Reprinted from:
ADVANCES IN COMMUNICATION SYSTEMS, VOL. 4
Theory and Applications
© 1975
ACADEMIC PRESS, INC.
New York San Francisco London

SOME ADVANCES IN BROADCAST CHANNELS

Thomas M. Cover†

I. INTRODUCTION 229

II. AN EXAMPLE OF A GAUSSIAN BROADCAST CHANNEL 232

III. BROADCAST CHANNEL FORMULATION 236

IV. REVIEW OF SOME EXTREME BROADCAST CHANNELS 240

A. Orthogonal Channels 240

B. Incompatible Channels 240

C. Bottleneck Channels 241

V. DEGRADED BROADCAST CHANNELS 242

VI. AN ACHIEVABLE REGION FOR THE GENERAL BROADCAST CHANNEL 244

A. Random Encoding for Theorem 2' 249

B. Decoding Rule 250

VII. THE GENERAL MULTIUSER FRAMEWORK 250

VIII. THE MULTIACCESS GAUSSIAN CHANNEL 251

IX. CONCLUDING REMARKS 257

REFERENCES 258

This work was supported under the National Science Foundation Grant GK-34363 and SRI ARPA Contract DAHC-15-73-C-0187.

INTRODUCTION

The broadcast channel [1] is an attempt to model the situation of a broadcaster with many receivers or a lecturer with many listeners. The general broadcast channel with k receivers is depicted in Fig. 1. The basic problem is to find

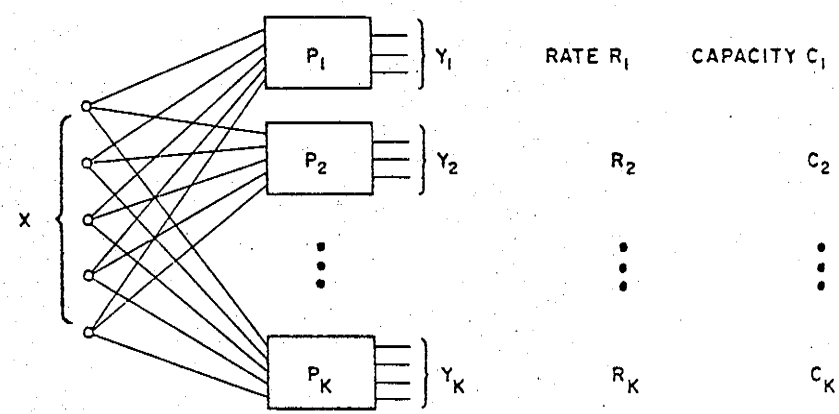

Fig. 1. Broadcast channel.

the set of simultaneously achievable rates $(R_1, R_2, \ldots, R_k)$. To date this problem has not been solved. The special case of sequentially degraded channels has, however, been solved by Bergmans [2] and Gallager [3]. An achievable rate region for the general broadcast channel has been put forth by van der Meulen [4] and Cover [5], but is not known to be the true capacity region. The above work will be characterized here. Also, related work on multiple-user channels will be briefly characterized. A nice review on related topics may be found in the work of Wyner [6]. Relevant papers include [19-31].

The basic background for broadcast communication are the maximin and time-sharing approaches. Suppose that the transmission channels to the receivers have respective channel capacities $C_1, C_2, \ldots, C_k$ bits per second. The first approach

that suggests itself is the maximin approach: send at rate $C_{min} = \min\{C_1, C_2, \ldots, C_k\}$. If the channels are "compatible," each receiver will understand perfectly at the rate $R = C_{min}$ bits/sec. Here the transmission rate is limited by the worst channel. At the other extreme, information could be sent at rate $R = C_{max}$, with resulting rates $R_i = 0$, $i = 1, 2, \ldots, k-1$, for all but the best channel, and rate $R_k = C_{max}$ for the best channel.

The next idea is that of time sharing. Allocate proportions of time $\lambda_1, \lambda_2, \ldots, \lambda_k$, $\lambda_i \geq 0$, $\Sigma \lambda_i = 1$, to sending at rates $C_1, C_2, \ldots, C_k$. Assuming compatibility of the channels and assuming $C_1 \leq C_2 \leq \cdots \leq C_k$, we find that the rate of transmission of information through the ith channel is given by

$$R_i = \sum_{j \leq i} \lambda_j C_j, \qquad i = 1, 2, \ldots, k$$

The reason for work on broadcast channels is that these regions of achievable rates can be exceeded by optimal encoding. It turns out that one should not transmit simultaneously to several channels at the rate of the worst channel, nor should one attempt to transmit information by a time-sharing or time-multiplexing method, but rather one should distribute the high-rate information across the low-rate message.

A brief outline of the paper follows. Section 2 provides a discussion of an example — the Gaussian broadcast channel — that contains some of the central ideas of the general broadcast channel. A review of the rigorous formulation of the problem is given in Section 3, and solved examples of extreme broadcast channels are given in Section 4. Section 5 presents the important case of the sequentially degraded broadcast channel. Some recent results on a lower bound for the capacity region are presented in Section 6. The broadcast channel and multiple access channel are put in a general multiuser context

in Section 7. Finally, Section 8 contains some new ideas on the multiple access Gaussian channel, in which it is shown that new users can come in on top of existing users without a diminution of rate or change in signalling strategy for the existing users.

II. AN EXAMPLE OF A GAUSSIAN BROADCAST CHANNEL

Before proceeding, let us consider from [1] a Gaussian broadcast channel with two receivers as shown in Fig. 2.

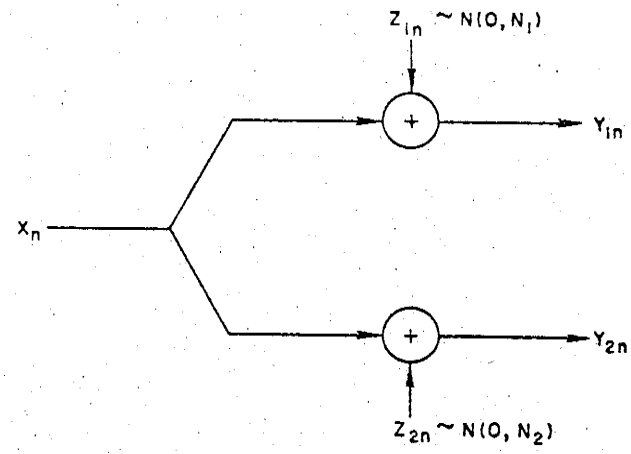

Fig. 2. Gaussian broadcast channel.

Let $\underline{z}_1 = (z_{11}, z_{12}, \ldots, z_{1n}, \ldots)$ be a sequence of independently identically distributed (i.i.d.) normal random variables with mean zero and variance N_1, and let $\underline{z}_2 = (z_{21}, z_{22}, \ldots, z_{2n}, \ldots)$ be i.i.d. normal random variables with mean zero and variance N_2. Let $N_1 < N_2$. At the ith transmission the real number x_i is sent and $y_{1i} = x_i + z_{1i}$, $y_{2i} = x_i + z_{2i}$ are received. In our analysis it is irrelevant whether z_{1i} and z_{2i} are correlated or not (although in the feedback case it may make a difference). Let there be a power constraint on the transmitted power, given for any n by

$$(1/n) \sum_{i=1}^{n} x_i^2 \leq S \tag{1}$$

for any signal $\underline{x} = (x_1, x_2, \ldots, x_n)$ of block length n.

