Random Coding Theorem for Broadcast Channels With Degraded Components

PATRICK P. BERGMANS

Abstract—**This paper generalizes Cover's results on broadcast channels with two binary symmetric channels (BSC) to the class of degraded channels with N components. A random code, and its associated decoding scheme, is shown to have expected probability of error going to zero for all components simultaneously as the codeword length goes to infinity, if the point representing the rates to the various receivers falls in the set of achievable rates described by this paper. A procedure to expurgate a good random broadcast code is given, leading to a bound on the maximum probability of error.**

Binary symmetric broadcast channels always fall in the class of degraded broadcast channels. The results of the paper are applied to this class of channels of potential practical importance.

I. Introduction

In a recent paper [1], Cover introduced the notion of a broadcast channel, through which one source sends information to two or more receivers. One of the results of Cover's paper is that in some situations there exists a coding scheme allowing the transmission of information to the different users at better rates than the so-called time-sharing rates. The present paper generalizes those results, and gives a rigorous proof of the coding theorem for degraded broadcast channels.

The notion of a degraded broadcast channel is introduced in Section II, together with the definitions of the various quantities used in the paper. In Section III, we exhibit a random coding scheme, describe the associated decoding rule, and find a set of sufficient conditions to guarantee that the expected probability of error goes to zero for all channels simultaneously when the length of the code goes to ∞. The rigorous proof of this statement is given in Appendix A. In Appendix B, we show that uniformly good broadcast codes exist, and that we can upperbound the maximum probability of error for all channels simultaneously. Some of the results of Section III were conjectured by Cover [1]. In Section IV, we introduce simple upper bounds to the capacity region, and discuss a generalized definition of broadcast rates. Finally, the case of the binary symmetric broadcast channel (BSBC) is treated in Section V.

II. Definitions and Preliminaries

A. Broadcast Channels

The most general representation of a discrete memoryless broadcast channel is given in Fig. 1. The input alphabet is

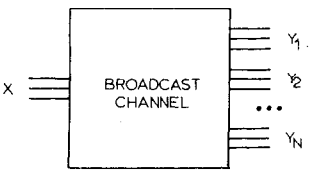

Fig. 1. Broadcast channel.

$\mathcal{A}$, and the output alphabet of the jth terminal is $\mathcal{B}_j$. The transition probability of the broadcast channel is $p(y_1,y_2,\cdots,y_N \mid x)$.

We impose a "no-collaboration" restriction between the receivers connected to the different terminals of the broadcast channel. Without this restriction, there is no real broadcast situation. The no-collaboration restriction allows us to factor the broadcast channel into its N component channels, since possible dependence between the Y_j conditioned on X is irrelevant. Hence, we need consider only the marginal transition probabilities $p(y_1 \mid x)$, $p(y_2 \mid x),\cdots$ of the component channels $A_1,A_2,\cdots,A_N$ (Fig. 2), and all broadcast channels with same marginals will be equivalent in this context.

C_j, the capacity of the jth component channel, is defined in the usual way as the maximum mutual information between the random variables X and Y_j. Without loss of generality, we shall assume that

$$C_1 > C_2 > \cdots > C_N. \tag{1}$$

There are no equalities in (1), but we shall show later that this is not a restriction.

We now wish to use this broadcast channel with N components to transmit the output of N independent sources $S_1,S_2,\cdots,S_N$ to N users, connected to the outputs of the component channels, on a one-to-one basis. By this, we mean that the output of the jth source is intended to be received by the jth receiver (Fig. 3). Subsequent arguments in Section IV will remove this restriction in interpretation.

B. Broadcast Codes

Let I_j be the set of possible outcomes of source S_j and let all the elements of I_j have same probability. This is not a loss of generality in a channel coding theorem, and, in fact, represents the worst case. The length of the code is n.

The size of I_j is M_j, and the rate of source S_j is given by

$$R_j = \frac{1}{n} \log_2 M_j \tag{2}$$

Manuscript received September 13, 1971; revised March 31, 1972. This work was supported by Air Force Contract AF 49(638)1517 and NSF Contract GK-34363.

The author was with Stanford University, Stanford, Calif. He is now with the Department of Electrical Engineering, Cornell University, Ithaca, N.Y.

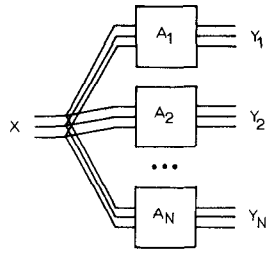

Fig. 2. Considering components only.

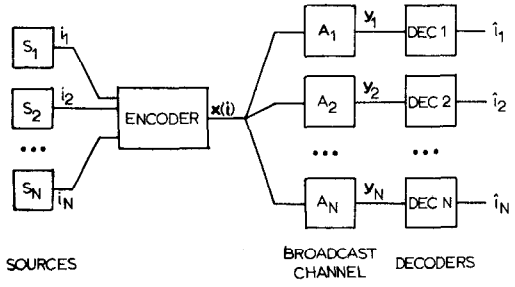

Fig. 3. Broadcast communication system.

For notational convenience, we define

$$\begin{aligned} \boldsymbol{i} &= (i_1,i_2,\cdots,i_N) \\ \boldsymbol{I} &= I_1 \times I_2 \times \cdots \times I_N \\ \boldsymbol{R} &= (R_1,R_2,\cdots,R_N) \\ \boldsymbol{M} &= (M_1,M_2,\cdots,M_N) \\ M &= |\boldsymbol{I}| = \prod_{j=1}^{N} M_j. \end{aligned} \tag{3}$$

A broadcast code consists of an encoding function

$$\boldsymbol{x}: \boldsymbol{I} \to \mathscr{A}^n \tag{4}$$

and N decoding functions

$$g_j: \mathscr{B}_j^n \to I_j. \tag{5}$$

When the source output is $\boldsymbol{i} = (i_1,i_2,\cdots,i_N)$, the jth receiver is in error if $g_j(\boldsymbol{y}_j) \neq i_j$. The probability of this event will be denoted

$$\lambda_j(\boldsymbol{i}) = \Pr\,[g_j(\boldsymbol{y}_j) \neq i_j \mid \boldsymbol{x}(\boldsymbol{i}) \text{ has been sent}]. \tag{6}$$

We introduce the following notation for the maximum and average probability of error

$$\lambda_j = \max_{\boldsymbol{i} \in \boldsymbol{I}} \lambda_j(\boldsymbol{i}) \tag{7}$$

and

$$\mu_j = \frac{1}{M} \sum_{\boldsymbol{i} \in \boldsymbol{I}} \lambda_j(\boldsymbol{i}). \tag{8}$$

Again, for notational convenience, let

$$\begin{aligned} \boldsymbol{\lambda} &= (\lambda_1,\lambda_2,\cdots,\lambda_N) \\ \boldsymbol{\mu} &= (\mu_1,\mu_2,\cdots,\mu_N). \end{aligned} \tag{9}$$

Using Wolfowitz' notation, we define a $(\boldsymbol{M},n,\boldsymbol{\lambda})$-code as a code of length n, size $\boldsymbol{M}$, and maximum probability of error $\boldsymbol{\lambda}$. $\boldsymbol{R}$ is an achievable rate if there exists a sequence of $(\boldsymbol{M},n,\boldsymbol{\lambda}^{(n)})$-codes, with $M_j = \lfloor \exp_2 (nR_j) \rfloor$, such that $\boldsymbol{\lambda}^{(n)} \to \boldsymbol{0}$ when $n \to \infty$. The set of achievable rates, or capacity region of the broadcast channel, will be denoted $\mathscr{C}$.

C. Time Sharing

Let $\{\tau_j\}$ be a set of numbers such that

$$\tau_j \geq 0 \quad \text{and} \quad \sum_{j=1}^{N} \tau_j = 1. \tag{10}$$

Clearly, the rate point

$$\boldsymbol{R} = (\tau_1 C_1,\tau_2 C_2,\cdots,\tau_N C_N) \tag{11}$$

is an achievable rate, by simple time sharing. Every point dominated by such an $\boldsymbol{R}$ is, of course, also achievable. Hence,

$$\mathscr{R}_{TS} = \{(R_1,\cdots,R_N): R_j \leq \tau_j C_j \; \forall j, \quad \text{for some } \{\tau_j\}\} \tag{12}$$

is the set of rates achievable by time sharing (Fig. 4). The boundary of $\mathscr{R}_{TS}$ is a hyperplane intersecting the R_j axis at C_j.

