3) Consider for a moment the subset $\mathcal{R}_0\subseteq\mathcal{R}$ consisting of

rate pairs corresponding to $W=$ const.:

$$\mathcal{R}_0=\{(R_x,R_z):\ R_x,R_z\geqslant 0,$$
$$R_x\leqslant I(U\wedge X),\ R_z\leqslant I(V\wedge Z),$$
$$R_x+R_z\leqslant I(U\wedge X)+I(V\wedge Z)-I(U\wedge V)$$
$$\text{for some } ((U,V),Y,X,Z)\in\mathcal{P}(F,G)\}.$$

We believe that the novelty of Theorem 2 is essentially in establishing that any rate pair in $\mathcal{R}_0$ is achievable. Let us give a heuristic reason why the rate pairs in $\mathcal{R}_0$ should be achievable. Let (U^n,Y^n,Z^n) denote the length n output of the discrete memoryless correlated source with generic variable (U,Y,Z). It can be shown by the method used in [14] that, for some sequence of positive numbers $\{\delta_n\}$ with $\delta_n\to 0$,

$$\lim_{n\to\infty} n^{-1}\max\{I(V^*\wedge Z^n):\ V^* \text{ is an r.v. such that } n^{-1}I(V^*\wedge U^n)<\delta_n \text{ and } (V^*,Y^n,Z^n) \text{ is a Markov chain}\}=\max\{I(V\wedge Z)-I(U\wedge V):\ (V,Y,Z) \text{ is a Markov chain}\}.$$

It is easy to see from this that, for $(U,Y,X,Z)\in\mathcal{P}(F,G)$ and large enough n, we can construct an r.v. V_n^* such that $((U^n,V_n^*),Y^n,X^n,Z^n)\in\mathcal{P}(F^n,G^n)$,

$$\lim_{n\to\infty} n^{-1}I(V_n^*\wedge Z^n)=\max\{I(V\wedge Z)-I(U\wedge V):\ ((U,V),Y,X,Z)\in\mathcal{P}(F,G)\}$$

and

$$\lim_{n\to\infty} n^{-1}I(V_n^*\wedge U^n)=0 \tag{2}$$

i.e., V_n^* is asymptotically independent of U^n. We also have $n^{-1}I(U^n\wedge X^n)=I(U\wedge X)$ for all n. Therefore, if we had $I(V_n^*\wedge U^n)=0$ for all n, instead of (2), the achievability of the pair $R_x=I(U\wedge X)$, $R_z=\max_V[I(V\wedge Z)-I(U\wedge V)]$ (and hence of $\mathcal{R}_0$) would follow from the Cover–van der Meulen theorem.

Theorem 3 (Pinsker–Marton): If (F,G) is a deterministic broadcast channel, i.e., if F and G are $(0,1)$ matrices, then the capacity region of (F,G) is

$$\mathcal{R}=\{(R_x,R_z):\ R_x,R_z\geqslant 0,\ R_x\leqslant H(X),\ R_z\leqslant H(Z),\ R_x+R_z\leqslant H(XZ) \text{ for some } (Y,X,Z)\in\mathcal{P}(F,G)\}.$$

The direct part of this theorem follows from Theorem 2 by defining $W=$ const, $U=X$, $V=Z$. The converse is trivial. Pinsker extended this theorem to the case of multicomponent broadcast channels, all components of which are deterministic.

Theorem 4 (Gelfand–Pinsker–Marton): If (F,G) is a DMBC for which F is a $(0,1)$ matrix then the capacity region of (F,G) is

$$\mathcal{R}=\{(R_x,R_z):\ 0\leqslant R_x\leqslant H(X),\ 0\leqslant R_z\leqslant I(V\wedge Z),\ R_x+R_z\leqslant H(X|V)+I(V\wedge Z) \text{ for some } (V,Y,X,Z)\in\mathcal{P}(F,G)\}.$$

Moreover, this region remains unchanged if V is allowed to take at most $\|\mathcal{Y}\|+2$ values.

