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 \( \infty \). 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 \( \mathcal{X} \), and the output alphabet of the \( j \)-th terminal is \( \mathcal{Y}_j \). The transition probability of the broadcast channel is

\[
p(y_1, y_2, \ldots, y_N | x).
\]

Fig. 1. Broadcast channel.

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 \( p(y_1 | x), p(y_2 | x), \ldots \) of the component channels \( A_1, A_2, \ldots, A_N \) (Fig. 2), and all broadcast channels with same marginals will be equivalent in this context.

\( C_j \), the capacity of the \( j \)-th 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 \( S_1, S_2, \ldots, S_N \) 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 \( j \)-th source is intended to be received by the \( j \)-th 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

\[
R_j = \frac{1}{n} \log_2 M_j \tag{2}
\]
Using Wolfowitz' notation, we define a \((M,n,\lambda)-\)code as a code of length \(n\), size \(M\), and maximum probability of error \(\lambda\). \(R\) is an achievable rate if there exists a sequence of \((M,n,\lambda^{(n)})\)-codes, with \(M_j = \lfloor \exp_2 (nR_j) \rfloor \), such that \(\lambda^{(n)} \to 0\) when \(n \to \infty\). The set of achievable rates, or capacity region of the broadcast channel, will be denoted \(\mathcal{C}\).

**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

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

is an achievable rate, by simple time sharing. Every point dominated by such an \(R\) is, of course, also achievable. Hence,

\[
\mathcal{R}_{TS} = \{(R_1, \cdots, R_N) : R_j \leq \tau_j C_j \forall j, \text{ for some } \{\tau_j\}\} \tag{12}
\]

is the set of rates achievable by time sharing (Fig. 4). The boundary of \(\mathcal{R}_{TS}\) 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 \(\mathcal{X}\), output alphabet \(\mathcal{Y}\), and transition probability \(p_1(y_i | x_i)\), and let \(A_2\) be another channel with same input alphabet \(\mathcal{X}\), output alphabet \(\mathcal{Y}_2\), and transition probability \(p_2(y_{ij} | x_i)\). The degradation is expressed by (Fig. 5)

\[
p_2(y_{ij} | x_i) = \sum_{y_{ij} \in \mathcal{Y}_2} p_3(y_{ij} | y_i) p_1(y_i | x_i), \tag{13}
\]

where \(p_3(y_{ij} | y_i)\) is the transition probability of the degrading channel \(D_2\), with input alphabet \(\mathcal{Y}_1\) and output alphabet \(\mathcal{Y}_2\).

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 \(C_1 \geq C_2\), where \(C_2\) is the capacity of \(A_2\).

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-
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

\[
1 \leq k \leq M_1 \text{ and } 1 \leq i \leq M_2.
\]

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 \( \mathcal{A} \) with same index \( i \) will be represented by a vector \( u_i \), in \( \mathcal{A}^n \), 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 \( q_1(x | u) \), defined for \( u, x \in \mathcal{A} \), are parameters that will be used in the random coding scheme. \( q_1(x | u) \) can be thought of as the transition probability of an artificial channel \( E \). \( p_1(y | x) \) and \( p_2(z | x) \) 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 \( \mathcal{A}^n \), with letters independently drawn according to \( q_2(u) \). To each cloud center, we append \( M_1 = 2^{nR_1} \) satellites in \( \mathcal{A}^n \), independently drawn according to \( q_1(x | u) \), 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 $q_1(x | u)$.

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

$$I_{q_2}(u; z) = \frac{1}{n} \log_2 \frac{p(z | u)}{p(z)},$$

where

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

and

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

Then, by definition [3]

$$E[I_{q_2}(u; z)] = I_{q_2}(U; Z).$$

Let

$$S(z) = \{ u \in \mathbb{A}^n: I_{q_2}(u; z) > \frac{R_2 + I_{q_2}(U; Z)}{2} \}$$

be the decoding set for $z$, and define

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

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

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

with

$$P_e^{(1)}(k, i) = \sum_{x \in \mathbb{A}^n} p_2(z | x_{ki}) d(u_i, z)$$

and

$$P_e^{(2)}(k, i) = \sum_{x \in \mathbb{A}^n} p_2(z | x_{ki}) \sum_{j \neq i}^{M_2} (1 - d(u_j, z)).$$