It is well known that the individual capacities are $C_1 = \frac{1}{2}\log(1 + (S/N_1))$ and $C_2 = \frac{1}{2}\log(1 + (S/N_2))$ bits/transmission, where all logarithms are to the base 2. Time sharing will achieve any convex combination of (C_2, C_2) and $(C_1, 0)$, as shown in Fig. 3.

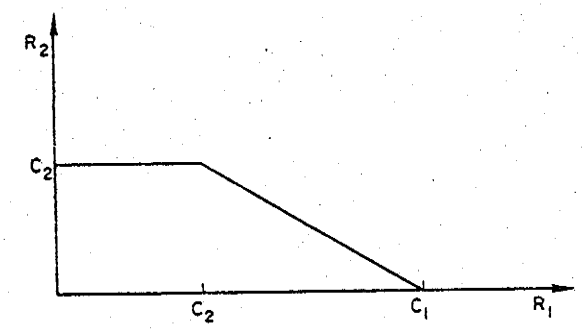

Fig. 3. Time-sharing rates for the Gaussian broadcast channel.

Now let us see how we can improve on this performance. Think of the signal $\underline{s}_2$ (intended for the high noise receiver Y_2) as a sequence of i.i.d. $N(0, \bar{\alpha}S)$ random variables. Superimposed on this sequence will be a sequence $\underline{s}_1$ that may be considered as a sequence of i.i.d. $N(0, \alpha S)$ random variables. Here $0 \leq \alpha \leq 1$ and $\bar{\alpha} = 1 - \alpha$. Thus the sequence $\underline{s} = \underline{s}_1 + \underline{s}_2$ will be a sequence of i.i.d. $N(0, S)$ random variables. The received sequences $\underline{y}_1 = \underline{s}_1 + \underline{s}_2 + \underline{z}_1$ and $\underline{y}_2 = \underline{s}_1 + \underline{s}_2 + \underline{z}_2$ are depicted in Fig. 4.

Now $\underline{s}_1$ and $\underline{z}_2$ are considered to be noise by receiver 2. We see that $s_{1i} + z_{2i}$ are i.i.d. $N(0, \alpha S + N_2)$ random

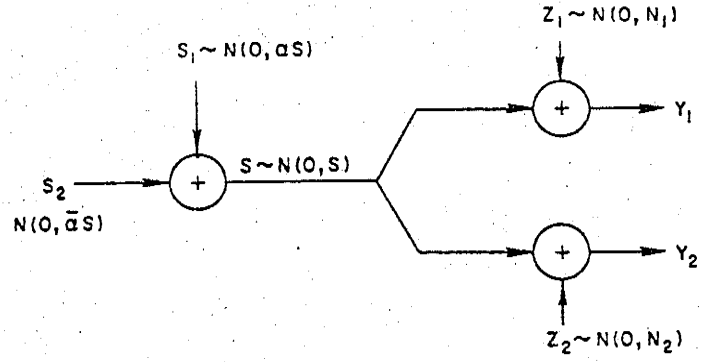

Fig. 4. Decomposition of the signal.

variables. Therefore, messages may be sent at rates less than

$$\frac{1}{2} \log \left(1 + \frac{\bar{\alpha} S}{\alpha S + N_2} \right) \triangleq C_2(\alpha) \tag{2}$$

to receiver Y_2 with probability of error near zero for sufficiently large block length n. That is, there exists a sequence of $\left(2^{n(C_2(\alpha)-\varepsilon)}, n\right)$ codes with average power constraint $\bar{\alpha} S$ and probability of error $\bar{p}_2^{(n)}(e) \to 0$.

Now, since $N_1 < N_2$, receiver Y_1 may also correctly determine the transmitted sequence $\underline{s}_2$ with arbitrarily low probability of error. Upon decoding of $\underline{s}_2$, given $\underline{y}_1$, receiver Y_1 then subtracts $\underline{s}_2$ from $\underline{y}_1$, yielding $\tilde{\underline{y}}_1 = \underline{y}_1 - \underline{s}_2 = \underline{s}_1 + \underline{z}_1$. At this stage channel 1 may be considered to be a Guassian channel with input power constraint αS and additive zero mean Gaussian noise with variance N_1. The capacity of this channel is $\frac{1}{2} \log(1 + (\alpha S/N_1)) = \tilde{C}_1(\alpha)$ bits/transmission and is achieved, roughly speaking, by choosing $2^{n\tilde{C}_1(\alpha)}$ independent n-sequences of i.i.d. $N(0,\alpha S)$ random variables as the code set for the possible sequences $\underline{s}_1$. Thus receiver Y_1 correctly receives both $\underline{s}_1$ and $\underline{s}_2$.

This informal argument indicates that rates

$$R_1 = \frac{1}{2} \log \left(1 + \frac{\bar{\alpha} S}{\alpha S + N_2} \right) + \frac{1}{2} \log \left(1 + \frac{\alpha S}{N_1} \right)$$

$$R_2 = \frac{1}{2} \log \left(1 + \frac{\bar{\alpha} S}{\alpha S + N_2} \right) \tag{3}$$

may simultaneously be ε-achieved for any $0 \le \alpha \le 1$. These rate pairs, shown in Fig. 5, dominate the time-sharing rates.

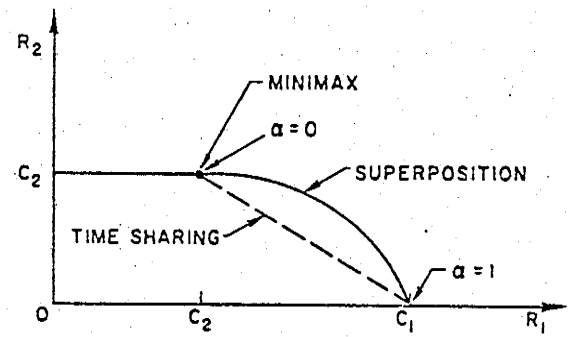

Fig. 5. Set of achievable rates for the Gaussian broadcast channel.