D. Degraded Channels

We shall say that a channel A_2 is a degraded version of a channel A_1 if there exists a third channel D_2 such that A_2 can be represented as the cascade of A_1 and D_2. Specifically, let A_1 be a channel with input alphabet $\mathscr{A}$, output alphabet $\mathscr{B}_1$, and transition probability $p_1(y_1 \mid x)$, and let A_2 be another channel with same input alphabet $\mathscr{A}$, output alphabet $\mathscr{B}_2$, and transition probability $p_2(y_2 \mid x)$. The degradation is expressed by (Fig. 5)

$$p_2(y_2 \mid x) = \sum_{y_1 \in \mathscr{B}_1} p_3(y_2 \mid y_1) p_1(y_1 \mid x), \tag{13}$$

where $p_3(y_2 \mid y_1)$ is the transition probability of the degrading channel D_2, with input alphabet $\mathscr{B}_1$ and output alphabet $\mathscr{B}_2$.

This can be considered as "post-degradation". It is a special case of channel inclusion described by Shannon in [2], where "predegradation" is also allowed. Shannon shows that $C_2 \leq C_1$, where C_i is the capacity of A_i.

By definition, if every component channel A_j of a broadcast channel is a degraded version of A_{j-1} $(j = 2,\cdots,N)$, the broadcast channel will be called cascade degraded, or, in this paper, simply degraded (although we can imagine other lattices of degraded channels).

Applying the no-collaboration restriction again, we can represent a degraded broadcast channel as a cascade formed by the best channel A_1, followed by successive degrading channels $D_2,D_3,\cdots,D_N$ (Fig. 6).

In the remainder of this paper, we shall consider only degraded broadcast channels. It will not be necessary that the structure of the channel be actually degraded, as long as there is an equivalent degraded broadcast channel (i.e., a broadcast channel with the same marginal transition prob-

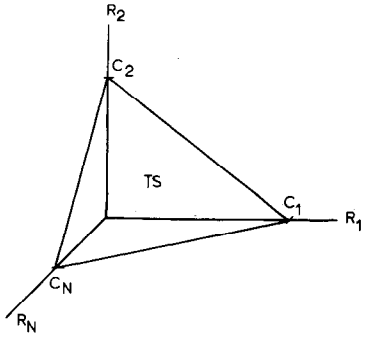

Fig. 4. Set of rates achievable by time-sharing.

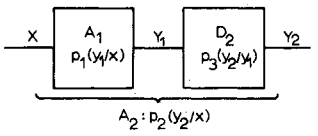

Fig. 5. A_2 is a degraded version of A_1.

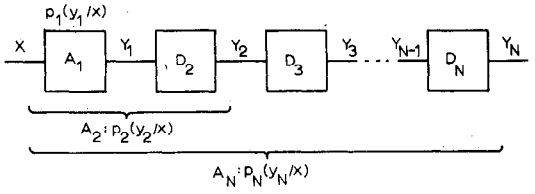

Fig. 6. Degraded broadcast channel.

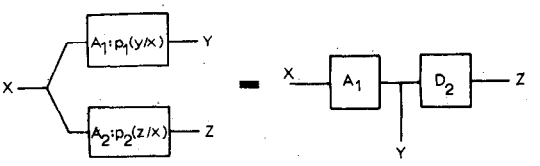

Fig. 7. Degraded broadcast channel with two components.

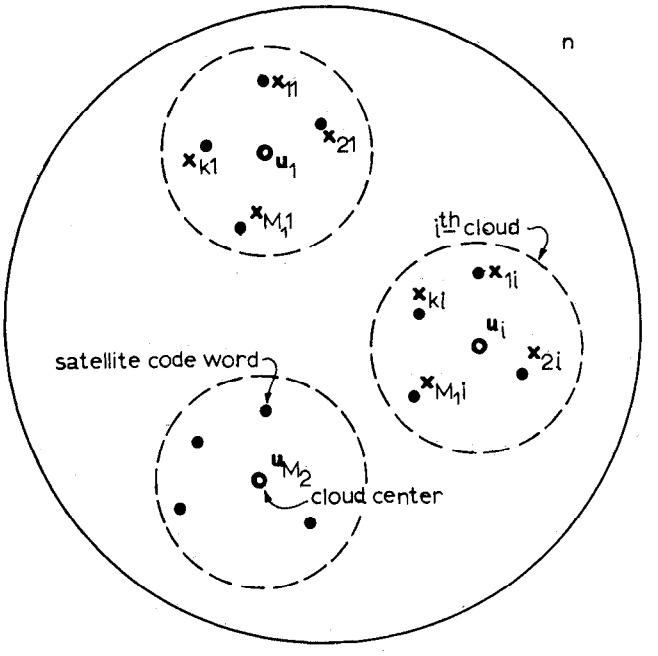

Fig. 8. Clouds and cloud centers of a broadcast code.

abilities). Many practical broadcast channels are degraded (an example is given in Section V), and the results obtained in the next section are widely applicable.

III. Coding Theorem for Degraded Broadcast Channels

A. Broadcast Channel With Two Components

The essence of the coding scheme for degraded broadcast channels can be most easily understood when applied to a broadcast situation with two sources.

We shall not use the definitions of Section II in their full generality, to avoid needlessly complicated notation. Instead, let k be the index emitted by source S_1, and i be the index emitted by source S_2. We have

$$1 \leq k \leq M_1 \text{ and } 1 \leq i \leq M_2. \tag{14}$$

Codeword x_{ki} will be used to transmit this joint message over the degraded channel represented in Fig. 7. User 1 should decode k correctly, while user 2 should decode i correctly.

The "cloud" of codewords x_{ki} in $\mathscr{A}^n$ with same index i will be represented by a vector u_i in $\mathscr{A}^n$, henceforth called the cloud center (Fig. 8). The meaning of this cloud center will be made clear later. The different x_{ki} in a given cloud will be called the satellite codewords of u_i. Since user 2 should decode correctly the index i of S_2, it is sufficient for him to determine the cloud to which the transmitted codeword x_{ki} belongs, or, in other words, its representative u_i.

In every cloud, the satellites are indexed by the index k of S_1. Hence, user 1 should only recover the index k of the satellite. To achieve this, we shall show that we can use u_i, because it is also available to user 1. However, knowledge of the index i of source S_2 does not give user 1 any information about source S_1. Hence, the rate R_2 as defined in (2) should depend only on M_2.

We now exhibit a random coding scheme for which the expected value of the average probability of error goes to zero for both channels simultaneously, as the block length n is allowed to increase without bound.

We shall need to consider the mutual information between some of the random variables U, X, Y, and Z, as illustrated in Fig. 9. $q_2(u)$ and $q_1(x \mid u)$, defined for $u,x \in \mathscr{A}$, are parameters that will be used in the random coding scheme. $q_1(x \mid u)$ can be thought of as the transition probability of an artificial channel E. $p_1(y \mid x)$ and $p_2(z \mid x)$ are the channel transition probabilities of channels A_1 and A_2. The various mutual informations are functions of the parameters q_1 and q_2.

Choose at random $M_2 = 2^{nR_2}$ cloud centers in $\mathscr{A}^n$, with letters independently drawn according to $q_2(u)$. To each cloud center, we append $M_1 = 2^{nR_1}$ satellites in $\mathscr{A}^n$, independently drawn according to $q_1(x \mid u)$, conditioned on the cloud center u. The effect is that of "passing" each

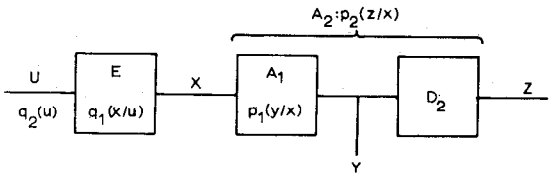

Fig. 9. Introduction of the artificial channel E.

cloud center vector M_1 times through the artificial channel E with transition probability $q_1(x \mid u)$.