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

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

Consequently, we have

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

and

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

$$I_{q_2}(x; y | u) = \frac{1}{n} \log_2 \frac{p(y | x, u)}{p(y | u)}$$

and

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

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

$$E[I_{q_2}(x; y | u)] = I_{q_2}(X; Y | U).$$

Let

$$T(y | u) \triangleq \begin{cases} 1, & x \in \mathbb{A}^n: I_{q_2}(x; y | u) > \frac{R_1 + I_{q_2}(X; Y | U)}{2} \\ 0, & \text{otherwise.} \end{cases}$$

be the conditional satellite decoding set for $y$, and define

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

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 $x_{kl}$ is not in $T(y | u_i)$ (error of the third type, with probability $P_e^{(3)}(k, l)$), or that there is some $x_{kl}$ in $T(y | u_i)$ for $l \neq k$ (error of the fourth type, with probability $P_e^{(4)}(k, l)$).

We have

$$\lambda_4(k, l) \leq \lambda_2(k, i) + P_e^{(3)}(k, l) + P_e^{(4)}(k, l)$$

with

$$P_e^{(3)}(k, l) = \sum_{y \in \mathbb{A}^n} p_1(y | x_{kl}) d(x_{kl}, y | u_i)$$

and

$$P_e^{(4)}(k, l) = \sum_{y \in \mathbb{A}^n} p_1(y | x_{kl}) \sum_{j \neq k}^{M_1} (1 - d(x_{kj}, y | u_j)).$$
Again, with a uniform distribution on $k$ and $i$, we have
\[ P_e^{(m)} = \frac{1}{M_1 M_2} \sum_{k=1}^{M_1} \sum_{i=1}^{M_2} P_e^{(m)}(k,i), \quad m = 3, 4 \] (31)
and
\[ \mu_1 \leq \frac{\mu_2 + P_e^{(3)} + P_e^{(4)}}{\mu_2} \leq \frac{\mu_1}{\mu_2} + \frac{P_e^{(3)}}{P_e^{(4)}}. \] (32)

In Appendix A, we prove rigorously that, if
\[ R_1 = I_{q_1}(U; Z) - 2\varepsilon \]
\[ R_2 = I_{q_2}(X; Y | U) - 2\varepsilon \] (33)
\[ \mu_1 \text{ and } \mu_2 \text{ go to zero simultaneously as } n \to \infty. \]

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 $S$.

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 $M_N = 2^{R_N}$ cloud centers in $d^n$ according to $q_N(x_N)$. Then, select $M_{N-1} = 2^{R_{N-1}}$ satellites per cloud center, according to $q_{N-1}(x_{N-1} | x_N)$, $M_{N-2} = 2^{R_{N-2}}$ subsatellites per satellite in each cloud, according to $q_{N-2}(x_{N-2} | x_{N-1})$, 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 $q_i(x_i | x_{i-1})$. This is illustrated in Fig. 10. The artificial satellizing channels are cascaded, and the broadcast channel is of the degraded type.

If, for some $q_1, q_2, \ldots, q_N$, we have
\[ R_N < I_{q_1}(X_N; Y_N) - \varepsilon \]
\[ R_{N-1} < I_{q_1} \cdots q_{N-1}(X_{N-1}; Y_{N-1} | X_N) - \varepsilon \]
\[ R_2 < I_{q_2}(X_2; Y_2 | X_1) - \varepsilon \]
\[ R_1 < I_{q_1} \cdots q_{N-1}(X_1; Y_1 | X_2) - \varepsilon \] (34)
\[ \mu \text{ goes to zero as } n \to \infty. \]