Summarizing the argument, we select a set of $2^{n(C_2(\alpha)-\varepsilon)}$ random n-sequences of i.i.d. $N(0,\alpha S)$ random variables and a set of $2^{n(\tilde{C}_1(\alpha)-\varepsilon)}$ random n-sequences of i.i.d. $N(0,\bar{\alpha} S)$ random variables. Now $2^{n(\tilde{C}_1(\alpha)+C_2(\alpha)-2\varepsilon)}$ n-sequences are formed by adding together pairs of sequences, in which the first sequence is chosen from the first set and the second sequence is chosen from the second set, and the pairs are chosen in all possible ways. A message

$$(r, s_1), \quad r \in \left\{1, 2, \ldots, 2^{n(C_2(\alpha)-\varepsilon)}\right\}, \quad s_1 \in \left\{1, 2, \ldots, 2^{n(\tilde{C}_1(\alpha)-\varepsilon)}\right\}$$

is transmitted by selecting the n-sequence corresponding to

the sum of the rth sequence in the first set and the s_1th sequence in the second set. Receiver 1 is intended to decode (r,s_1) correctly and receiver 2 is intended to decode r correctly, thus simultaneously achieving rates

$$R_1 = \tilde{C}_1(\alpha) + C_2(\alpha) - 2\varepsilon$$
$$R_2 = C_2(\alpha) - \varepsilon \tag{4}$$

as given in (3).

If $N_1 = 0$, and channel 1 is therefore perfect, we have $C_1 = \infty$ and $C_2 = \frac{1}{2}\log(1+(S/N_2))$. A compound channel or maximin approach would have us send at rates $(R_1,R_2) = (C_2,C_2)$. An arbitrarily small decrement in the rate for channel 2, corresponding to $0 < \alpha \ll 1$ in (3), yields $(R_1,R_2) = (\infty, C_2 - \varepsilon)$ as a pair of achievable rates. Although this rate pair does not dominate (C_2,C_2), it seems preferable.

The achievability of this Gaussian rate region was shown by Cover [1] and is a special case of Bergmans's theorem on degraded channels [2]. The proof of the converse theorem establishing this region as the capacity region is given by Bergmans [7]. Implications of the rate region to broadcasting are given by Bergmans and Cover [8], who show that the usual methods of time- and frequency-sharing are suboptimal. Some new ideas on multiaccess Gaussian channels will be treated in Section 8.

III. BROADCAST CHANNEL FORMULATION

We shall define a two-receiver memoryless <u>broadcast channel</u>, denoted by $(X, p(Y_1,Y_2 \mid x), Y_1 \times Y_2)$ or by $p(y_1,y_2 \mid x)$, to consist of three finite sets X, Y_1, Y_2 and a collection of probability distributions $p(\cdot,\cdot \mid x)$ on $Y_1 \times Y_2$, one for each $x \in X$. The interpretation is that x is an input to the channel and y_1 and y_2 are the respective outputs at receiver terminals

1 and 2 as shown in Fig. 6. The problem is to communicate

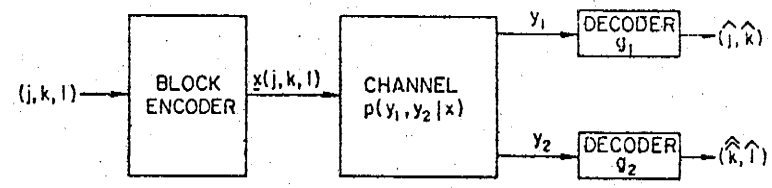

Fig. 6. Two-receiver broadcast channel.

simultaneously with receivers 1 and 2 as efficiently as possible.

The nth extension for a broadcast channel is the broadcast channel

$$\left(X^n,\ p(\underline{y}_1,\underline{y}_2 \mid \underline{x}),\ Y_1^n \times Y_2^n\right) \tag{5}$$

where $p(\underline{y}_1,\underline{y}_2 \mid \underline{x}) = \Pi_{j=1}^n\ p(y_{1j},y_{2j} \mid x_j)$, for $\underline{x} \in X^n$, $\underline{y}_1 \in Y_1^n$, $\underline{y}_2 \in Y_2^n$.

An $((M_{11},M_{22},M_{12}),n)$ code for a broadcast channel consists of three sets of integers

$$\underset{\sim}{M}_{11} = \{1,2,\ldots,M_{11}\}, \quad \underset{\sim}{M}_{12} = \{1,2,\ldots,M_{12}\},$$
$$\underset{\sim}{M}_{22} = \{1,2,\ldots,M_{22}\} \tag{6}$$

an encoding function

$$\underline{x} : \underset{\sim}{M}_{11} \times \underset{\sim}{M}_{12} \times \underset{\sim}{M}_{22} \to X^n \tag{7}$$

and two decoding functions

$$g_1 : Y_1^n \to \underset{\sim}{M}_{11} \times \underset{\sim}{M}_{12}, \qquad g_1(\underline{y}_1) = (\hat{j},\hat{k})$$
$$g_2 : Y_2^n \to \underset{\sim}{M}_{12} \times \underset{\sim}{M}_{22}, \qquad g_2(\underline{y}_2) = (\hat{\hat{k}},\hat{\ell}) \tag{8}$$

The set $\{\underline{x}(j,k,\ell) \mid (j,k,\ell) \in \underset{\sim}{M}_{11} \times \underset{\sim}{M}_{12} \times \underset{\sim}{M}_{22}\}$ is called the "set of codewords." As illustrated in Fig. 6, we think of integers j and ℓ as being arbitrarily chosen by the transmitter to be sent to receivers 1 and 2, respectively. The integer k is also chosen by the transmitter and is intended to be received by both receivers. Thus k is the "common" part of the message, and j and ℓ are the "independent" parts of the message.

If the message (j,k,ℓ) is sent, let

$$\lambda(j,k,\ell) = \Pr\{g_1(\underline{Y}_1) \neq (j,k) \quad \text{or} \quad g_2(\underline{Y}_2) \neq (k,\ell)\} \tag{9}$$

denote the probability of error, where we note that $\underline{Y}_1, \underline{Y}_2$ are the only chance variables in the above expression. We define the "<u>average probability of error</u>" of the code as

$$P(e) = P_n(e) = (1/M) \sum_{j,k,\ell} \lambda(j,k,\ell) \tag{10}$$

where

$$M = M_{11} M_{22} M_{12} \tag{11}$$

This probability of error is calculated under a special distribution; namely, the uniform distribution over the codewords.

We shall define the <u>rate</u> (R_{11}, R_{12}, R_{22}) of an $((M_{11}, M_{12}, M_{22}), n)$ code by

$$\begin{aligned} R_{11} &= (1/n) \log M_{11} \\ R_{12} &= (1/n) \log M_{12} \\ R_{22} &= (1/n) \log M_{22} \end{aligned} \tag{12}$$

all defined in bits per transmission. Thus R_{ii} is the rate of transmission of independent information to receiver i, for $i = 1,2$, and R_{12} is the portion of the information common to both receivers.

DEFINITION. The rate (R_{11}, R_{12}, R_{22}) is said to be <u>achievable</u> by a broadcast channel if, for any $\varepsilon > 0$ and for all n sufficiently large, there exists an $((M_{11}, M_{12}, M_{22}), n)$ code with

$$M_{11} \geq 2^{nR_{11}}, \qquad M_{12} \geq 2^{nR_{12}}, \qquad M_{22} \geq 2^{nR_{22}} \tag{13}$$

such that $P_n(e) < \varepsilon$.