The direct part of Theorem 4 follows from Theorem 2 with $W=$ const, $U=X$. The converse is a special case of the following outer bound for the capacity region of a DMBC [16].

Theorem 5 (Körner–Marton): The capacity region $\mathcal{R}$ of the DMBC (F,G) satisfies

$$\mathcal{R}\subseteq\{(R_x,R_z):\ 0\leqslant R_x\leqslant I(Y\wedge X),\ 0\leqslant R_z\leqslant I(V\wedge Z),\ R_x+R_z\leqslant I(Y\wedge X|V)+I(V\wedge Z) \text{ for some } (V,Y,X,Z)\in\mathcal{P}(F,G)\}. \tag{3}$$

Moreover V can be assumed to take at most $\|\mathcal{Y}\|+2$ values.

Remarks: 1) A similar outer bound can be obtained for the rate region of codes all codewords of which have the same composition, say Q. Combining the bound for fixed Q with that obtained by reversing the roles of F and G, and than letting Q vary over the distributions on $\mathcal{Y}$, the bound (3) can be improved. 2) Theorem 5 can be proved either by the method of "images of a set via two channels" used in [6], or by using only a single-letter technique for information quantities as in [18]. (Such a technique is also implicitly contained in the first method.) Here we give a proof of the second type, since it is simpler. However, the method used in [6] would give a stronger result: namely, that (1) holds also with $\mathcal{R}(\epsilon)$ (the region of the so-called ϵ-achievable rates) instead of $\mathcal{R}$. This means that Theorem 4 holds with a strong converse.

Theorem 5 is proved in Appendix I.

In the proof of Theorem 2 we use the following notation:

P_X	distribution of the r.v. X;
$P_{U\|W}$	conditional distribution of the r.v. U given the r.v. W;
P_X^n and $P_{U\|W}^n$	denote the nth memoryless extensions of these distributions;

for a finite set $\mathcal{W}$, $a\in\mathcal{W}$, $n=1,2\cdots$ and $w^n=w_1w_2\cdots w_n\in\mathcal{W}^n$, $n(a|w^n)\triangleq\{i:\ w_i=a\}\|$; for an r.v. W with range $\mathcal{W}$, $\eta>0$ and $n=1,2,\cdots$,

$$\mathcal{T}_W^n(\eta)\triangleq\{w^n\in\mathcal{W}^n:\ |n^{-1}n(a|w^n)-P_W(a)|<\eta, \text{ all } a\in\mathcal{W}\};$$

for a pair of r.v.'s (W,X) with range $\mathcal{W}\times\mathcal{X}$, for $\eta_1>0$, for a sequence $w^n\in\mathcal{T}_W^n(\eta_1)$ and for $\eta_2\geqslant\eta_1$,

$$\mathcal{T}_X(w^n,\eta_2)\triangleq\{X^n\in\mathcal{X}^n:\ |n^{-1}n(ab|w^nx^n)-P_{WX}(ab)|<\eta_2, \text{ all } (a,b)\in\mathcal{W}\times\mathcal{X};\ n(ab|w^nx^n)=0 \text{ for } P_{WX}(a,b)=0\}.$$

III. Proof of Theorem 2

It suffices to prove that the pair

$$R_x\triangleq I(WU\wedge X)=I(W\wedge X)+I(U\wedge X|W),$$
$$R_z\triangleq\min\{I(W\wedge X),I(W\wedge Z)\}+I(U\wedge X|W)+I(V\wedge Z|W)-I(U\wedge V|W)-R_x$$
$$=I(V\wedge Z|W)-I(U\wedge V|W)-|I(W\wedge X)-I(W\wedge Z)|_+$$

is achievable for $((W,U,V),Y,X,Z)\in\mathcal{P}(F,G)$. We may assume $R_z>0$, i.e.,