In the following, we shall use an overbar to denote the expectation over the ensemble of codes generated according to preceding procedure.

The decoding sets are determined as follows. Let

$$I_{q_1q_2}(u; z) \triangleq \frac{1}{n} \log_2 \frac{p(z \mid u)}{p(z)}, \tag{15}$$

where

$$p(z \mid u) = \sum_{x \in \mathscr{A}^n} p_2(z \mid x) q_1(x \mid u)$$

and

$$p(z) = \sum_{u \in \mathscr{A}^n} p(z \mid u) q_2(u).$$

Then, by definition [3]

$$E[I_{q_1q_2}(u; z)] \triangleq I_{q_1q_2}(U; Z). \tag{16}$$

Let

$$S(z) = \left\{ u \in \mathscr{A}^n : I_{q_1q_2}(u; z) > \frac{R_2 + I_{q_1q_2}(U; Z)}{2} \right\} \tag{17}$$

be the decoding set for z, and define

$$d(u,z) \triangleq \begin{cases} 1, & u \notin S(z) \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

When x_{ki} has been sent, and z has been received, we shall say that the cloud center detector makes an error if either u_i is not in $S(z)$ (error of the first type, occurring with probability $P_e^{(1)}(k,i)$) or if u_j is in $S(z)$ for $j \neq i$ (error of the second type, occurring with probability $P_e^{(2)}(k,i)$).

Hence

$$\lambda_2(k,i) \leq P_e^{(1)}(k,i) + P_e^{(2)}(k,i) \tag{19}$$

with

$$P_e^{(1)}(k,i) = \sum_{z \in \mathscr{C}^n} p_2(z \mid x_{ki})\, d(u_i,z) \tag{20}$$

and

$$P_e^{(2)}(k,i) = \sum_{z \in \mathscr{C}^n} p_2(z \mid x_{ki}) \sum_{\substack{j=1 \\ j \neq i}}^{M_2} (1 - d(u_j,z)). \tag{21}$$

With a uniform distribution in the independent indices k and i, we define

$$P_e^{(m)} = \frac{1}{M_1 M_2} \sum_{k=1}^{M_1} \sum_{i=1}^{M_2} P_e^{(m)}(k,i), \qquad m = 1,2. \tag{22}$$

Consequently, we have

$$\mu_2 \leq P_e^{(1)} + P_e^{(2)}$$

and

$$\overline{\mu_2} \leq \overline{P_e^{(1)}} + \overline{P_e^{(2)}}. \tag{23}$$

Quite obviously, the better receiver can detect the cloud center with a probability of error no greater than the probability of error for the more degraded channel receiver.

A trivial way to achieve an equal probability of error is to append an additional degrading channel to obtain the input–output statistics of the degraded channel and to use the decoding rule for receiver 2. In addition, user 1 will now attempt to recover the satellite codeword itself. Let

$$I_{q_1q_2}(x; y \mid u) = \frac{1}{n} \log_2 \frac{p(y \mid x, u)}{p(y \mid u)} = \frac{1}{n} \log_2 \frac{p_1(y \mid x)}{p(y \mid u)}, \tag{24}$$

where

$$p(y \mid u) = \sum_{x \in \mathscr{A}^n} p_1(y \mid x) q_1(x \mid u).$$

Then, in analogy with the first part of the decoding procedure, we have

$$E[I_{q_1q_2}(x; y \mid u)] = I_{q_1q_2}(X; Y \mid U). \tag{25}$$

Let

$$T(y \mid u) \triangleq \left\{ x \in \mathscr{A}^n : I_{q_1q_2}(x; y \mid u) > \frac{R_1 + I_{q_1q_2}(X; Y \mid U)}{2} \right\} \tag{26}$$

be the conditional satellite decoding set for y, and define

$$d(x, y \mid u) \triangleq \begin{cases} 1, & x \notin T(y \mid u) \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

The satellite will now be detected as follows. The receiver first detects the cloud center u_i making errors of the first or the second type (see above). Even if the cloud center has been decoded correctly, it is still possible that x_{ki} is not in $T(y \mid u_i)$ (error of the third type, with probability $P_e^{(3)}(k,i)$), or that there is some x_{li} in $T(y \mid u_i)$, for $l \neq k$ (error of the fourth type, with probability $P_e^{(4)}(k,i)$).

We have

$$\lambda_1(k,i) \leq \lambda_2(k,i) + P_e^{(3)}(k,i) + P_e^{(4)}(k,i) \tag{28}$$

with

$$P_e^{(3)}(k,i) = \sum_{y \in \mathscr{B}^n} p_1(y \mid x_{ki})\, d(x_{ki}, y \mid u_i) \tag{29}$$

and

$$P_e^{(4)}(k,i) = \sum_{y \in \mathscr{B}^n} p_1(y \mid x_{ki}) \sum_{\substack{l=1 \\ l \neq k}}^{M_1} (1 - d(x_{li}, y \mid u_i)). \tag{30}$$

Again, with a uniform distribution on k and i, we have

$$P_e^{(m)} = \frac{1}{M_1 M_2} \sum_{k=1}^{M_1} \sum_{i=1}^{M_2} P_e^{(m)}(k,i), \qquad m = 3,4 \quad (31)$$

and

$$\mu_1 \leq \mu_2 + P_e^{(3)} + P_e^{(4)}$$
$$\overline{\mu_1} \leq \overline{\mu_2} + \overline{P_e^{(3)}} + \overline{P_e^{(4)}}. \quad (32)$$

In Appendix A, we prove rigorously that, if

$$R_1 = I_{q_1 q_2}(U;Z) - 2\varepsilon$$
$$R_2 = I_{q_1 q_2}(X; Y \mid U) - 2\varepsilon \quad (33)$$

$\overline{\mu_1}$ and $\overline{\mu_2}$ go to zero simultaneously as $n \to \infty$.

The proof is similar to Shannon's original random coding proof [4]. The main difference, of course, is the introduction of the artificial "satellizing" channel E.

If we consider the transmission of the "cloud message" to receiver 2 (through the worst channel), it is clear that a message is not transmitted by a single codeword, but by a codeword chosen at random (at transmission time) in the cloud. This can only be done at the expense of the maximum rate at which reliable transmission can be achieved. The purpose of the introduction of the satellizing channel E is to make it possible to consider this extra randomization as being part of the channel, and, by doing so, to find the maximum rate just described as the capacity of the cascade of E and A_2. The first part of Appendix A is the verification that this can be done, and that channel E asymptotically represents the perturbation introduced by the superposition of the output of source S_1.

The second part of Appendix A is concerned with the information that can be sent to receiver 1, under the constraint that the codeword to be used must lie in a cloud determined by the message to receiver 2. The decoding method involves the notion of a conditional decoding set. The conditional independence of the satellite codewords is essential to the second part of the proof.

B. Broadcast Situation With N Sources

An obvious generalization of the previous argument yields a random-coding scheme for a degraded broadcast channel with N components.

First, choose $M_N = 2^{nR_N}$ cloud centers in $\mathcal{A}^n$ according to $q_N(x_N)$. Then, select $M_{N-1} = 2^{nR_{N-1}}$ satellites per cloud center, according to $q_{N-1}(x_{N-1} \mid x_N)$, $M_{N-2} = 2^{nR_{N-2}}$ subsatellites per satellite in each cloud, according to $q_{N-2}(x_{N-2} \mid x_{N-1})$, and so forth, until

$$M = \prod_{j=1}^{N} M_j$$

codewords have been selected. At each level, the satellization process can be represented as the result of "passing" the n-vectors generated so far (not yet codewords) M_i times through an artificial channel with transition probability $q_i(x_i \mid x_{i+1})$. This is illustrated in Fig. 10. The artificial satellizing channels are cascaded, and the broadcast channel is of the degraded type.