COMMENT. Note that the total number $M = M_{11} M_{12} M_{22}$ of codewords for a code satisfying (13) must exceed $2^{n(R_{11}+R_{12}+R_{22})}$.

DEFINITION. The <u>capacity region</u> $\mathcal{R}^*$ for a broadcast channel is the set of all achievable rates (R_{11}, R_{12}, R_{22}).

The extension of the definition of the broadcast channel from 2 receivers to k receivers is notationally cumbersome but straightforward, given the following comment. The index sets M_{11}, M_{12}, M_{22} are replaced by $2^k - 1$ index sets $I(\Theta)$, $\Theta \in \{0,1\}^k$, $\Theta \neq 0$, with the interpretation that the integer $i(\Theta)$ selected in index set $I(\Theta) = \{1,2,\ldots,M(\Theta)\}$ is intended (by the proper code selection) to be received correctly by every receiver j for which $\Theta_j = 1$ in $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_k)$. Then, for example, the rate of transmission over the nth extension of a broadcast channel to the ith receiver will be given by

$$R_i = (1/n) \log \sum_{\Theta \in \{0,1\}^k, \Theta_i = 1} M(\Theta) = (1/n) \sum_{\Theta_i = 1} \log M(\Theta) \tag{14}$$

In the two-receiver broadcast channel, the corresponding sets in the new notation are $M_{12} = I(1,1)$, $M_{11} = I(1,0)$, $M_{22} = I(0,1)$.

IV. REVIEW OF SOME EXTREME BROADCAST CHANNELS

A. Orthogonal Channels

Consider a broadcast channel in which efficient communication to one receiver in no way interferes with communication to the other. A movie designed to be shown simultaneously to a blind person and a deaf person would be such an example.

Consider the broadcast channel with $X = \{1,2,3,4\}$, $Y_1 = \{1,2\}$, $Y_2 = \{1,2\}$, with channel matrices

$$P_1 = \begin{vmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{vmatrix}, \qquad P_2 = \begin{vmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{vmatrix} \tag{15}$$

We easily calculate $C_1 = C_2 = 1$ bit/transmission. Clearly, from the standpoint of receiver y_1, inputs $x = 1$ and $x = 2$ both result in $y_1 = 1$ with probability 1 and can therefore be merged. Proceeding with this analysis, we find that Y_1 can determine only $x \in \{1,2\}$ versus $x \in \{3,4\}$, while Y_2 can determine only $x \in \{1,3\}$ versus $x \in \{2,4\}$. Thus, as shown by Cover [1], all rates $0 \le R_1 \le C_1 = 1$, $0 \le R_2 \le C_2 = 1$, for arbitrary $0 \le R_{12} \le 1$ are achievable.

B. Incompatible Channels

Let

$$X = \{1,2,3,4\}, \qquad Y_1 = \{1,2\}, \qquad Y_2 = \{1,2\}$$

and let

$$P_1 = \begin{vmatrix} 1 & 0 \\ 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{vmatrix}, \qquad P_2 = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \\ 0 & 1 \end{vmatrix} \tag{16}$$

be the channel matrices. Thus if X wishes to communicate with Y_1 over the perfect channel $x \in \{1,2\} \to Y_1$, he must send pure noise to Y_2, i.e., $\Pr\{Y_2 = 1 \mid x \in \{1,2\}\} = \frac{1}{2}$. A similar statement holds for X communicating with Y_2.

Here it can be shown that only rates $0 \le R_1 + R_2 \le 1$ can be achieved. This is an example in which the two channels are so incompatible that one can do no better than time sharing, i.e., using one channel efficiently part of the time and the other channel the remainder.

C. Bottleneck Channel

Consider the broadcast channel in which the two channels have the same structure, i.e.,

$$p_1(y_1 \mid x) = p_2(y_2 \mid x), \qquad \forall x \in X, \ \forall y_1, y_2 \in Y_1 = Y_2 = Y$$

We shall term this the "bottleneck channel."

Here, we note that any code for receiver Y_1 is also a code with the same error properties for receiver Y_2. Thus Y_1 and Y_2 both perceive correctly the transmitted sequence x with low probability of error.

Let the capacity of channel P be denoted by $C_1 = C_2 = C$ bits/transmission. Now, since both receivers receive the same information about X, it follows that both receivers 1 and 2 will be able to recover correctly r, s_1, and s_2 if and only if (R_1, R_2, R_{12}) is an achievable rate. The capacity region is given by Cover [1] as

$$R_1 + R_2 - R_{12} \le C, \quad 0 \le R_1 < C, \quad 0 \le R_2 < C, \quad 0 \le R_{12} < C \tag{17}$$

Let us now compare the orthogonal channel with the bottleneck channel. The orthogonal channel achieves $(R_1, R_2) = (1,1)$ with arbitrary joint rate $0 \le R_{12} \le 1$. Thus fully independent

messages $(R_{11} = 0)$ or maximally dependent messages $(R_{11} = 1)$ can be sent simultaneously to receivers 1 and 2.

At the other extreme, in the case of the bottleneck channel with capacity $C = 1$, we can simultaneously achieve $R_1 = 1$, $R_2 = 1$. Here, however, it may be seen that achieving $(R_1, R_2) = (1,1)$ implies $R_{12} = 1$. Thus the messages sent to 1 and 2 must be maximally dependent, and in fact equal.

V. DEGRADED BROADCAST CHANNELS

The degraded broadcast channel models the situation in which one receiver is "farther away" than the other receiver. 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 $\mathcal{A}$, output alphabet $\mathcal{B}_1$, and transition probability $p_1(y_1 \mid x)$, and let A_2 be another channel with same input alphabet $\mathcal{A}$, output alphabet $\mathcal{B}_2$, and transition probability $p_2(y_2 \mid x)$. The degradation is expressed by

$$p_2(y_2 \mid x) = \sum_{y_1 \in \mathcal{B}_1} p_3(y_2 \mid y_1)\, p_1(y_1 \mid x) \tag{18}$$

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

By definition, if every component channel A_j of a broadcast channel is a degraded version of A_{j-1} $(j = 1,\ldots,N)$, the broadcast channel will be called "degraded". We can represent a degraded broadcast channel as a cascade formed by the best channel A_1, followed by successive degrading channels $D_2, D_3, \ldots, D_N$.

The capacity region for the degraded broadcast channel is

now completely understood. The achievability of a certain natural region had been conjectured [1] and has been proved by Bergmans [2] in full generality, including the continuous alphabet case. Wyner and Ziv [9,10] proved the converse for certain degraded binary symmetric channels, Bergmans [7] proved the converse for Gaussian channels, and Gallager [3] proved the converse completely for general degraded channels. A subsequent alternative proof of the converse can be found in the work of Ahlswede [11].

Bergmans [2] considers the following random code for the N receiver degraded broadcast channel. 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 $q_i(x_i \mid x_{i+1})$. The artificial satellizing channels are cascaded, and the broadcast channel is of the degraded type.