$$I(U\wedge V|W)<I(V\wedge Z|W)-|I(W\wedge X)-I(W\wedge Z)|_+.$$

Fix the numbers $\epsilon,\delta,\eta>0$. For n fixed, define[1]

$$I=[\exp(n(I(W\wedge X)-\delta))] \tag{4}$$

$$J=[\exp(n(I(U\wedge X|W)-\delta))] \tag{5}$$

$$L=[\exp(n(I(U\wedge V|W)+\delta))] \tag{6}$$

$$K=[\exp(n(I(V\wedge Z|W)-|I(W\wedge X)-I(W\wedge Z)|_+ - I(U\wedge V|W)-2\delta))]. \tag{7}$$

We then have

$$KL\leqslant\exp(n(I(V\wedge Z|W)-\delta)) \tag{8}$$

$$IKL\leqslant\exp(n(I(WV\wedge Z)-\delta)). \tag{9}$$

Using a random coding method, we shall define length n codewords y_{ijk} $(1\leqslant i\leqslant I,\ 1\leqslant j\leqslant J,\ 1\leqslant k\leqslant K)$ and decoding sets $\mathcal{A}_{ij}\subseteq\mathcal{X}^n$, $\mathcal{B}_k\subseteq\mathcal{Y}^n$ so that (1) holds with the index j replaced by the pair of indices (ij) (and ϵ replaced by const ϵ).

Select the length n sequences w_i $(1\leqslant i\leqslant I)$ independently of each other and according to the distribution P_W^n. Then, for fixed i, select the length n sequences u_{ij} $(1\leqslant j\leqslant J)$ and v_{ikl} $(1\leqslant k\leqslant K,\ 1\leqslant l\leqslant L)$ independently of each other and according to the distributions $P_{U|W}^n(\cdot|w_i)$ and $P_{V|W}^n(\cdot|w_i)$, respectively. The set

$$\{w_i,u_{ij},v_{ikl}\colon 1\leqslant i\leqslant I,\ 1\leqslant j\leqslant J,\ 1\leqslant k\leqslant K,\ 1\leqslant l\leqslant L\}$$

will be called an auxiliary random code.

For i,j,k fixed, let χ_{ijk} denote the indicator variable of the event that the triple of sequences $(w_iu_{ij}v_{ikl})$ belongs to $\mathcal{T}_{WUV}^n(\eta/2)$ for at least one value of l. This value of l will be denoted by $l(ijk)$. From the rule for selecting our auxiliary random code, and from (6) it follows that, for sufficiently large n,

$$\Pr\{\chi_{ijk}=0\}<\epsilon. \tag{10}$$

If $\chi_{ijk}=1$, let the codeword y_{ijk} be any sequence in $\mathcal{T}_Y(w_iu_{ij}v_{ikl(ijk)},\eta)$ and let y_{ijk} be arbitrary for $\chi_{ijk}=0$. Set

$$\hat{\mathcal{A}}_{ij}=\mathcal{T}_X(w_iu_{ij},2\eta)$$

$$\hat{\mathcal{B}}_{ikl}=\mathcal{T}_Z(w_iv_{ikl},2\eta),$$

$$\hat{\mathcal{B}}_{ik}=\bigcup_l\hat{\mathcal{B}}_{ikl},$$

$$\hat{\mathcal{B}}_k=\bigcup_{i,l}\hat{\mathcal{B}}_{ikl}=\bigcup_i\hat{\mathcal{B}}_{ik}.$$

We define the decoding sets $\mathcal{A}_{ij}$ and $\mathcal{B}_k$ by

$$\mathcal{A}_{ij}=\hat{\mathcal{A}}_{ij}\setminus\cup\{\mathcal{A}_{i'j'}\colon(i'j')\neq(ij)\}$$

$$\mathcal{B}_k=\hat{\mathcal{B}}_k\setminus\cup\{\hat{\mathcal{B}}_{k'}\colon k'\neq k\}.$$

For i,j,k fixed, we shall estimate the expected value of the quantities

$$1-F^n(\mathcal{A}_{ij}|y_{ijk})\quad\text{and}\quad 1-G^n(\mathcal{B}_k|y_{ijk})$$

over the ensemble of the auxiliary random codes.