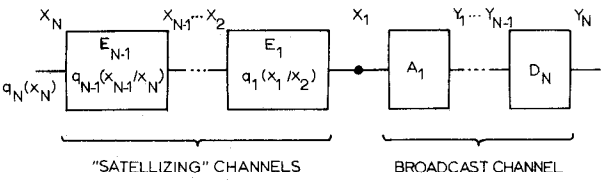

Fig. 10. Introduction of $N-1$ artificial channels.

If, for some $q_1, q_2, \cdots, q_N$, we have

$$R_N < I_{q_1 \cdots q_N}(X_N; Y_N) - \varepsilon$$
$$R_{N-1} < I_{q_1 \cdots q_N}(X_{N-1}; Y_{N-1} \mid X_N) - \varepsilon$$
$$\cdots\cdots\cdots$$
$$R_2 < I_{q_1 \cdots q_N}(X_2; Y_2 \mid X_3) - \varepsilon$$
$$R_1 < I_{q_1 \cdots q_N}(X_1; Y_1 \mid X_2) - \varepsilon \quad (34)$$

$\bar{\mu}$ goes to zero as $n \to \infty$.

So far, we have proved that with certain conditions on $\boldsymbol{R}$, there exists a random coding scheme for which the expected average probability of error $\bar{\mu}$ goes to zero as $n \to \infty$. In Appendix B, we prove that, with identical conditions on $\boldsymbol{R}$, there exists a sequence of codes for which the maximum probability of error λ goes to zero as $n \to \infty$. Hence, the set $\mathcal{I}$ of all $\boldsymbol{R}$ described parametrically by $q_1, \cdots, q_N$ in (34) is completely contained in the capacity region. Special conditions on $q_1, \cdots, q_N$ will yield points on the boundary of $\mathcal{I}$. In particular, if E_k is a useless channel ($q_k(x_k \mid x_{k-1})$ is only a function $q_k(x_k)$ of x_k), and if $E_1, \cdots, E_{k-1}$ are perfect channels, we have

$$R_k \leq I(X_k; Y_k \mid X_{k+1}) = I(X_k; Y_k) = I(X_1; Y_k). \quad (35)$$

Letting $q_k(x_k)$ be the probability mass function maximizing the information between the input of the channel and the kth terminal, we can achieve the rate point $\boldsymbol{R} = (0, \cdots, 0, C_k, 0, \cdots, 0)$.

If $\mathcal{R}$ is the convex hull of $\mathcal{I}$, we can also achieve any rate point in $\mathcal{R}$ by time sharing between some points in $\mathcal{I}$ (Fig. 11). Consequently, $\mathcal{R}_{TS}$ is completely contained in $\mathcal{R}$.

In view of the results of all examples (some of them are presented in Section V), we have been led to believe that $\mathcal{R}$ might very well coincide with $\mathcal{I}$ in all cases. In other words, we do not need "explicit" time sharing to achieve all the points in $\mathcal{R}$.

Finally, at the time of first writing of this paper, we had conjectured that the set $\mathcal{R}$ described by (34) is the capacity region $\mathcal{C}$, i.e., that no rate point outside $\mathcal{R}$ can be achieved with arbitrarily small probability of error. Wolfowitz has stated that he has a proof of this fact [5].

IV. Bounds on $\mathcal{R}$; Generalized Definition of Broadcast Rates

A. Bounds on $\mathcal{R}$

An obvious bound on the set $\mathcal{I}$ described in (34) is the set

$$\mathcal{I}_u = \{(I(X_1; Y_1), \cdots, I(X_1; Y_N)) : p(x_1) \geq 0, \sum p(x_1) = 1\}. \quad (36)$$

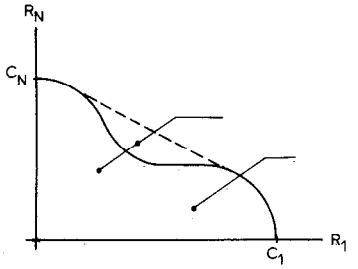

Fig. 11. Set of achievable rates.

Indeed, we have

$$\begin{aligned} I(X_k; Y_k \mid X_{k+1}) &= H(Y_k \mid X_{k+1}) - H(Y_k \mid X_k) \\ &\le H(Y_k) - H(Y_k \mid X_k) \\ &= I(X_k; Y_k) \le I(X_1; Y_k). \end{aligned} \tag{37}$$

Consequently, the convex hull $\mathscr{R}$ of $\mathscr{I}$ is contained in the convex hull $\mathscr{R}_u$ of $\mathscr{I}_u$. $\mathscr{R}_u$ is the generalization to several receivers of the upper bound introduced by Cover in [1]. Considering the inequalities in (37), it should be clear that for a degraded broadcast channel, the only possible common points of the boundaries of $\mathscr{R}$ and $\mathscr{R}_u$ are rate points of the form $\boldsymbol{R} = (0,\cdots,0,C_k,0,\cdots,0)$. A trivial consequence of (37) is $R_k \le C_k$.

Because the users are not allowed to collaborate in the decoding process, we may represent the broadcast channel by a cascade of successive degradation, as mentioned before (Fig. 12). The total throughput rate of the channel is $R_1 + R_2 + \cdots + R_N$. Since information is transmitted reliably, this total rate should not exceed the capacity of the channel in its cascade representation. An upper bound to this capacity is the capacity C_1 of channel A_1. Hence

$$\sum_{i=1}^{N} R_i \le C_1. \tag{38}$$

Another way to show this is to note that

$$\begin{aligned} C_1 &\ge I_{q_1\cdots q_N}(X_1; Y_1) = I_{q_1\cdots q_N}(X_N,X_{N-1},\cdots,X_1; Y_1) \\ &= I_{q_1\cdots q_N}(X_N; Y_1) + I_{q_1\cdots q_N}(X_{N-1}; Y_1 \mid X_N) + \cdots \\ &\quad + I_{q_1\cdots q_N}(X_1; Y_1 \mid X_N,X_{N-1},X_{N-2},\cdots,X_2) \\ &= I_{q_1\cdots q_N}(X_N; Y_1) + I_{q_1\cdots q_N}(X_{N-1}; Y_1 \mid X_N) + \cdots \\ &\quad + I_{q_1\cdots q_N}(X_1; Y_1 \mid X_2) \\ &\ge I_{q_1\cdots q_N}(X_N; Y_N) + I_{q_1\cdots q_N}(X_{N-1}; Y_{N-1} \mid X_N) + \cdots \\ &\quad + I_{q_1\cdots q_N}(X_1; Y_1 \mid X_2) \\ &\ge R_N + R_{N-1} + \cdots + R_1. \end{aligned} \tag{39}$$

Equation (38) proves that, for a broadcast channel composed of identical components with capacity C, the coding scheme for degraded broadcast channels, though still applicable, does not dominate time-sharing coding. It does, however, achieve time-sharing rates provided the probability mass function of X_1 maximizes the mutual information between X_1 and any of the Y_j. Indeed

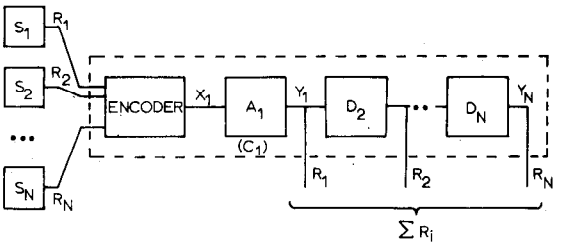

Fig. 12. Total information flow in a degraded broadcast channel.

$$\begin{aligned} \sum_{i=1}^{N} R_i &= \sum_{i=1}^{N-1} I(X_i; Y_i \mid X_{i+1}) + I(X_N; Y_N) \\ &= \sum_{i=1}^{N-1} I(X_i; Y_j \mid X_{i+1}) + I(X_N; Y_j) \\ &= I(X_1,X_2,\cdots,X_N; Y_j) = C. \end{aligned} \tag{40}$$

If not all components of the broadcast channel are equal, the preceding fact can easily be generalized. For example, let

$$C_1 = C_2 < C_3 < C_4 = C_5. \tag{41}$$

Fig. 10 can be simplified to group components with same transition probability. In this case, we have (Fig. 13):

$$\begin{aligned} R_1 + R_2 &< I_{q_1q_3q_4}(X_1; Y_1 \mid X_3) \\ R_3 &< I_{q_1q_3q_4}(X_3; Y_3 \mid X_4) \\ R_4 + R_5 &< I_{q_1q_3q_4}(X_4; Y_4) \end{aligned} \tag{42}$$

with $R_k \ge 0$.