THEOREM 1. (Bergmans and Gallager) The capacity region is given by all $(R_1, R_2, \ldots, R_N)$ such that there exists a probability assignment $q_1, q_2, \ldots, q_N$ such that

$$
\begin{aligned}
R_N &< I_{q_1 \cdots q_N}(X_N;\ Y_N) \\
R_{N-1} &< I_{q_1 \cdots q_N}(X_{N-1};\ Y_{N-1} \mid X_N) \\
&\ \ \vdots \\
R_2 &< I_{q_1 \cdots q_N}(X_2;\ Y_2 \mid X_3) \\
R_1 &< I_{q_1 \cdots q_N}(X_1;\ Y_1 \mid X_2)
\end{aligned}
\tag{19}
$$

VI. AN ACHIEVABLE REGION FOR THE GENERAL BROADCAST CHANNEL

The capacity region for the general one-sender two-receiver broadcast channel is not now known. Van der Meulen [4] and Cover [5], in as yet unpublished research, however, have independently put forth the achievable rate region described in this section. The development in this section follows Cover [5].

Let $p(y_1,y_2 \mid x)$ denote a discrete memoryless channel with single source X and two independent receivers Y_1 and Y_2. We exhibit an achievable region of rates (R_{11},R_{12},R_{22}) at which independent information can be sent, respectively, to receiver 1, to both receivers 1 and 2, and to receiver 2. The achievability of the region is shown by using a version of the asymptotic equipartition property involving many simultaneous "typicality" constraints. These results immediately generalize to yield an achievable rate region for the m-sender n-receiver channel in terms of standard mutual information quantities.

First we shall define three auxiliary random variables U,R,V taking values in finite sets $\mathcal{U},\mathcal{R},\mathcal{V}$. Let $x:\mathcal{U}\times\mathcal{R}\times\mathcal{V}\to\mathcal{X}$ denote an arbitrary mapping of the auxiliary random variables into the input alphabet. The picture we have in mind is given in Fig. 7.

For each assignment of probability distributions $p(u)$,

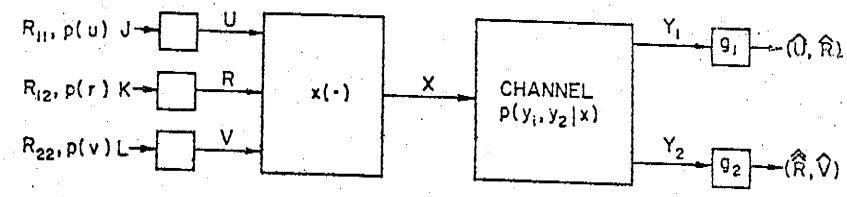

Fig. 7. Auxiliary random variables.

$p(r)$, $p(v)$ and mapping function $x(\cdot)$, we associate the joint probability distribution function

$$p(u,r,v,y_1,y_2) = p(u)p(r)p(v)p(y_1,y_2 \mid x(u,r,v)) \tag{20}$$

Mutual information quantities like

$$I(U,R;Y_1) = \sum_{u,r,y_1} p(u,r,y_1)\ \log[p(u,r,y_1)/p(u,r)p(y_1)] \tag{21}$$

are defined in the usual way.

Define

$$
\begin{aligned}
\underline{I} &= (I_1,I_2,\ldots,I_6) \\
&= \big(I(U;R,Y_1),I(V;R,Y_2),I(R;U,Y_1),I(R;V,Y_2),I(U,R;Y_1),I(V,R;Y_2)\big)
\end{aligned}
\tag{22}
$$

Let $\mathcal{I}$ denote the set of all $\underline{I} \in \mathbb{R}^6$ generated by all assignments of $p(u)$, $p(r)$, $p(v)$, $x(\cdot)$. Let $C_0(\mathcal{I})$ denote the convex hull of $\mathcal{I}$. Let $R(I)$ denote the set of all $(R_{11},R_{12},R_{22}) \in \mathbb{R}^3$ satisfying the six inequalities

$$
\begin{aligned}
R(\underline{I}):\quad & R_{11} < I_1,\quad R_{22} < I_2,\quad R_{12} < I_3,\quad R_{12} < I_4 \\
& R_{11}+R_{12} < I_5,\quad R_{12}+R_{22} < I_6
\end{aligned}
\tag{23}
$$

THEOREM 2. The region

$$\mathcal{R} = \sum_{\underline{I} \in C_0(\mathcal{J})} R(\underline{I}) \tag{24}$$

is achievable.

We can express the capacity region in another form. Observe that an arbitrary point $\underline{I}$ on the boundary of $C_0(\mathcal{J})$ can always be expressed as the convex combination of no more than six (extreme) points of $\mathcal{J}$. For a given $\underline{I}$, let

$$\underline{I} = \sum_{i=1}^{6} p(q^{(i)})\, \underline{I}_{q^{(i)}}, \qquad \underline{I}_{q^{(i)}} \in \mathcal{J} \tag{25}$$

be the desired convex combination, where $q^{(i)} = (p(u \mid q^{(i)}), p(r \mid q^{(i)}), p(v \mid q^{(i)}), x_{q^{(i)}}(\cdot))$ is an element in the set of all assignments $(p, x(\cdot))$ and $\underline{I}_{q^{(i)}}$ is the vector of mutual informations induced by this assignment. Let Q denote a random variable with $\Pr\{Q = q^{(i)}\} = p(q^{(i)})$, $p(q^{(i)}) \geq 0$, $\sum p(q^{(i)}) = 1$, $i = 1,2,\ldots,6$. It then follows from inspection of the definition of mutual information that, for example,

$$\sum_{i=1}^{6} p(q^{(i)})\, I_{q^{(i)}}(U;R,Y_1) = I(U;R,Y_1 \mid Q) \tag{26}$$

Thus $\mathcal{R}$ can be expressed as follows.

THEOREM 2. $\mathcal{R}$ is the union of all $(R_{11}, R_{12}, R_{22}) \in \mathbb{R}^3$ satisfying the inequalities

$$\begin{aligned} &R_{11} < I(U;R,Y_1 \mid Q), && R_{22} < I(V;R,Y_2 \mid Q), \\ &R_{12} < I(R;V,Y_1 \mid Q), && R_{12} < I(R;V,Y_2 \mid Q) \\ &R_{11} + R_{12} < I(U,R;Y_1 \mid Q), && R_{12} + R_{22} < I(V,R;Y_2 \mid Q) \end{aligned} \tag{27}$$

where the union is over all random variables Q such that Q takes on no more than six values in the set of all assignments

$(p, x(\cdot))$. This is the same region as in Theorem 2.

Before proceeding with the argument, we shall define a simultaneous asymptotic equipartition property. This will allow us to decode messages at the receiver by simply checking to see which of the possible input messages is jointly "typical" with the received output. If there is one and only one such candidate, we shall declare this to be the message.