We have

$$\begin{aligned}\chi_{ijk}[1-F^n(\mathcal{A}_{ij}|y_{ijk})]&=\chi_{ijk}[1-F^n(\hat{\mathcal{A}}_{ij}|y_{ijk})]\\&\quad+\chi_{ijk}F^n(\hat{\mathcal{A}}_{ij}\cap\cup\{\hat{\mathcal{A}}_{i'j'}\colon(i'j')\neq(ij)\}|y_{ijk})\\&=\chi_{ijk}[1-F^n(\hat{\mathcal{A}}_{ij}|y_{ijk})]\\&\quad+\chi_{ijk}\sum_{x^n\in\hat{\mathcal{A}}_{ij}}F^n(x^n|y_{ijk})\tau\Big(x^n,\bigcup_{(i'j')\neq(ij)}\mathcal{A}_{i'j'}\Big)\end{aligned} \tag{11}$$

where

$$\tau(x^n,\mathcal{A})=\begin{cases}1, & \text{if } x^n\in\mathcal{A}\\ 0, & \text{otherwise.}\end{cases}$$

Furthermore, since $\hat{\mathcal{B}}_k\supseteq\hat{\mathcal{B}}_{ik}$, and, consequently,

$$\mathcal{B}_k\supseteq\hat{\mathcal{B}}_{ik}\setminus\cup\{\hat{\mathcal{B}}_{k'}\colon k'\neq k\}$$

we have, as in (11)

$$\chi_{ijk}[1-G^n(\mathcal{B}_k|y_{ijk})]\leqslant\chi_{ijk}[1-G^n(\hat{\mathcal{B}}_{ik}|y_{ijk})]+\sum_{z^n\in\hat{\mathcal{B}}_{ik}}G^n(z^n|y_{ijk})\tau\Big(z^n,\bigcup_{k'\neq k}\hat{\mathcal{B}}_{k'}\Big). \tag{12}$$

If $\chi_{ijk}=1$, then $(w_iu_{ij}v_{ikl})\in\mathcal{T}_{WUV}^n(\eta/2)$ and $y_{ijk}\in\mathcal{T}_Y(w_iu_{ij}v_{ikl},\eta)$ for some l. Therefore, recalling the definition of $\hat{\mathcal{A}}_{ij}$ and $\hat{\mathcal{B}}_{ikl}$, it easily follows from the law of large numbers and from the relation $((W,U,V),Y,X,Z)\in\mathcal{P}(F,G)$ that for large enough n,

$$\chi_{ijk}[1-F^n(\hat{\mathcal{A}}_{ij}|y_{ijk})]<\epsilon \tag{13}$$

and, similarly,

$$\chi_{ijk}[1-G^n(\hat{\mathcal{B}}_{ik}|y_{ijk})]\leqslant\chi_{ijk}[1-G^n(\hat{\mathcal{B}}_{ikl}|y_{ijk})]<\epsilon. \tag{14}$$

Furthermore, we have

$$\begin{aligned}\tau\Big(x^n,\bigcup_{(i'j')\neq(ij)}\hat{\mathcal{A}}_{i'j'}\Big)&\leqslant\sum_{i'\neq i}\tau\Big(x^n,\bigcup_{j'}\hat{\mathcal{A}}_{i'j'}\Big)+\sum_{j'\neq j}\tau(x^n,\hat{\mathcal{A}}_{ij'})\\&\leqslant\sum_{i'\neq i}\tau(x^n,\hat{\mathcal{A}}_{i'})+\sum_{j'\neq j}\tau(x^n,\hat{\mathcal{A}}_{ij'})\end{aligned} \tag{15}$$