So far, we have proved that with certain conditions on $R$, there exists a random coding scheme for which the expected average probability of error $\mu$ goes to zero as $n \to \infty$. In Appendix B, we prove that, with identical conditions on $R$, there exists a sequence of codes for which the maximum probability of error $\lambda$ goes to zero as $n \to \infty$. Hence, the set $\mathcal{J}$ of all $R$ described parametrically by $q_1, \ldots, q_N$ in (34) is completely contained in the capacity region. Special conditions on $q_1, \ldots, q_N$ will yield points on the boundary of $\mathcal{J}$. In particular, if $E_i$ is a useless channel ($q_k(x_k | x_{k-1})$ is only a function $q_k(x_k)$ of $x_k$), and if $E_1, \ldots, E_{k-1}$ are perfect channels, we have
\[ R_k \leq I(X_k; Y_k | X_{k+1}) = I(X_k; Y_k) = I(X_1; Y_k). \] (35)

Letting $q_k(x_k)$ be the probability mass function maximizing the information between the input of the channel and the $k$th terminal, we can achieve the rate point $R = (0, \ldots, 0, C_k, 0, \ldots, 0)$.

If $\mathcal{R}$ is the convex hull of $\mathcal{J}$, we can also achieve any rate point in $\mathcal{R}$ by time sharing between some points in $\mathcal{J}$ (Fig. 11). Consequently, $\mathcal{R}_{\text{FS}}$ 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 $\mathcal{J}$ 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 $\mathcal{W}$, 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 $\mathcal{J}$ described in (34) is the set
\[ \mathcal{J}_e = \{(I(X_1; Y_1), \ldots, I(X_1; Y_N)); p(x_1) \geq 0, \]
\[ \sum p(x_1) = 1 \}. \] (36)
Indeed, we have
\[ I(X_k; Y_k | X_{k+1}) = H(Y_k | X_{k+1}) - H(Y_k | X_k) \]
\[ \leq H(Y_k) - H(Y_k | X_k) = I(X_k; Y_k) \leq I(X_1; Y_1). \]

Consequently, the convex hull \( \mathcal{R}_a \) of \( \mathcal{J} \) is contained in the convex hull \( \mathcal{R}_u \) of \( \mathcal{J} \). \( \mathcal{R}_u \) 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 \( \mathcal{R} \) and \( \mathcal{R}_u \) are rate points of the form \( R = (0, \ldots, 0, c_i, 0, \ldots, 0) \). A trivial consequence of (37) is
\[ R_i \leq c_i. \]

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
\[ \sum_{i=1}^{N} R_i \leq C_1. \]

Another way to show this is to note that
\[ C_1 \geq I_{q_1, \ldots, q_N}(X_1; Y_1) = I_{q_1, \ldots, q_N}(X_N; X_{N-1}, \ldots, X_1; Y_1) \]
\[ = I_{q_1, \ldots, q_N}(X_N; Y_1) + I_{q_1, \ldots, q_N}(X_{N-1}; Y_1 | X_N) + \cdots \]
\[ + I_{q_1, \ldots, q_N}(X_1; Y_1 | X_{N-1}, \ldots, X_2) = I_{q_1, \ldots, q_N}(X_N; Y_1) + I_{q_1, \ldots, q_N}(X_{N-1}; Y_1 | X_N) + \cdots \]
\[ + I_{q_1, \ldots, q_N}(X_1; Y_1 | X_2) \geq I_{q_1, \ldots, q_N}(X_N; Y_N) + I_{q_1, \ldots, q_N}(X_{N-1}; Y_{N-1} | X_N) + \cdots \]
\[ + I_{q_1, \ldots, q_N}(X_1; Y_1 | X_2) \geq R_N + R_{N-1} + \cdots + R_1. \]

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 \( X_i \) maximizes the mutual information between \( X_i \) and any of the \( Y_j \). Indeed
\[ \sum_{i=1}^{N} R_i = \sum_{i=1}^{N-1} I(X_i; Y_i | X_{i+1}) + I(X_N; Y_N) \]
\[ = \sum_{i=1}^{N-1} I(X_i; Y_i | X_{i+1}) + I(X_N; Y_j) = I(X_1, X_2, \ldots, X_N; Y_1) = C. \]

If not all components of the broadcast channel are equal, the preceding fact can easily be generalized. For example, let
\[ C_1 = C_2 < C_3 < C_4 = C_5. \]

Fig. 10 can be simplified to group components with same transition probability. In this case, we have (Fig. 13):
\[ R_1 + R_2 < I_{q_1, q_2}(X_1; Y_1 | X_2) \]
\[ R_3 < I_{q_3}(X_3; Y_3 | X_4) \]
\[ R_4 + R_5 < I_{q_3, q_4}(X_4; Y_4) \]
with \( R_k \geq 0 \).