Let $\{X^{(1)}, X^{(2)}, \ldots, X^{(k)}\}$ denote a finite collection of discrete random variables with some fixed joint distribution $p(x^{(1)}, x^{(2)}, \ldots, x^{(k)})$. Let S denote an ordered subset of these random variables, and consider n independent copies of S. Thus,

$$\Pr\{\underline{S} = \underline{s}\} = \prod_{i=1}^{n} \Pr\{S_i = s_i\} \tag{28}$$

For example, if $S = (X^{(j)}, X^{(k)})$, then

$$\begin{aligned} \Pr\{\underline{S} = \underline{s}\} &= \Pr\left\{\left(\underline{X}^{(j)}, \underline{X}^{(k)}\right) = \left(\underline{x}^{(j)}, \underline{x}^{(k)}\right)\right\} \\ &= \prod_{i=1}^{n} p\left(x_i^{(j)}, x_i^{(k)}\right) \end{aligned} \tag{29}$$

We know by the law of large numbers that for a given subset S of random variables,

$$-(1/n) \log p(S_1, S_2, \ldots, S_n) = -\sum_{i=1}^{n} \log p(S_i) \to H(S) \tag{30}$$

with probability 1. This convergence takes place simultaneously with probability one for all 2^k subsets

$$S \subseteq \{X^{(1)}, X^{(2)}, \ldots, X^{(k)}\} \tag{31}$$

DEFINITION. The set A_ε of <u>jointly ε-typical</u> n-sequences $(\underline{x}^{(1)}, \underline{x}^{(2)}, \ldots, \underline{x}^{(k)})$ is defined by

$$A_\varepsilon\left(X^{(1)},X^{(2)},\ldots,X^{(k)}\right)$$

$$A_\varepsilon = \left\{\left(\underline{x}^{(1)},\underline{x}^{(2)},\ldots,\underline{x}^{(k)}\right) \in \left(\mathcal{X}^{(1)}\right)^n \times \left(\mathcal{X}^{(2)}\right)^n \times \cdots \times \left(\mathcal{X}^{(k)}\right)^n : \left|-(1/n)\log p(\underline{S}) - H(S)\right| \le \varepsilon,\ \forall S \subseteq \left\{X^{(1)},X^{(2)},\ldots,X^{(k)}\right\}\right\}, \tag{32}$$

where $\underline{S}$ denotes the ordered set of sequences in $\{\underline{x}^{(1)},\ldots,\underline{x}^{(k)}\}$ corresponding to S. Let $A_\varepsilon(S)$ denote the restriction of A_ε to the coordinates corresponding to S.

Thus, for example,

$$A_\varepsilon\left(X^{(1)},X^{(2)}\right) = \left\{\left(\underline{x}^{(1)},\underline{x}^{(2)}\right) : \left|-(1/n)\log p\left(\underline{x}^{(1)},\underline{x}^{(2)}\right) - H\left(X^{(1)},X^{(2)}\right)\right| < \varepsilon : \left|-(1/n)\log p\left(\underline{x}^{(1)}\right) - H\left(X^{(1)}\right)\right| < \varepsilon : \left|-(1/n)\log p\left(\underline{x}^{(2)}\right) - H\left(X^{(2)}\right)\right| < \varepsilon\right\} \tag{33}$$

The proof requires a bound on the probability that conditionally independent sequences are jointly typical. Let the discrete random variables W,Z,Q have joint distribution $p(w,z,q)$. Let W',Z' be conditionally independent given Q, with the marginals

$$p(w|q) = \sum_z p(w,z,q)/p(q), \qquad p(z|q) = \sum_w p(w,z,q)/p(q) \tag{34}$$

The unconditional version of the following lemma has been observed and proved by Forney [12] as crucial in giving the natural proof of Shannon's second theorem. This lemma has also been used by the author on source compression for dependent ergodic sources [13].

LEMMA 1. Let $(\underline{W},\underline{Z},\underline{Q}) \sim \Pi_{i=1}^n p(w_i,z_i,q_i)$ and $(\underline{W}',\underline{Z}',\underline{Q}) \sim \Pi_{i=1}^n p(w_i|q_i)p(z_i|q_i)p(q_i)$. Then, for n such that $\Pr\{A_\varepsilon\} > 1-\varepsilon$,

$$(1-\varepsilon)\, 2^{-n(I(W;Z|Q)+7\varepsilon)} \le \Pr\left\{(\underline{W}',\underline{Z}',\underline{Q}) \in A_\varepsilon(W,Z,Q)\right\} \le 2^{-n(I(W;Z|Q)-7\varepsilon)} \tag{35}$$

For any $\underline{I} \in C_0(\mathcal{S})$, we shall show how to achieve any rate (R_{11},R_{12},R_{22}) satisfying the six inequalities in Eq. (27). Consider a given assignment $p(u|q)$, $p(r|q)$, $p(v|q)$, $x(\cdot|q)$, $p(q)$, $q \in \{q^{(1)},q^{(2)},\ldots,q^{(6)}\}$ and the associated $\underline{I} \in C_0(\mathcal{S})$ given in Eq. (26).

A. Random Encoding for Theorem 2'

First, generate a sequence of n i.i.d. random variables $\underline{Q} = (Q_1,Q_2,\ldots,Q_n)$. Here, $\underline{Q} = \underline{q}$ plays the role of a time-sharing parameter which at each time k informs the transmitter and both receivers that the mode of operation is $Q_k = q_k$, where q_k is one of the six modes in $\{q^{(1)},q^{(2)},\ldots,q^{(6)}\}$. Conditioned on $\underline{Q} = \underline{q}$, generate $2^{nR_{11}}$ random n-sequences of random variables drawn according to $p(\underline{u}|\underline{q})$; $2^{nR_{22}}$ random n-sequences of random variables drawn according to $p(\underline{v}|\underline{q})$. Index the strings by $j = 1,2,\ldots,2^{nR_{11}}$, $k = 1,2,\ldots,2^{nR_{12}}$, $\ell = 1,2,\ldots,2^{nR_{22}}$, respectively. Thus, for example, the jth n-sequence (word) $\underline{U}(j)$ has probability

$$\Pr\left\{\underline{U}(j) = \underline{u}(j) \,|\, \underline{Q} = \underline{q}\right\} = \prod_{i=1}^n p(u_i(j)|q_i) \tag{36}$$

Also, $\underline{U}(j)$, $\underline{R}(k)$, $\underline{V}(\ell)$, $\forall j,k,\ell$, are conditionally independent given $\underline{Q}$.

To each (j,k,ℓ) there corresponds a triple of n-sequences $(\underline{u}(j),\underline{r}(k),\underline{v}(\ell))$ and the codeword

$$\underline{x}(j,k,\ell) = \left(x_1(j,k,\ell), x_2(j,k,\ell), \ldots, x_n(j,k,\ell)\right) \tag{37}$$

where

$$x_m(j,k,\ell) = x\left(u_m(j), r_m(k), v_m(\ell) \,|\, q_m\right) \tag{38}$$

The codebook consists of the M n-sequences $\underline{x}(i,j,k)$, $(j,k,\ell) \in \underset{\sim}{M}_{11} \times \underset{\sim}{M}_{12} \times \underset{\sim}{M}_{22}$.