(where $\hat{\mathcal{A}}_{i'}=\mathcal{T}_X(w_{i'},2\eta)$), and

$$\begin{aligned}\tau\Big(z^n,\bigcup_{k'\neq k}\hat{\mathcal{B}}_{k'}\Big)&=\tau\Big(z^n,\bigcup_{i',k'\neq k,l'}\hat{\mathcal{B}}_{i'k'l'}\Big)\\&\leqslant\sum_{i'\neq i,k',l'}\tau(z^n,\hat{\mathcal{B}}_{i'k'l'})+\sum_{k'\neq k,l'}\tau(z^n,\hat{\mathcal{B}}_{ik'l'}).\end{aligned} \tag{16}$$

The rule for selecting the auxiliary code provides an estimate of the conditional expectations of the r.v.'s $\tau(X^n,\hat{\mathcal{A}}_{i'})$, $\tau(X^n,\hat{\mathcal{A}}_{ij'})$, $\tau(Z^n,\hat{\mathcal{B}}_{i'k'l'})$, and $\tau(Z^n,\hat{\mathcal{B}}_{ik'l'})$, given the values of the auxiliary codewords W_i, U_{ij}, V_{ik1}, $V_{ik2},\cdots,V_{ikL}$. For $X^n\in\hat{\mathcal{A}}_{ij}$, $Z^n\in\hat{\mathcal{B}}_{ik}$, sufficiently small

[1] In (4)–(7) $[M]$ denotes the integer part of the number M.

η, and large enough n, we have

$$E\{\tau(x^n,\hat{\mathcal{Q}}_{i'})|w_i,u_{ij},v_{ik1},\cdots,v_{ikL}\} \leqslant \exp[-n(I(W\wedge X)-\delta/2)], \quad \text{for } i'\neq i \quad (17)$$

$$E\{\tau(x^n,\hat{\mathcal{Q}}_{ij'})|w_i,u_{ij},v_{ik1},\cdots,v_{ikL}\} \leqslant \exp[-n(I(U\wedge X|W)-\delta/2)], \quad \text{for } j'\neq j \quad (18)$$

$$E\{\tau(z^n,\mathcal{B}_{i'k'l'})|w_i,u_{ij},v_{ik1},\cdots,v_{ikL}\} \leqslant \exp[-n(I(WV\wedge Z)-\delta/2)], \quad \text{for } i'\neq i \text{ and any } k',l' \quad (19)$$

$$E\{\tau(z^n,\hat{\mathcal{B}}_{ik'l'})|w_i,u_{ij},v_{ik1},\cdots,v_{ikL}\} \leqslant \exp[-n(I(V\wedge Z|W)-\delta/2)], \quad \text{for } k'\neq k \text{ and any } l'. \quad (20)$$

Substituting (13) and (15) into (11), taking mathematical expectation and using (17) and (18), we get

$$E\{\chi_{ijk}[1-F^n(\mathcal{Q}_{ij}|y_{ijk})]\} \leqslant \epsilon + I\exp[-n(I(W\wedge X)-\delta/2)] + J\exp[-n(I(U\wedge X|W)-\delta/2)]. \quad (21)$$

Similarly, from (14), (16), (12), (19), and (20) we get

$$E\{\chi_{ijk}[1-G^n(\mathcal{B}_k|y_{ijk})]\} \leqslant \epsilon + IKL\exp[-n(I(WV\wedge Z)-\delta/2)] + KL\exp[-n(I(V\wedge Z|W)-\delta/2)]. \quad (22)$$

Using (4), (5), (8), (9), and (10) it follows from (21) and (22) that for all i,j,k and for large enough n,

$$E\{2-F^n(\mathcal{Q}_{ij}|y_{ijk})-G^n(\mathcal{B}_k|y_{ijk})\}<8\epsilon.$$

This implies the existence of an $(n,8\epsilon)$-code with rate pair $(R_x-2\delta,R_z-2\delta)$ for any $\epsilon,\delta>0$. The theorem is proved.