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 \( I_1, I_2, \ldots, I_N \) of the one-to-one case are replaced by \( 2^N - 1 \) index sets \( I_\theta \) such that
\[ \theta = (\theta_1, \theta_2, \ldots, \theta_N) \in \Theta \cong \{0,1\}^N - 0. \]

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 \( i_\theta \) of source \( S_\theta \) is intended to be received by all users \( U_j \) such that \( \theta_j = 1 \). The rate \( R_\theta \) of source \( S_\theta \) is defined by
\[ R_\theta = \frac{1}{n} \log_2 M_\theta. \]
where $M_\theta$ is the size of the output alphabet $I_\theta$ of $S_\theta$. The total information rate to user $j$ is the sum of the rates of the sources $S_\theta$ such that $\theta_j = 1$. Hence

$$R_j = \sum_{\theta \in \Theta} \theta_j R_\theta \quad (45)$$

The one-to-one case we have considered until now (one source index to a single user) is characterized by the fact that $R_{\theta_j} > 0$ if and only if $\theta$ has weight one (i.e., $\theta$ 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 $\mathcal{R}$ of achievable rates in the one-to-one case, it is easy to determine whether a particular rate $R = (R_1, \ldots, R_N)$ is achievable. Let

$$R_N' = \sum_{\theta \in \Theta} \theta_j R_\theta$$
$$R_{N-1}' = \sum_{\theta \in \Theta} \theta_{N-1} \theta_j R_\theta$$
$$R_{N-2}' = \sum_{\theta \in \Theta} \theta_{N-2} \theta_j \theta_{N-1} R_\theta$$
$$\vdots$$
$$R_1' = \sum_{\theta \in \Theta} \theta_1 \theta_2 \cdots \theta_{N-1} \theta_j R_\theta \quad (46)$$

We claim that $R \in R_{2N-1}^*$ is achievable if the rate point $(R_1', R_2', \ldots, R_N') \in \mathcal{R}$. The sufficient condition just described is also necessary if $\mathcal{R}$ is the capacity region $\mathcal{C}$ 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

$$R_j' = \sum_{\theta \in \Theta} \theta_j R_\theta \quad (47)$$

A superposition code is then used to encode the information just described, with rates $R_{\theta_j}', R_{\theta_{N-1}}', \ldots, R_1'$, on the successive levels of satellization. The proof of the converse (assuming that $\mathcal{R} = \mathcal{C}$) is by contradiction. It is shown that if it is possible to achieve some rate point $R \in R_{2N-1}^*$ such that the rate point defined in (46) does not lie in $\mathcal{C}$, 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: $R_{\theta_j} > 0$ if and only if $\theta$ is of the form $(1,1,\ldots,1,0,0,\ldots,0)$. 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 subprocesses representing the different degrees of refinement, with the $N$th source representing the coarsest information. User $k$ would then receive arbitrarily well the output of sources $S_1, \ldots, S_{k-1}$ (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 $R_{2N}^*$.

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

$$p_3 = p_1 p_3 + p_1 p_2 = p_1 + p_2 \quad (48)$$

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

$$I(B; C | A) = H(C | A) - H(C | B, A) = \mathcal{H}(\alpha * \beta) - \mathcal{H}(\beta), \quad (49)$$

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

$$\mathcal{H}(p) = -p \log_2 p - \bar{p} \log_2 \bar{p} \quad (50)$$

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

$$R_N = 1 - \mathcal{H}(q_1 * q_2 * \cdots * q_{N-1} * p_N)$$
$$R_{N-1} = \mathcal{H}(q_1 * q_2 * \cdots * q_{N-1} * p_{N-1}) - \mathcal{H}(q_1 * q_2 * \cdots * q_{N-2} * p_{N-1})$$
$$\vdots$$
$$R_2 = \mathcal{H}(q_1 * q_2 * p_2) - \mathcal{H}(q_1 * p_2)$$
$$R_1 = \mathcal{H}(q_1 * p_1) - \mathcal{H}(p_1) \quad (51)$$