Again, time sharing among equal channels can be achieved by the coding scheme for degraded broadcast channels or by explicit time sharing.

B. Generalized Definition of Broadcast Rates

In the previous sections, we have considered a broadcast situation where different and independent information was sent to the different users. The model for this case was introduced in Section II (Fig. 3). In his original paper, Cover considers a broadcast situation where there is some common information intended for two or more users [1, p. 5]. The N index sets $I_1,I_2,\cdots,I_N$ of the one-to-one case are replaced by $2^N - 1$ index sets I_θ such that

$$\theta = (\theta_1,\theta_2,\cdots,\theta_N) \in \Theta \triangleq \{0,1\}^N - \mathbf{0}. \tag{43}$$

In analogy with Fig. 3, we can represent this case as the simultaneous transmission of the random output of $2^N - 1$ sources to N users, over a broadcast channel with N components.

The output i_θ of source S_θ is intended to be received by all users U_j such that $\theta_j = 1$. The rate R_θ of source S_θ is defined by

$$R_\theta = \frac{1}{n} \log_2 M_\theta, \tag{44}$$

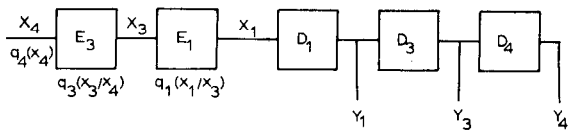

Fig. 13. Simplifying the original broadcast channel.

where M_θ is the size of the output alphabet I_θ of S_θ. The total information rate to user j is the sum of the rates of the sources S_θ such that $\theta_j = 1$. Hence

$$R_j = \sum_{\theta \in \Theta} \theta_j R_\theta \tag{45}$$

The one-to-one case we have considered until now (one source index to a single user) is characterized by the fact that $R_\theta > 0$ if and only if θ has weight one (i.e., θ has only one $\theta_j = 1$).

We now show how the results of the previous sections can be extended to the more general situation described previously for the class of degraded broadcast channels. Given the set $\mathscr{R}$ of achievable rates in the one-to-one case, it is easy to determine whether a particular rate $\boldsymbol{R} = (R_{0\cdots01}, R_{0\cdots10}, \cdots, R_{1\cdots11}) \in R_+^{2^N-1}$ is achievable. Let

$$\begin{aligned} R_N' &= \sum_{\theta \in \Theta} \theta_N R_\theta \\ R_{N-1}' &= \sum_{\theta \in \Theta} \theta_{N-1}\bar{\theta}_N R_\theta \\ R_{N-2}' &= \sum_{\theta \in \Theta} \theta_{N-2}\bar{\theta}_{N-1}\bar{\theta}_N R_\theta \\ &\cdots\cdots\cdots \\ R_1' &= \sum_{\theta \in \Theta} \theta_1\bar{\theta}_2 \cdots \bar{\theta}_{N-1}\bar{\theta}_N R_\theta. \end{aligned} \tag{46}$$

We claim that $\boldsymbol{R} \in R_+^{2^N-1}$ is achievable if the rate point $(R_1', R_2', \cdots, R_N') \in \mathscr{R}$. The sufficient condition just described is also necessary if $\mathscr{R}$ is the capacity region $\mathscr{C}$ for the one-to-one case, as we have conjectured.

A simple proof of this fact can be found in [6] and can be summarized as follows.

User N has to receive the information of all sources such that $\theta_N = 1$. Because of the degraded structure of the channel, the additional information to be sent to user j is the output of the sources intended for user j ($\theta_j = 1$), and which have not yet been encoded for a user with a channel of lower capacity ($\theta_k = 0$, $k > j$). This leads to the expression

$$R_j' = \sum_{\theta \in \Theta} \theta_j\bar{\theta}_{j+1}\bar{\theta}_{j+2} \cdots \bar{\theta}_N R_\theta. \tag{47}$$

A superposition code is then used to encode the information just described, with rates $R_N', R_{N-1}', \cdots, R_1'$, on the successive levels of satellization. The proof of the converse (assuming that $\mathscr{R} = \mathscr{C}$) is by contradiction. It is shown that if it is possible to achieve some rate point $\boldsymbol{R} \in R^{2^N-1}$ such that the rate point defined in (46) does not lie in $\mathscr{C}$, it is also possible to achieve this rate point in a one-to-one communication problem, which is impossible.

In addition to the one-to-one case, there is another particular case of special interest: $R_\theta > 0$ if and only if θ is of the form $(1,1,\cdots,1,0,0,\cdots,0)$. This means that we have only N sources of rate greater than 0, and that the information intended for a given user is also intended for all users with better channels. We shall have a situation like this when we wish to transmit the same information with different degrees of refinement to different users through a degraded broadcast channel. One could conceivably factor the source into subsources representing the different degrees of refinement, with the Nth source representing the coarsest information. User k would then receive arbitrarily well the output of sources $S_{11\cdots10\cdots0}, \cdots, S_{111\cdots11}$ (all sources with $\theta_k = 1$).

In his treatment of the binary symmetric broadcast channel with two components, Cover originally established the set of rates for this interpretation.

V. Binary Symmetric Broadcast Channels (BSBC)

A. Many Components—General Case

There exists a linear ordering on the set of binary symmetric channels (BSC), in the sense that of two BSC's, one channel can always be represented as a degraded version of the other one. Moreover, from symmetry considerations, it should be clear that the artificial channels have also to be symmetric to generate points on the boundary of the set of achievable rates, and that the initial $q_N(x_N)$ should be $(\frac{1}{2},\frac{1}{2})$. This leaves us with $N-1$ scalar parameters to describe the boundary surface in R_+^N.

The BSC resulting from the cascade of two BSC's with parameters p_1 and p_2 has a parameter

$$p_3 = p_1\bar{p}_2 + \bar{p}_1 p_2 \triangleq p_1 * p_2. \tag{48}$$

The results of Sections III and IV are now applicable. For the case shown in Fig. 14, we have

$$\begin{aligned} I(B; C \mid A) &= H(C \mid A) - H(C \mid B, A) \\ &= \mathscr{H}(\alpha * \beta) - \mathscr{H}(\beta), \end{aligned} \tag{49}$$

where $\mathscr{H}(p)$ has the usual definition

$$\mathscr{H}(p) = -p \log_2 p - \bar{p} \log_2 \bar{p}. \tag{50}$$

Referring to the BSBC of Fig. 15, the set of achievable rates has a boundary surface given by

$$\begin{aligned} R_N &= 1 - \mathscr{H}(q_1 * q_2 * \cdots * q_{N-1} * p_N) \\ R_{N-1} &= \mathscr{H}(q_1 * q_2 * \cdots * q_{N-1} * p_{N-1}) \\ &\quad - \mathscr{H}(q_1 * q_2 * \cdots * q_{N-2} * p_{N-1}) \\ &\cdots\cdots\cdots \\ R_2 &= \mathscr{H}(q_1 * q_2 * p_2) - \mathscr{H}(q_1 * p_2) \\ R_1 &= \mathscr{H}(q_1 * p_1) - \mathscr{H}(p_1), \end{aligned} \tag{51}$$

where $q_1, q_2, \cdots, q_{N-1}$ are parameters allowed to vary between 0 and $\frac{1}{2}$.

An alternate notation for the R_j is given by

$$R_j = \mathscr{H}(r_j * p_j) - \mathscr{H}(r_{j-1} * p_j), \tag{52}$$

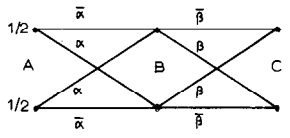

Fig. 14. Cascade of two BSC's.