B. Decoding Rule

Both receivers know $\underline{q}$. If $\underline{y}_1$ is received, declare that $(\hat{j},\hat{k}) = (j,k)$ was sent if there is one and only one pair $(j,k) \in \underset{\sim}{M}_{11} \times \underset{\sim}{M}_{12}$ such that $(\underline{u}(j), \underline{r}(k), \underline{y}_1, \underline{q}) \in A_\varepsilon(U,R,Y_1,Q)$, i.e., if there is only one input pair (j,k) that is jointly typical with the output.

If $\underline{y}_2$ is received, declare $(\hat{\hat{k}},\hat{\ell}) = (k,\ell)$ was sent if there is one and only one pair (k,ℓ) such that $(\underline{r}(k), \underline{v}(\ell), \underline{y}_2, \underline{q}) \in A_\varepsilon(R,V,Y_2,Q)$.

The proof of the achievability of $\mathcal{R}$ follows from Lemma 1 which allows bounding error probabilities for error events like $E(j,k,\ell) : (\underline{U}(j), \underline{R}(k), \underline{Y}_1, \underline{Q}) \in A_\varepsilon(U,R,Y_1,Q)$. Details are provided in Cover [5].

VII. THE GENERAL MULTIUSER FRAMEWORK

A multiple-user channel, denoted by $(X_1 \times X_2 \times \cdots \times X_m, p(y_1,y_2,\ldots,y_n|x_1,x_2,\ldots,x_m), Y_1 \times Y_2 \times \cdots \times Y_n)$ is defined to consist of m finite sets X_i, $i = 1,2,\ldots,m$; n finite sets Y_j, $J = 1,2,\ldots,n$; and a collection of probability distributions $p(\cdot,\cdot,\ldots,\cdot|x_1,x_2,\ldots,x_m)$ on $Y_1 \times Y_2 \times \cdots \times Y_n$, one for each $(x_1,x_2,\ldots,x_m) \in X_1 \times \cdots \times X_n$. The interpretation is that $x_1,x_2,\ldots,x_m$ are the respective inputs from m senders and $y_1,y_2,\ldots,y_n$ are the respective outputs at receiver terminals $1,2,\ldots,n$. The channel is assumed to be memoryless. It is assumed that the messages are independent. Let $R_i(S)$,

$S \subseteq \{1,2,\ldots,n\}$ be the rate at which independent information is sent from sender i to precisely the receivers $j \in S$. The problem is to characterize the capacity region $\mathcal{R}^*$ of simultaneously achievable rates.

Most network communication problems are of the above type. An exception to this formulation is the 2-way channel of Shannon [14] in which subsequent channel uses for transmitter i may depend on the past received signal for receiver i.

CASE 1. $m = n = 1$. This is the single channel problem solved by Shannon [15] in 1948.

CASE 2. $m \geq 2$, $n = 1$. This is the multiaccess channel completely solved by Ahlswede [16] and Liao [17]. The Gaussian version of this channel will be treated in the next section.

CASE 3. $m = 1$, $n \geq 2$. This is the broadcast channel, the primary subject of this chapter.

CASE 4. $m \geq 2$, $n \geq 2$. This is the general multiuser network communication problem without feedback. Ulrey [18] has some results on this, which unfortunately do not include the broadcast formulation. Ulrey's results, however, include the previously known multiaccess work and shed some light on achievable rate regions for the broadcast channel (see also van der Meulen [4]).

VIII. THE MULTIACCESS GAUSSIAN CHANNEL

This section describes a simple optimal collection of coding schemes for the problem of several independent transmitters attempting to communicate information under a power constraint to one receiver using a common frequency band of bandwidth W. The resulting capacity region dominates that which can be achieved by dividing the band into independent subchannels for the various transmitters.

Liao [17] and Ahlswede [16] have solved the problem of multiple access for two transmitters and one receiver in the finite alphabet case. We consider the continuous amplitude signal case, but it would be correct to say that these results are an application of the theorem of Liao and Ahlswede, when the power constraint is included in the proof in a straightforward way.

The proof of optimality given here could have been achieved in 1948. The proof makes use of the special properties of Gaussian channels to obviate the technical details necessary to prove optimality in the finite-alphabet case of Liao and Ahlswede. By coincidence, this result for the Gaussian channel was first presented in adjacent talks in a session in a Communication Theory Workshop, Solvang, California, 1973, by A. Wyner and T. Cover. Wyner's presentation appears elsewhere in [6] and the author's presentation appears here. Other work on encoding sources for multiaccess channels can be found in [18,20,21,25].

The basic idea is that all transmitters transmit at once at different rates at constant power over the entire bandwidth. Decoding consists of first finding the signal sent by the first transmitter, subtracting it out, then finding the signal transmitted by the second transmitter, subtracting it out, and so on. We demonstrate that the sum of the rates is as high as if all the transmitters cooperated and pooled their power initially for the use of an omniscient super transmitter. The code books that are used are no different than those that are used in the normal Gaussian channel case of signaling in the presence of white additive Gaussian noise. No cooperation of the sources is required.

Suppose that a transmitter has power P_1 and must communicate over an additive white Gaussian noise (AWGN) channel of bandwidth W and noise power spectral density N. Then

Shannon's basic theorem implies that the transmitter can send at rates R_1 up to

$$C_1 = W \ln\left(1 + (P_1/NW)\right) \quad \text{nats/sec} \tag{39}$$

Now suppose that another transmitter comes on the scene, finds the first transmitter is sending at channel capacity, but also wishes to send information to the same receiver. Naive considerations suggest that a decrease in the rate (or power, or time, or frequency bandwidth) for the first receiver must be agreed on before the second transmitter can send any information. This is not the case, as the following analysis shows.

Let the second transmitter use the channel as if P_1 were noise power. Transmitter 2 then sends at rate

$$R_2 = W \ln\left[1 + \left(P_2/(NW + P_1)\right)\right] \tag{40}$$

No change whatsoever is assumed in the transmitted waveform for transmitter 1, that is, transmitter 1 proceeds as if transmitter 2 is not transmitting on top of him.

Before investigating the decoding, note the pretty fact that

$$\begin{aligned} R_1 + R_2 &= W \ln\left((P_1 + NW)/NW\right) + W \ln\left((P_1 + NW + P_2)/(P_1 + NW)\right) \\ &= W \ln\left(\frac{NW + P_1}{NW}\right)\left(\frac{P_1 + NW + P_2}{P_1 + NW}\right) \\ &= W \ln\left(1 + \frac{P_1 + P_2}{NW}\right) \end{aligned} \tag{41}$$

This is precisely the channel cpacity of a single channel with <u>combined</u> power $P_1 + P_2$.