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

An alternate notation for the $R_j$ is given by

$$R_j = \mathcal{H}(r_j * p_j) - \mathcal{H}(r_{j-1} * p_j) \quad (52)$$
where
\[ 0 = r_0 \leq r_1 \leq r_2 \leq \ldots \leq r_{N-1} \leq r_N = \frac{1}{2}. \] (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 \( R_1 = (0,0,\ldots,0,C_i,0,\ldots,0) \) (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 \( R = (C_i,C_i,\ldots,C_i,0,0,\ldots,0) \). 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
\[
R_2 = 1 - \mathcal{H}(q_1 * p_2) \\
R_1 = \mathcal{H}(q_1 * p_1) - \mathcal{H}(p_1). \] (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
\[
\frac{dR_2}{dR_1(q_1,p_2)} = \frac{2}{\mathcal{H}(q_1 * p_1)} \left( \frac{p_2 - p_1}{p_2/p_1} \right) \] (55)
\[
\frac{dR_2}{dR_1(C_i,0)} = \frac{2}{\mathcal{H}(q_1 * p_1)} \left( \frac{p_2 - p_1}{p_2/p_1} \right)^2 \] (56)

The slope in (55) will be zero if \( p_1 \) 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
\[
E = \frac{R_1}{C_i} + \frac{R_2}{C_i} - 1 \] (57)

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

**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
\[
R_2 = I_{x|z}(U;Z) - 2\varepsilon \\
R_1 = I_{u|x}(X;Y | U) - 2\varepsilon \] (A1)

then
\[
\lim_{n \to \infty} \frac{P_e^{(n)}}{m} = 0, \quad m = 1,2,3,4 \] (A2)

and hence
\[
\lim_{n \to \infty} \bar{P} = 0.\]
1) \( m = 1 \): We have

\[
\bar{P}_e^{(1)} = \frac{1}{M_1 M_2} \sum_{k=1}^{M_1} \sum_{i=1}^{M_2} P_e^{(1)}(k,i)
\]

\[=
\sum_{z \in \mathbb{Z}_n} \sum_{x \in \mathbb{Z}_n} q_2(z|x)q_1(x|u)p_2(z|x) d(u,z)
\]

\[= F[d(u,z)] = \Pr \left[ d(u,z) = 1 \right]
\]

\[= \Pr \left[ I_{q_1q_2}(u; z) \leq \frac{R_2 + I_{q_1q_2}(U; Z)}{2} \right]. \tag{A3}
\]

With \( R_2 = I_{q_1q_2}(U; Z) - 2\varepsilon \), (A3) becomes

\[
\bar{P}_e^{(1)} = \Pr \left[ I_{q_1q_2}(u; z) \leq I_{q_1q_2}(U; Z) - \varepsilon \right]. \tag{A4}
\]

Since \( I_{q_1q_2}(u; z) \) is the sum of independent finite mean random variables, then, by the law of large numbers, we have

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

2) \( m = 2 \):

\[
\bar{P}_e^{(2)} = \frac{1}{M_1 M_2} \sum_{k=1}^{M_1} \sum_{i=1}^{M_2} P_e^{(2)}(k,i)
\]

\[=
\sum_{z \in \mathbb{Z}_n} \sum_{j=1}^{M_0} \sum_{i=1}^{M_0} p_2(z|x_i)(1 - d(u_j,z)). \tag{A6}
\]

Now, because the cloud centers have been generated independently, for \( i \neq j \), the random variable \( p_2(z|x_i) \) is independent of the random variable \( (1 - d(u_j,z)) \), and we can write

\[
\bar{P}_e^{(2)} = \frac{1}{M_1 M_2} \sum_{z \in \mathbb{Z}_n} \sum_{i=1}^{M_0} p_2(z|x_i)(1 - d(u_j,z)). \tag{A7}
\]