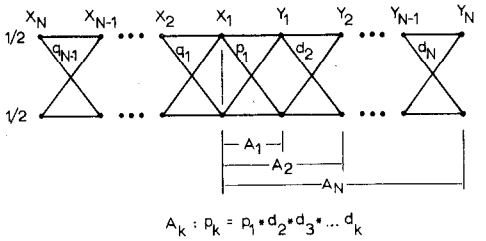

Fig. 15. Binary symmetric broadcast channel with the artificial channels for random coding.

where

$$0 = r_0 \leq r_1 \leq r_2 \leq \cdots \leq r_{N-1} \leq r_N = \tfrac{1}{2}. \quad (53)$$

The two notations are clearly equivalent for corresponding values of the q_j and r_j, i.e., for $r_{j+1} = r_j * q_{j+1}$.

The point $\boldsymbol{R}_i = (0,0,\cdots,0,C_i,0,\cdots,0)$ (transmission at capacity C_i to user i) is achieved for $q_i = \frac{1}{2}$ and $q_j = 0$ for all $j < i$; q_j, for $j > i$ is of no importance. In the case where users with better channels also use the information intended for users with worse channels, the same q yield $\boldsymbol{R} = (C_i,C_i,\cdots,C_i,0,0,\cdots,0)$. This is a "maximin" rate point for the i best channels.

B. BSBC With Two Components

In the case of two BSC's, we find

$$R_2 = 1 - \mathcal{H}(q_1 * p_2)$$
$$R_1 = \mathcal{H}(q_1 * p_1) - \mathcal{H}(p_1). \quad (54)$$

This is Cover's result when $p_1 = 0$, except for the fact that, in the original paper, R_1 is augmented with R_2 as discussed in Section IV.

Some quick calculations show that

$$\left.\frac{dR_2}{dR_1}\right|_{(0,C_2)} = -\frac{(\bar{p}_2 - p_2)\log_2(\bar{p}_2/p_2)}{(\bar{p}_1 - p_1)\log_2(\bar{p}_1/p_1)} \quad (55)$$

$$\left.\frac{dR_2}{dR_1}\right|_{(C_1,0)} = -\frac{(\bar{p}_2 - p_2)^2}{(\bar{p}_1 - p_1)^2}. \quad (56)$$

The slope in (55) will be zero iff p_1 is zero, as shown in [1]. It is possible to show that, for all p_1 and p_2, the slopes in (55) and (56) are such that the boundary of the capacity region dominates the time-sharing line (Fig. 16).

Finally, an accurate (R_1,R_2) curve is given in Fig. 17. The variable E is defined by

$$E = \frac{R_1}{C_1} + \frac{R_2}{C_2} - 1 \quad (57)$$

and is zero on the time-sharing line. The linear upper bound discussed in Section IV is also shown.

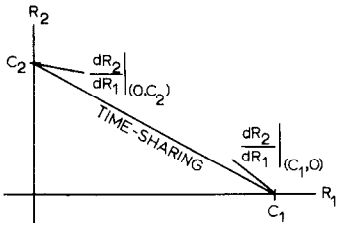

Fig. 16. Slopes of the boundary of the set of achievable rates at the points $(0,C_2)$ and $(C_1,0)$.

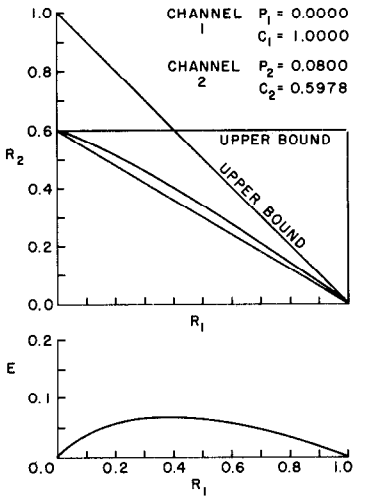

Fig. 17. Set of achievable rates for $p_1 = 0.00$ and $p_2 = 0.08$.

VI. Conclusions

In this paper, we have described a (random) coding scheme and the associated decoding method for the class of degraded broadcast channels. Sufficient conditions for arbitrarily reliable simultaneous communication have been found, which, if proven necessary, will characterize the capacity region for this class of channels.

Acknowledgment

The author wishes to thank Prof. T. M. Cover, who motivated this work and conjectured some of the results, for his many helpful suggestions during this research. He is also indebted to Dr. A. D. Wyner, who suggested the linear upper bound of Section IV, and to Prof. R. M. Gray for many interesting discussions.

Appendix A

Theorem 1

If

$$R_2 = I_{q_1q_2}(U;Z) - 2\varepsilon$$
$$R_1 = I_{q_1q_2}(X;Y \mid U) - 2\varepsilon \quad (A1)$$

then

$$\lim_{n\to\infty} \overline{P_e^{(m)}} = 0, \qquad m = 1,2,3,4 \quad (A2)$$

and hence

$$\lim_{n\to\infty} \bar{\mu} = \mathbf{0}.$$

1) $m = 1$: We have

$$\begin{aligned}\overline{P_e^{(1)}} &= \frac{1}{M_1 M_2} \sum_{k=1}^{M_1} \sum_{i=1}^{M_2} \overline{P_e^{(1)}(k,i)} \\ &= \sum_{z \in \mathscr{C}^n} \overline{p_2(z \mid x_{ki})\, d(u_i,z)} \\ &= \sum_{u \in \mathscr{A}^n} \sum_{x \in \mathscr{A}^n} \sum_{z \in \mathscr{C}^n} q_2(u) q_1(x \mid u) p_2(z \mid x)\, d(u,z) \\ &= E[d(u,z)] = \Pr\,[d(u,z) = 1] \\ &= \Pr\left[I_{q_1 q_2}(u;z) \le \frac{R_2 + I_{q_1 q_2}(U;Z)}{2} \right]. \end{aligned} \tag{A3}$$

With $R_2 = I_{q_1 q_2}(U;Z) - 2\varepsilon$, (A3) becomes

$$\overline{P_e^{(1)}} = \Pr\,[I_{q_1 q_2}(u;z) \le I_{q_1 q_2}(U;Z) - \varepsilon]. \tag{A4}$$

Since $I_{q_1 q_2}(u;z)$ is the sum of independent finite mean random variables, then, by the law of large numbers, we have

$$\lim_{n \to \infty} \overline{P_e^{(1)}} = 0. \tag{A5}$$

2) $m = 2$:

$$\begin{aligned}\overline{P_e^{(2)}} &= \frac{1}{M_1 M_2} \sum_{k=1}^{M_1} \sum_{i=1}^{M_2} \overline{P_e^{(2)}(k,i)} \\ &= \sum_{z \in \mathscr{C}^n} \sum_{\substack{j=1 \\ j \ne i}}^{M_2} \overline{p_2(z \mid x_{ki})(1 - d(u_j,z))}. \end{aligned} \tag{A6}$$

Now, because the cloud centers have been generated independently, for $i \ne j$, the random variable $p_2(z \mid x_{ki})$ is independent of the random variable $(1 - d(u_j,z))$, and we can write

$$\overline{P_e^{(2)}} = \sum_{z \in \mathscr{C}^n} \sum_{\substack{j=1 \\ j \ne i}}^{M_2} \overline{p_2(z \mid x_{ki})}\; \overline{(1 - d(u_j,z))}. \tag{A7}$$

Now, in (A7) we have

$$\begin{aligned}\overline{p_2(z \mid x_{ki})} &= \sum_{u \in \mathscr{A}^n} \sum_{x \in \mathscr{A}^n} q_2(u) q_1(x \mid u) p_2(z \mid x) \\ &= p(z). \end{aligned} \tag{A8}$$

Hence, (A6) becomes

$$\begin{aligned}\overline{P_e^{(2)}} &= \sum_{z \in \mathscr{C}^n} \sum_{\substack{j=1 \\ j \ne i}}^{M_2} p(z) \sum_{u \in \mathscr{A}^n} \sum_{x \in \mathscr{A}^n} q_2(u) q_1(x \mid u)(1 - d(u,z)) \\ &= (M_2 - 1) \sum_{z \in \mathscr{C}^n} \sum_{u \in S(z)} q_2(u) p(z). \end{aligned} \tag{A9}$$