Acknowledgment

The author is indebted to J. Körner, who agreed to publish Theorem 5 in this paper.

Appendix I

Proof of Theorem 5

Let the r.v. Y^n be uniformly distributed over the codewords $\{y_{jk}: 1\leqslant j\leqslant J, 1\leqslant k\leqslant K\}$ of an (n,ϵ)-code for (F,G), and denote by V^* the r.v. taking the value k if $Y^n=y_{jk}$. By Fano's lemma

$$n^{-1}\log J=n^{-1}H(Y^n|V^*)<n^{-1}I(Y^n\wedge X^n|V^*)+\text{const.}\,\epsilon,$$
$$n^{-1}\log K=n^{-1}H(V^*)<n^{-1}I(V^*\wedge Z^n)+\text{const.}\,\epsilon;$$

where X^n and Z^n are the n-length outputs of the discrete memoryless channels F and G, respectively, corresponding to the input Y^n. In order to prove (3), it suffices to show that for some $(V,Y,X,Z)\in\mathcal{P}(F,G)$ we have

$$n^{-1}I(Y^n\wedge X^n|V^*)\leqslant I(Y\wedge X) \quad (23)$$

$$n^{-1}I(V^*\wedge Z^n)\leqslant I(V\wedge Z) \quad (24)$$

and

$$n^{-1}[I(Y^n\wedge X^n|V^*)+I(V^*\wedge Z^n)] \leqslant I(Y\wedge X|V)+I(V\wedge Z) \quad (25)$$

provided that Y^n is any r.v. with values in $\mathcal{Y}^n$, X^n and Z^n are the length n outputs of channels F and G, respectively, corresponding to the input Y^n, and V^*,Y^n and (X^n,Z^n) form a Markov chain.

It is clear that

$$n^{-1}I(Y^n\wedge X^n|V^*)\leqslant n^{-1}I(Y^n\wedge X^n) \leqslant n^{-1}\sum_{i=1}^{n}I(Y_i\wedge X_i) \quad (25)$$

$$n^{-1}I(V^*\wedge Z^n)\leqslant n^{-1}\sum_{i=1}^{n}I(V^*X^{i-1}Z_i^n\wedge Z_i) = n^{-1}\sum_{i=1}^{n}I(V_i\wedge Z_i) \quad (26)$$

where $X^{i-1}\triangleq X_1X_2\cdots X_{i-1}$, $Z_i^n=Z_{i+1}\cdots Z_n$, $V_i=V^*X^{i-1}Z_i^n$, $1\leqslant i\leqslant n$. Moreover,

$$\begin{aligned}&I(V^*\wedge Z^n)-I(V^*\wedge X^n)\\&=[I(V^*\wedge Z^n)-I(V^*\wedge X_1Z_1^n)]\\&\quad+[I(V^*\wedge X_1Z_1^n)-I(V^*\wedge X_1X_2Z_2^n)]\\&\quad+\cdots+[I(V^*\wedge X^{n-1}Z_n)-I(V^*\wedge X^n)]\\&=\sum_{i=1}^{n}[I(V^*\wedge Z_i|X^{i-1}Z_i^n)-I(V^*\wedge X_i|X^{i-1}Z_i^n)].\end{aligned} \quad (27)$$

Applying (27) to Y^n instead of V^* yields

$$I(Y^n\wedge Z^n)-I(Y^n\wedge X^n) =\sum_{i=1}^{n}[I(Y_i\wedge Z_i|X^{i-1}Z_i^n)-I(Y_i\wedge X_i|X^{i-1}Z_i^n)], \quad (27')$$

which is equivalent to

$$\sum_{i=1}^{n}I(Z_i^n\wedge X_i|X^{i-1})=\sum_{i=1}^{n}I(X^{i-1}\wedge Z_i|Z_i^n). \quad (28)$$