Now, in (A7) we have

\[
\frac{p_2(z|x_i)}{p_2(z|x_i)} = \sum_{u \in \mathbb{Z}_n \cap d \mathbb{Z}_n} q_2(u)q_1(x|u)p_2(z|x)
\]

\[= p(z). \tag{A8}
\]

Hence, (A6) becomes

\[
\bar{P}_e^{(2)} = \frac{1}{M_1 M_2} \sum_{z \in \mathbb{Z}_n} \sum_{j=1}^{M_0} \sum_{i=1}^{M_0} q_2(z|x_i)q_1(x|u)(1 - d(u_j,z))
\]

\[= (M_2 - 1) \sum_{x \in \mathbb{Z}_n} q_2(z|x)p(z). \tag{A9}
\]

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, if \( u \notin S(z) \). But, whenever \( u \in S(z) \)

\[
I_{q_1q_2}(u; z) = \frac{1}{n} \log_2 \frac{p(z|u)}{p(z)} \geq \frac{R_2 + I_{q_1q_2}(U; Z)}{2}
\]

or

\[
p(z) \leq p(z|u)2^{-n(R_2 + I_{q_1q_2}(U; Z))/2}. \tag{A10}
\]

Consequently,

\[
\bar{P}_e^{(2)} < (M_2 - 1) \sum_{z \in \mathbb{Z}_n} \sum_{u \in S(z)} q_2(z|u)p(z|u)2^{-n(R_2 + I_{q_1q_2}(U; Z))/2}.
\]

\[
\leq M_22^{-n(R_2 - I_{q_1q_2}(U; Z)/2)} \sum_{z \in \mathbb{Z}_n} \sum_{u \in S(z)} q_2(z|u)p(z|u)
\]

\[= 2^{-n(R_2 - I_{q_1q_2}(U; Z)/2)}.
\]

\[= 2^{-n_2}. \tag{A11}
\]

Hence, for all positive \( \varepsilon \),

\[
\lim_{n \to \infty} \bar{P}_e^{(2)} = 0. \tag{A12}
\]

3) \( m = 3 \):

\[
\bar{P}_e^{(3)} = \frac{1}{M_1 M_2} \sum_{k=1}^{M_1} \sum_{j=1}^{M_0} P_e^{(3)}(k,i)
\]

\[=
\sum_{y \in \mathbb{Z}_n} \sum_{x \in \mathbb{Z}_n} \sum_{i=1}^{M_0} q_2(y|x)p_2(z|x_j)(1 - d(u_i,y)). \tag{A13}
\]

Now, conditioned on \( u_i \), the random variables \( p_2(y|x_i) \) and \( (1 - d(x_i,y) \) are independent for \( k \neq i \). 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

\[
E[AB] = E[E[AB | C]] = E[E[A | C]E[B | C]]. \tag{A15}
\]

Hence, in (A15)

\[
\bar{P}_e^{(3)} = \frac{1}{M_1 M_2} \sum_{k=1}^{M_1} \sum_{i=1}^{M_0} P_e^{(3)}(k,i)
\]

\[=
\sum_{y \in \mathbb{Z}_n} \sum_{i=1}^{M_0} \sum_{l=1}^{M_0} p_1(y|x_i)(1 - d(x_i,y|u_i)). \tag{A16}
\]

Now, conditioned on \( u_i \), the random variables \( p(y|x_i) \) and \( (1 - d(x_i,y|u_i)) \) are independent for \( k \neq i \). 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

\[
E[AB] = E[E[AB | C]] = E[E[A | C]E[B | C]]. \tag{A15}
\]