In the last step, we have summed over the x and restricted the sum over the u to $S(z)$, since $1 - d(u,z)$ is 0, iff $u \notin S(z)$. But, whenever $u \in S(z)$

$$I_{q_1 q_2}(u;z) = \frac{1}{n} \log_2 \frac{p(z \mid u)}{p(z)} > \frac{R_2 + I_{q_1 q_2}(U;Z)}{2}$$

or

$$p(z) < p(z \mid u) 2^{-n(R_2 + I_{q_1 q_2}(U;Z)/2)}. \tag{A10}$$

Consequently,

$$\begin{aligned}\overline{P_e^{(2)}} &< (M_2 - 1) \sum_{z \in \mathscr{C}^n} \sum_{u \in S(z)} q_2(u) p(z \mid u) 2^{-n(R_2 + I_{q_1 q_2}(U;Z)/2)} \\ &< M_2 2^{-n(R_2 + I_{q_1 q_2}(U;Z)/2)} \sum_{z \in \mathscr{C}^n} \sum_{u \in S(z)} q_2(u) p(z \mid u) \\ &\le 2^{-n[R_2 - (R_2 - I_{q_1 q_2}(U;Z)/2)]} \sum_{z \in \mathscr{C}^n} \sum_{u \in \mathscr{A}^n} q_2(u) p(z \mid u) \\ &= 2^{-n(I_{q_1 q_2}(U;Z) - R_2/2)} \\ &= 2^{-n\varepsilon}. \end{aligned} \tag{A11}$$

Hence, for all positive ε,

$$\lim_{n \to \infty} \overline{P_e^{(2)}} = 0. \tag{A12}$$

3) $m = 3$:

$$\begin{aligned}\overline{P_e^{(3)}} &= \frac{1}{M_1 M_2} \sum_{k=1}^{M_1} \sum_{i=1}^{M_2} \overline{P_e^{(3)}(k,i)} \\ &= \sum_{y \in \mathscr{B}^n} \overline{p_1(y \mid x_{ki})\, d(x_{ki}, y \mid u_i)} \\ &= \sum_{u \in \mathscr{A}^n} \sum_{x \in \mathscr{A}^n} \sum_{y \in \mathscr{B}^n} q_2(u) q_1(x \mid u) p_1(y \mid x)\, d(x, y \mid u) \\ &= E[d(x, y \mid u)] = \Pr\,[d(x, y \mid u) = 1] \\ &= \Pr\left[I_{q_1 q_2}(x, y \mid u) \le \frac{R_1 + I_{q_1 q_2}(X; Y \mid U)}{2} \right]. \end{aligned} \tag{A13}$$

With $R_1 = I_{q_1 q_2}(X; Y \mid U) - 2\varepsilon$, we can again apply the law of large numbers and conclude that

$$\lim_{n \to \infty} \overline{P_e^{(3)}} = 0. \tag{A14}$$

4) $m = 4$:

$$\begin{aligned}\overline{P_e^{(4)}} &= \frac{1}{M_1 M_2} \sum_{k=1}^{M_1} \sum_{i=1}^{M_2} \overline{P_e^{(4)}(k,i)} \\ &= \sum_{y \in \mathscr{B}^n} \sum_{\substack{l=1 \\ l \ne k}}^{M_1} \overline{p_1(y \mid x_{ki}\,(1 - d(x_{li}, y \mid u_i))}. \end{aligned} \tag{A15}$$

Now, conditioned on u_i, the random variables $p(y \mid x_{ki})$ and $(1 - d(x_{li}, y \mid u_i))$ are independent for $k \ne l$. We now use the fact that, if A and B are two random variables conditionally independent given a third random variable C, we can write

$$E[AB] = E[E[AB \mid C]] = E[E[A \mid C]E[B \mid C]].$$

Hence, in (A15)

$$\begin{aligned}&\overline{p_1(y \mid x_{ki})(1 - d(x_{li}, y \mid u_i))} \\ &\quad = \sum_{u \in \mathscr{A}^n} q_2(u) \left[\sum_{x \in \mathscr{A}^n} q_1(x \mid u) p_1(y \mid x) \right] \\ &\qquad \cdot \left[\sum_{x \in \mathscr{A}^n} q_1(x \mid u)(1 - d(x, y \mid u)) \right]. \end{aligned} \tag{A16}$$

Now, in (A16), we can write

$$\sum_{x \in \mathscr{A}^n} q_1(x \mid u) p_1(y \mid x) = p(y \mid u). \tag{A17}$$

Substitution of (A17) in (A16), and of (A16) in (A15) yields

$$\begin{aligned}\overline{P_e^{(4)}} &= (M_1 - 1) \sum_{y \in \mathscr{B}^n} \sum_{u \in \mathscr{A}^n} q_2(u) p(y \mid u) \sum_{x \in T(y \mid u)} q_1(x \mid u) \\ &= (M_1 - 1) \sum_{y \in \mathscr{B}^n} \sum_{u \in \mathscr{A}^n} \sum_{x \in T(y \mid u)} q_2(u) p(y \mid u) q_1(x \mid u).\end{aligned} \tag{A18}$$

In (A18), we have restricted the summation over x to $T(y \mid u)$, since $(1 - d(x, y \mid u))$ is zero outside this set. But, whenever $x \in T(y \mid u)$, we have

$$p(y \mid u) < p_1(y \mid x) 2^{-n(R_1 + I_{q_1 q_2}(X;Y|U)/2)}.$$

Thus (A18) becomes

$$\begin{aligned}\overline{P_e^{(4)}} &< M_1 2^{-n(R_1 + I_{q_1 q_2}(X;Y|U)/2)} \\ &\quad \cdot \sum_{y \in \mathscr{B}^n} \sum_{u \in \mathscr{A}^n} \sum_{x \in T(y|u)} q_2(u) q_1(x \mid u) p_1(y \mid x) \\ &\leq 2^{-n(I_{q_1 q_2}(X;Y|U) - R_1/2)} \\ &\quad \cdot \sum_{y \in \mathscr{B}^n} \sum_{u \in \mathscr{A}^n} \sum_{x \in \mathscr{A}^n} q_2(u) q_1(x \mid u) p_1(y \mid x) \\ &= 2^{-n\varepsilon}.\end{aligned} \tag{A19}$$

And

$$\lim_{n \to \infty} \overline{P_e^{(4)}} = 0. \tag{A20}$$

5) *Conclusion:* Collecting (23), (32), and (A2) we conclude that

$$\lim_{n \to \infty} \bar{\mu} = \mathbf{0}.$$

Appendix B

In this appendix, we complete the proof of the coding theorem by showing that there exists a sequence of broadcast codes such that the maximum probability of error goes to the zero vector as $n \to \infty$. The proof is not restricted to degraded broadcast channels, since no assumption is made about the channel itself. We shall need the following lemma, in which we assume that there exists a random coding scheme such that the expected arithmetic average probability of error goes to zero when the block length of the code is allowed to increase without bound. The existence of such a sequence of codes is proved in Appendix A.

Lemma: If there exists a sequence of random codes such that

$$\lim_{n \to \infty} \overline{\boldsymbol{\mu}^{(n)}} = \mathbf{0},$$

then there exists a sequence of codes of identical rates, such that

$$\boldsymbol{\mu}^{(n)} \to \mathbf{0}, \qquad \text{as } n \to \infty. \tag{B1}$$

Proof: If

$$\lim_{n \to \infty} \overline{\boldsymbol{\mu}^{(n)}} = \mathbf{0},$$

there exists an n for which

$$\bar{\boldsymbol{\mu}} < \boldsymbol{\eta} = (\eta, \eta, \cdots, \eta) \tag{B2}$$

no matter how small the $\boldsymbol{\eta}$-vector is. We use the shorthand notation

$$\boldsymbol{a} < \boldsymbol{b} \leftrightarrow a_j < b_j, \qquad 1 \leq j \leq N.$$

Then, there exists a code of length n, to which positive probability is assigned by the random coding procedure, and such that

$$\boldsymbol{\mu} < N\boldsymbol{\eta}. \tag{B3}$$

This is a well-known fact and it can easily be proved. Indeed, let

$$Q_j \triangleq \Pr\,[\{\text{codes}\}: \mu_j \geq N\eta]. \tag{B4}$$

Then, from (B2) and (B4),

$$\eta > \mu_j \geq Q_j \cdot N\eta$$

or

$$Q_j < \frac{1}{N}. \tag{B5}$$

Consequently

$$\begin{aligned}\Pr\,[\{\text{codes}\}: \boldsymbol{\mu} < N\boldsymbol{\eta}] &\geq 1 - \sum_{j=1}^{N} \Pr\,[\{\text{codes}\}: \mu_j \geq N\eta] \\ &= 1 - \sum_{j=1}^{N} Q_j > 0\end{aligned} \tag{B6}$$

and there exists at least one code such that (B3) is satisfied. Since $\boldsymbol{\eta}$ can be made arbitrarily small (if there is not bound on n), we have proved the lemma.