We use (27) and (28) to overbound the left-hand side of (24):

$$\begin{aligned}&n^{-1}[I(Y^n\wedge X^n|V^*)+I(V^*\wedge Z^n)]\\&=n^{-1}[I(Y^n\wedge X^n)+I(V^*\wedge Z^n)-I(V^*\wedge X^n)]\\&=n^{-1}\sum_{i=1}^{n}[I(Y_i\wedge X_i|X^{-1}Z_i^n)+I(Z_i^n\wedge X_i|X^{i-1})\\&\quad+I(V^*\wedge Z_i|X^{i-1}Z_i^n)-I(V^*\wedge X_i|X^{i-1}Z_i^n)]\\&=n^{-1}\sum_{i=1}^{n}[I(Y_i\wedge X_i|V_i)+I(X^{i-1}\wedge Z_i|Z_i^n)\\&\quad+I(V^*\wedge Z_i|X^{i-1}Z_i^n)]\\&\leqslant n^{-1}\sum_{i=1}^{n}[I(Y_i\wedge X_i|V_i)+I(V_i\wedge Z_i)].\end{aligned} \quad (29)$$

The third equality in (29) follows from

$$(V_i,Y_i,X_i,Z_i)\in\mathcal{P}(F,G), \qquad i=1,2,\cdots,n \quad (30)$$

which can be easily verified.

Equations (25), (26), and (29) together with (30) imply (23)–(25) if we define

$$V=(I,V_I),\ Y=Y_I,\ X=X_I,\ Z=Z_I,$$

where I is an r.v. uniformly distributed over the set $\{1,2,\cdots,n\}$ and independent of (V^*,Y^n,X^n,Z^n).

The fact that the region in (1) does not decrease if V is allowed to take at most $\|\mathcal{Y}\|+2$ values can be seen using [5, lemma 3].

APPENDIX II

Let (F,G) be the Blackwell channel, i.e., $\mathcal{Y}=\{0,1,2\}$, $\mathcal{X}=\mathcal{Z}=\{0,1\}$, and

$$F=\begin{pmatrix}1&0\\0&1\\0&1\end{pmatrix} \qquad G=\begin{pmatrix}0&1\\1&0\\0&1\end{pmatrix}.$$

Proposition:

$$\max\{R_x+R_z:(R_x,R_z)\in\hat{\mathcal{R}}\}<\log 3 = \max\{R_x+R_z:(R_x,R_z)\in\mathcal{R}\}. \quad (31)$$

Proof: From Theorem 3 it follows that the right-most side of (31) equals

$$\max\{H(X,Z):(Y,X,Z)\in\mathcal{P}(F,G)\}=\log 3$$

and the maximum is achieved if and only if the joint distribution of the pair (X,Z) is

$$\Pr\{X=1,Z=1\}=\Pr\{X=1,Z=0\}=\Pr\{X=0,Z=1\}=\tfrac{1}{3}. \quad (32)$$

Since in [12] size constraints are proved for the cardinality of the auxiliary r.v.'s W, U, V figuring in the definition of $\hat{\mathcal{R}}$, it is enough to prove that for $((U,V,W),Y,X,Z)\in\mathcal{P}(F,G)$ such that W, U, and V are independent, the inequality

$$\min\{I(W\wedge X),I(W\wedge Z)\}+I(U\wedge X|W)+I(V\wedge Z|W)<\log 3 \quad (33)$$

holds.