Hence, in (A15)

\[
\bar{P}_e^{(3)} = \frac{1}{M_1 M_2} \sum_{k=1}^{M_1} \sum_{i=1}^{M_0} P_e^{(3)}(k,i)
\]

\[=
\sum_{y \in \mathbb{Z}_n} \sum_{i=1}^{M_0} \sum_{l=1}^{M_0} p_1(y|x_i)(1 - d(x_i,y|u_i)). \tag{A16}
\]
Now, in (A16), we can write
\[ \sum_{x \in S^n} q_2(x | u) p(y | x) = p(y | u). \quad (A17) \]
Substitution of (A17) in (A16), and of (A16) in (A15) yields
\[ P_e^{(5)} = (M_1 - 1) \sum_{x \in S^n} q_2(u) p(y | u) \sum_{x \in T(y | u)} q_1(x | u) \]
\[ = (M_1 - 1) \sum_{x \in S^n} \sum_{u \in S^n} q_2(u) p(y | u) q_1(x | u). \quad (A18) \]
In (A18), we have restricted the summation over \( x \) to \( T(y | u) \), since \((1 - d(x, y | u)) \) is zero outside this set. But, whenever \( x \in T(y | u) \), we have
\[ p(y | u) < p_1(y | x) 2^{-nR_1 + I_{1,2}(X;Y | U)/2}. \]
Thus (A18) becomes
\[ P_e^{(5)} < M_1 2^{-nR_1 + I_{1,2}(X;Y | U)/2} \]
\[ \leq 2^{-nI_{1,2}(X;Y | U)/2} \]
\[ \leq 2^{-nR_1 + I_{1,2}(X;Y | U)/2} \]
\[ = 2^{-nR_1 + I_{1,2}(X;Y | U)/2} \quad (A19) \]
And
\[ \lim_{n \to \infty} P_e^{(5)} = 0. \quad (A20) \]
5) Conclusion: Collecting (23), (32), and (A2) we conclude that
\[ \lim_{n \to \infty} \mu = 0. \]

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
\[ \lim_{n \to \infty} \mu^{(n)} = 0, \]
then there exists a sequence of codes of identical rates, such that
\[ \mu^{(n)} \to 0, \quad \text{as } n \to \infty. \quad (B1) \]

Proof: If
\[ \lim_{n \to \infty} \mu^{(n)} = 0, \]
there exists an \( n \) for which
\[ \mu < \eta = (\eta, \eta, \cdots, \eta) \quad (B2) \]
no matter how small the \( \eta \)-vector is. We use the shorthand notation
\[ a < b \leftrightarrow a_j < b_j, \quad 1 \leq j \leq N. \]
Then, there exists a code of length \( n \), to which positive probability is assigned by the random coding procedure, and such that
\[ \mu < N\eta. \quad (B3) \]
This is a well-known fact and it can easily be proved. Indeed, let
\[ Q_j \triangleq \Pr \{ \text{[codes]}: \mu_j \geq N\eta \}. \quad (B4) \]
Then, from (B2) and (B4),
\[ \eta > \mu_j \geq Q_j \cdot N\eta \]
or
\[ Q_j < \frac{1}{N}. \quad (B5) \]
Consequently
\[ \Pr \{ \text{[codes]}: \mu \leq N\eta \} \geq 1 - \sum_{j=1}^{N} Q_j > 0 \quad (B6) \]
and there exists at least one code such that (B3) is satisfied. Since \( \eta \) 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
\[ \lim_{n \to \infty} \mu^{(n)} = 0 \]
then there exists a sequence of codes of identical rates such that
\[ \lambda^{(n)} \to 0, \quad \text{as } n \to \infty. \]

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
\[ S^{(N)} \triangleq \{ i: \lambda(i) < 2^N N^2 \eta \} \quad (B7) \]
with
\[ \sigma^{(N)} \triangleq |S^{(N)}|. \]
We claim that
\[ \sigma^{(N)} > M \frac{2^N - 1}{2^N}. \quad (B8) \]
Let
\[ T_j \triangleq \{ i: \lambda_j(i) \geq 2^N N^2 \eta \} \quad (B9) \]
with \( t_j \triangleq |T_j| \). We have, from (B3) and (B9)
\[ N\eta > \frac{1}{M} \sum_{i=1}^{M} \lambda_j(i) \geq \frac{1}{M} t_j 2^N N^2 \eta \]
or
\[ t_j < \frac{M}{N \cdot 2^N}. \quad (B10) \]
Since
\[ I = S^{(N)} \cup T_1 \cup T_2 \cup \cdots \cup T_N \quad (B11) \]
and
\[ S^{(N)} \cap T_j = \emptyset \]  
(B12)

we have
\[ s^{(N)} \geq M - \sum_{j=1}^{N} t_j > M - N \frac{M}{N^2} \]

or
\[ s^{(N)} > M\frac{2^N - 1}{2^N} \]