We finally prove the following theorem.

Theorem 2: If there exists a sequence of random codes such that

$$\lim_{n \to \infty} \overline{\boldsymbol{\mu}^{(n)}} = \mathbf{0}$$

then there exists a sequence of codes of identical rates such that

$$\boldsymbol{\lambda}^{(n)} \to \mathbf{0}, \qquad \text{as } n \to \infty.$$

Proof: The notation used in this proof was introduced in Section II. We consider a code such that (B3) is satisfied. The existence of such a code is guaranteed by the lemma.

For that code, consider

$$S^{(N)} \triangleq \{\boldsymbol{i}: \boldsymbol{\lambda}(\boldsymbol{i}) < 2^N N^2 \boldsymbol{\eta}\} \tag{B7}$$

with

$$s^{(N)} \triangleq |S^{(N)}|.$$

We claim that

$$s^{(N)} > M \frac{2^N - 1}{2^N}. \tag{B8}$$

Let

$$T_j \triangleq \{\boldsymbol{i}: \lambda_j(\boldsymbol{i}) \geq 2^N N^2 \eta\} \tag{B9}$$

with $t_j \triangleq |T_j|$. We have, from (B3) and (B9)

$$N\eta > \frac{1}{M} \sum_{\boldsymbol{i} \in \boldsymbol{I}} \lambda_j(\boldsymbol{i}) \geq \frac{1}{M} t_j 2^N N^2 \eta$$

or

$$t_j < \frac{M}{N \cdot 2^N}. \tag{B10}$$

Since

$$\boldsymbol{I} = S^{(N)} \cup T_1 \cup T_2 \cup \cdots \cup T_N \tag{B11}$$

and

$$S^{(N)} \cap T_j = \varnothing \tag{B12}$$

we have

$$s^{(N)} \geq M - \sum_{j=1}^{N} t_j > M - N\frac{M}{N2^N}$$

or

$$s^{(N)} > M\frac{2^N - 1}{2^N}. \tag{B13}$$

We now consider "shortened" i-vectors. Let

$$i^{(k)} \triangleq (i_1, i_2, \cdots, i_k), \qquad k = 1, \cdots, N-1. \tag{B14}$$

We define $S^{(k-1)}$ recursively as follows. Let $S^{(k-1)}$ be the set of $i^{(k-1)}$ having the property that, for each $i^{(k-1)}$ in $S^{(k-1)}$, there are at least $(M_k/2)$ choices of i_k yielding an $i^{(k)}$ in $S^{(k)}$. The set $S^{(N)}$ has been defined previously. The size of $S^{(k)}$ will be denoted $s^{(k)}$. We shall show that

$$s^{(k)} > \frac{2^k - 1}{2^k} \cdot \prod_{j=1}^{k} M_j. \tag{B15}$$

The proof is by induction. We have already shown that (B15) holds for $k = N$. Assume now that the size $s^{(k)}$ of $S^{(k)}$ satisfies (B15). By definition, if $i^{(k-1)} \in S^{(k-1)}$, there are between $(M_k/2)$ and M_k choices of i_k which we can append to $i^{(k-1)}$ to yield an $i^{(k)}$ in $S^{(k)}$. The number of those $i^{(k-1)}$ is $s^{(k-1)}$. On the other hand, if $i^{(k-1)} \notin S^{(k-1)}$, there are less than $(M_k/2)$ choices of i_k with the same property. The number of $i^{(k-1)}$ not in $S^{(k-1)}$ is clearly

$$\prod_{j=1}^{k-1} M_j - s^{(k-1)}. \tag{B16}$$

Hence, we find

$$s^{(k-1)}M_k + \left[\prod_{j=1}^{k-1} M_j - s^{(k-1)}\right]\frac{M_k}{2} > s^{(k)} > \frac{2^k - 1}{2^k}\prod_{j=1}^{k} M_j$$

or

$$s^{(k-1)}\frac{M_k}{2} > \frac{2^k - 1}{2^k}\prod_{j=1}^{k} M_j - \frac{M_k}{2}\prod_{j=1}^{k-1} M_j$$

or

$$s^{(k-1)} > \frac{2^{k-1} - 1}{2^{k-1}}\prod_{j=1}^{k-1} M_j. \tag{B17}$$

This recursive property, together with (B13) proves (B15) in general. In particular,

$$s^{(1)} > \frac{2^1 - 1}{2^1} M_1 = \frac{M_1}{2}.$$

There are at least $(M_1/2)$ indices i_1, to each of which we can append at least $(M_2/2)$ different indices i_2 to yield pairs in a set $S^{(2)}$. To each pair in $S^{(2)}$, we can append at least $(M_3/2)$ different i_3's to yield triples in a set $S^{(3)}$. If we keep appending successive indices like this, we shall end up with the set $S^{(N)}$ described previously, and defined in (B7). Consequently, we have separated a subset $S^{(N)}$ of I, such that the code words with indices $i \in S^{(N)}$ all satisfy

$$\lambda(i) < 2^N N^2 \eta. \tag{B18}$$

Let

$$I^* = I_1^* \times I_2^* \times \cdots \times I_N^*, \tag{B19}$$

where

$$I_j^* \subset I_j \qquad \text{and} \qquad |I_j^*| = \frac{M_j}{2}. \tag{B20}$$

From the above considerations, it should be clear that there is a one-to-one mapping ϕ from I^* into $S^{(N)}$

$$\phi: I^* \to S^{(N)}. \tag{B21}$$

Let

$$\{x(i): i \in I\} \tag{B22}$$

be a code satisfying (B3). We have shown that such a code exists. Then, the code

$$\{x(\phi(i)): i \in I^*\} \tag{B23}$$

is a code satisfying (B18). The rate point R^* of this code is given by

$$R_k^* = \frac{1}{n}\log_2\frac{M_k}{2} = R_k - \frac{1}{n}, \qquad k = 1, \cdots, N \tag{B24}$$

and can be made arbitrarily close to R when $n \to \infty$.

In conclusion, we have proved the existence of a code such that the maximum probability of error is bounded by an arbitrarily small number $2^N N^2 \eta$ for all component channels simultaneously. Letting $n \to \infty$, we can make η arbitrarily small, and bound λ down to **0**.

The condition for existence of a sequence of codes with maximum probability of error going to zero when $n \to \infty$ is the same as the condition for existence of a sequence of random codes with expected arithmetic average probability of error going to zero, by (B24).

As mentioned before, the proof of the theorem does not rely on the fact that the broadcast channel is degraded. Hence, it may be said that, for any broadcast channel, sufficient conditions on the rates for the existence of a random coding theorem are also sufficient conditions to guarantee the achievability of these rates, in the sense of maximum probability of error.

References

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

[2] C. E. Shannon, "A note on a partial ordering for communication channels," *Inform. Contr.*, vol. 1, pp. 390–397, 1958.

[3] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968, pp. 17–19.

[4] C. E. Shannon, "Certain results in coding theory for noisy channels," *Inform. Contr.*, vol. 1, pp. 6–25, 1957.

[5] J. Wolfowitz, personal communication.

[6] P. Bergmans, Ph.D. dissertation, Dep. Elec. Eng., Stanford Univ., Stanford, Calif., 1972.