It can be easily seen that

$$\begin{aligned}&I(U\wedge X|W)+I(V\wedge Z|W)\\&\leqslant I(U\wedge X|VW)+I(V\wedge Z|W)\\&\leqslant I(UV\wedge XZ|W)\leqslant H(XZ|W)\end{aligned} \quad (34)$$

and it is obvious that

$$\min\{I(W\wedge X),I(W\wedge Z)\}\leqslant I(W\wedge XZ) \quad (35)$$

so the left-hand side of (33) can be upper-bounded by

$$I(W\wedge XZ)+H(XZ|W)=H(XZ)\leqslant\log 3. \quad (36)$$

Suppose that in (33) we have equality instead of strict inequality. Then the following conditions must be satisfied:

$$\text{the pair } (X,Z) \text{ has distribution (32)} \quad (37)$$

$$I(W\wedge X)=I(W\wedge Z)=I(W\wedge XZ) \quad (38)$$

$$I(U\wedge X|VW)+I(V\wedge Z|W)=I(UV\wedge XZ|W)=H(XZ|W) \quad (39)$$

c.f., (36), (35), and (34). From (37) it follows that the distribution of (X,Z) is indecomposable (see [19]). Therefore, (38) implies that W is independent of (X,Z) (see [20]). From this it follows that we can restrict attention to the case $W=\text{const}$. Then (39) implies that

$$(X,Z) \text{ is a deterministic function of } (U,V) \quad (40)$$

$$I(V\wedge X|Z)=I(U\wedge Z|XV)=0. \quad (41)$$

Equations (40) and (41) imply that $H(Z|XV)\leqslant H(Z|UV)=0$, which together with (32) means that we also have $H(Z|V)=0$. Similarly $H(X|U)=0$ and consequently $I(U\wedge V)\geqslant I(X\wedge Z)>0$, which contradicts the assumption that U and V are independent.

REFERENCES

[1] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.

[2] J. Körner and K. Marton, "Comparison of two noisy channels," *Colloquia Mathematica Societatis János Bolyai, 16, Topics in Information Theory*, I. Csiszár and P. Elias, Eds. Amsterdam: North-Holland, 1977, pp. 411–423.

[3] A. El Gamal, "The capacity of a class of broadcast channels," Dep. of Statistics, Standford Univ., Stanford, CA, Tech. Rep. 24, 1978.

[4] R. G. Gallager, "Coding for degraded broadcast channels," *Problemy Peredachi Informatsii*, vol. X, pp. 3–14, Sept. 1974, (in Russian).

[5] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 629–37, Nov. 1975.

[6] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 60–64, Jan. 1977.

[7] M. Pinsker, "The capacity region of noiseless broadcast channels," *Problemy Peredachi Informatsii*, vol. XIV, pp. 28–34, 1978 (in Russian).

[8] S. Gelfand, "The capacity region of a broadcast channel," *Problemy Peredachi Informatsii*, vol. XIII, pp. 106–108, 1977 (in Russian).

[9] K. Marton, "The capacity region of deterministic broadcast channels," to appear in *Trans. Int. Symp. on Information Theory*, Paris-Cachan, 1977.

[10] E. van der Meulen, "Random coding theorems for the general discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 180–190, Mar. 1975.

[11] T. Cover, "An achievable rate region for the broadcast channel, *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 399–404, July 1975.

[12] B. E. Hajek and M. B. Pursley, "Evaluation of an achievable rate region for the broadcast channel," to appear in *IEEE Trans. Inform. Theory*.

[13] P. Bergmans, "Coding theorems for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 197–207, Mar. 1973.

[14] J. Körner and K. Marton, "Images of a set via two channels and their role in multi-user communication," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 751–761, Nov. 1977.

[15] G. Poltirev, "The capacity region of parallel broadcast channels with degraded components," *Problemy Peredachi Informatsii*, vol. XIII, pp. 23–35, 1977 (in Russian).

[16] J. Körner and K. Marton, unpublished result.

[17] S. Gelfand and M. Pinsker, paper to appear in *Problemy Peredachi Informatsii*.

[18] I. Csiszar and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 339–348, May 1978.

[19] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems Contr. Inform.*, vol. 2, pp. 149–162, 1973.

[20] R. Ahlswede and J. Körner "On common information and related characteristics of correlated information sources," Preprint presented at the 7th Prague Conf. Information Theory, Sept. 1974.