(B13)

We now consider "shortened" i-vectors. Let
\[ f^{(k)} \triangleq (i_1, i_2, \ldots, i_k), \quad k = 1, \ldots, N - 1. \]  
(B14)

We define \( S^{(k-1)} \) recursively as follows. Let \( S^{(k-1)} \) be the set of \( f^{(k-1)} \) having the property that, for each \( f^{(k-1)} \) in \( S^{(k-1)} \), there are at least \( (M/2) \) choices of \( i_k \) yielding an \( f^{(k)} \) in \( S^{(k)} \). The set \( S^{(k)} \) has been defined previously. The size of \( S^{(k)} \) will be denoted \( s^{(k)} \). We shall show that
\[ s^{(k)} > \frac{2^{k-1} - 1}{2^{k-1}} \sum_{j=1}^{k-1} M_j \]  
(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 \( f^{(k-1)} \in S^{(k-1)} \), there are between \((M/2)\) and \( M_1 \) choices of \( i_k \) which we can append to \( f^{(k-1)} \) to yield an \( f^{(k)} \) in \( S^{(k)} \). The number of those \( f^{(k-1)} \) is \( s^{(k-1)} \). On the other hand, if \( f^{(k-1)} \notin S^{(k-1)} \), there are less than \((M/2)\) choices of \( i_k \) with the same property. The number of \( f^{(k-1)} \) not in \( S^{(k-1)} \) is clearly
\[ \frac{k-1}{2} \sum_{j=1}^{k-1} M_j = s^{(k-1)}. \]  
(B16)

Hence, we find
\[ s^{(k-1)}M_k + \left[ \frac{k-1}{2} \sum_{j=1}^{k-1} M_j - s^{(k-1)} \right] M_k > s^{(k)} > \frac{2^{k-1} - 1}{2^{k-1}} \sum_{j=1}^{k-1} M_j \]

or
\[ s^{(k-1)} \frac{M_k}{2} > \frac{2^{k-1} - 1}{2^{k}} \sum_{j=1}^{k-1} M_j - \frac{M_k}{2} \sum_{j=1}^{k-1} M_j \]

or
\[ s^{(k-1)} > \frac{2^{k-1} - 1}{2^{k}} \sum_{j=1}^{k-1} M_j. \]  
(B17)

This recursive property, together with (B13) proves (B15) in general. In particular,
\[ s^{(1)} = \frac{21 - 1}{21} M_1 = \frac{M_1}{2}. \]

There are at least \((M/2)\) indices \( i_1 \), to each of which we can append at least \((M/2)\) different indices \( i_2 \) to yield pairs in a set \( S^{(2)} \). To each pair in \( S^{(2)} \), we can append at least \((M/2)\) 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 \( I \) each such that the code words with indices \( i \in S^{(N)} \) all satisfy
\[ \lambda(i) < 2^N N^2 \eta. \]  
(B18)

Let
\[ I^* = I_1^* \times I_2^* \times \cdots \times I_N^*, \]  
(B19)

where
\[ I_j^* \subset I_j \quad \text{and} \quad |I_j^*| = \frac{M_j}{2}. \]  
(B20)

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

Let
\[ (x(i)): i \in I^* \]  
(B21)

be a code satisfying (B3). We have shown that such a code exists. Then, the code
\[ (x(\phi(i))): i \in I^* \]  
(B23)

is a code satisfying (B18). The rate point \( R^* \) of this code is given by
\[ R_k^* = \frac{1}{n} \log_2 \frac{M_k}{2} = R_k - \frac{1}{n}, \quad k = 1, \ldots, N \]  
(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 \( \eta \) arbitrarily small, and bound \( \lambda \) 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