The decoding at the receiver is simple. The receiver observes $\underline{y} = \underline{x}_1 + \underline{x}_2 + \underline{n}$. (We shall discuss the discrete time case.) Passage to continuous time follows the work of Gallager [19].) The receiver first duplicates the action of a receiver

for transmitter 2 and determines the code signal $\hat{\underline{x}}_2$ precisely (with probability of error $\leq \varepsilon$) and subtracts it from $\underline{y}$. He then duplicates the action of a receiver for transmitter 1 and chooses the closest code signal $\hat{\underline{x}}_1$ to $\underline{y} - \hat{\underline{x}}_2$. The receiver then declares $\hat{\underline{x}}_1$, $\hat{\underline{x}}_2$ to be the transmitted code signals.

The proof that the probabilities of error in each of these cases are arbitrarily near zero in the limit as time tends to infinity is very similar to the arguments of Cover [1], Bergmans [2], and especially Bergmans and Cover [8] and will not be repeated here. The outline is as follows. We shall use a random coding argument.

Generate a sequence of $n = 2TW$ independent identically distributed Gaussian random variables with mean zero and variance $\sigma^2 = P_1/2W$. This is the first code word in the randomly generated code book. Continue to independently generate such n-sequences until $e^{R_1 T}$ words are generated, thus constituting the code book for transmitter 1. Now generate $e^{R_2 T}$ independent identically distributed n-sequences, the components of which are i.i.d. zero mean Gaussian with variance $P_2/2W$. This is the code book for transmitter 2.

The transmission of information proceeds as follows: Receiver 1 chooses an index i from the set of integers $\{1,2,\ldots e^{R_1 T}\}$. Independently transmitter 2 chooses an index j from the set of integers $\{1,2,\ldots,e^{R_2 T}\}$. The corresponding waveforms $\underline{x}^{(i)}$ and $\underline{x}^{(j)}$ are then transmitted simultaneously over the channel. The received waveform is then $\underline{y} = \underline{x}^{(i)} + \underline{x}^{(j)} + \underline{n}$. The decoding procedure above is used.

Actually a more straightforward decoding system can be used, since the receiver will ultimately know (and must know) both signals $\underline{x}^{(i)}$ and $\underline{x}^{(j)}$. (This is distinct from the case studied by Bergmans and Cover [8] in which the noise receiver

powers were different for two receivers, thus obscuring for the worst receiver some of the information that was intended for the better receiver.) The overall scheme which will work is the following: Let $(\hat{i},\hat{j})$ be the pair of indices minimizing $\|y - x^{(i)} - x^{(j)}\|^2$.

The proof of the achievability of rates $R_1 < W \ln(1 + (P_1/NW))$, $R_2 < W \ln[1 + (P_2/(P_1 + NW))]$ follows simply from the arguments in Section II, using Lemma 1. Permuting the roles of transmitters 1 and 2 yields another set of rates, and time-sharing yields the line of rates between.

Now consider the general case of a possibly countably infinite number of transmitters with respective powers $P_1, P_2, \ldots$. Let S be an arbitrary subset of $\{1,2,\ldots\}$, and define

$$P(S) = \sum_{i \in S} P_i, \qquad R(S) = \sum_{i \in S} R_i \tag{42}$$

We then have the following theorem:

THEOREM 3. The capacity region for the multiaccess AWGN broadcast channel is given by the set of all $(R_1, R_2, \ldots)$ satisfying

$$R(S) \leq W \ln\left(1 + \frac{P(S)}{NW}\right) \tag{43}$$

for all subsets $S \subseteq \{1,2,\ldots\}$.

We see that we can send information at a total rate

$$\Sigma R_i = W \ln\left(1 + \frac{\Sigma P_i}{NW}\right) \tag{44}$$

which is equal to that which could be achieved if all the power were pooled and the best available code for the pooled source were used. Certainly sources acting separately, and independently, cannot hope to do better. Thus we see that k

transmitters can use the same channel to send to a single receiver as effectively as if they had pooled their power and information requirements a priori. This of course is a useful result if the transmitters are located at different geographical positions. There is a certain robustness to this procedure, and the independence of the transmitting signals is a nice bonus.

Finally, we consider the total received information when we have a number of users k tending to infinity, each transmitter having the same power P. Clearly then, the total information flow to the receiver which is being multiaccessed will then be $W \ln(1+(kP/NW))$, which tends to infinity asymptotically like $\ln k$. Thus the bandwidth limitation does not prevent the flow of information to the receiver from going to infinity as the number of users grows to infinity.

One might ask how this scheme might be implemented in practice using standard modulation techniques. For example, can one use an FM channel of a certain bandwidth and expect to have another transmitter come in on top of the first without degrading the performance of the first transmitter? We think the answer may be "no", at least for modulation schemes, and the reason is that the best estimates of transmitted signals in white noise are still linear estimates. There is no possibility of determining signal $\underline{x}_2$ without error and thus subtracting it off, leaving the received waveform free of error except for the noise $\underline{n}$. It may be true that the only way to use these results is with coding-theoretic transmission schemes. Here we think that superimposed convolutional codes will be very useful. (Bergmans has investigated this.) The Viterbi algorithm provides a very efficient means of decoding first $\underline{x}_2$, then $\underline{x}_1$, given $\underline{y}$.

It should be noted that although the technique of superposition of information which we have used here tends to give

strict dominance over time sharing and frequency sharing [1,2,8], there is at least one point on the frequency multiplex curve which lies on the boundary in the multiaccess problem treated here. If the available bandwidth is divided up into k disjoint frequency bands, proportional to the powers $P_1, P_2, \ldots, P_k$, respectively, then a point on the boundary $R = W \ln(1+(\Sigma P_i/NW))$ can be achieved. Apart from some nuisance problems with guard bands, which have been neglected, there is a somewhat more serious practical objection that this requires that the powers and the number of users of the multiaccess channel must be known a priori and the band allocations must be adhered to. Thus if several of the users fall off the system, the total rate is seriously diminished below capacity because of the nonallocation of certain frequency channels.

IX. CONCLUDING REMARKS

The m-sender n-receiver network capacity region is currently unknown, although the $m \times 1$ region is known, and the $1 \times n$ region is known for the special case of degraded broadcast channels. When the $1 \times n$ case is completely understood, it seems likely that the entire $m \times n$ solution will easily follow, because of the compatibility of the $m \times 1$ and $1 \times n$ formulations.

The basic problem motivating research on the broadcast $1 \times n$ channel is somewhat larger even than the formulation given in Section III. The problem, mentioned by Cover [1, p.8] is as follows.

Let $p(y_1, \ldots, y_n | x)$ be a broadcast channel with n receivers. Consider n discrete-time finite-alphabet stochastic processes $\{V_j^{(i)}\}_{j=1}^{\infty}$, $i = 1, 2, \ldots, n$. Suppose that $V^{(1)}, V^{(2)}, \ldots, V^{(n)}$ are jointly ergodic. It is desired to send the process $V^{(i)} = (V_1^{(i)}, V_2^{(i)}, \ldots)$ to receiver i

simultaneously for $i = 1,2,\ldots,n$, with arbitrarily small probability of error. What is the necessary and sufficient characterization of the entropy of $V^{(i)}$, $i = 1,2,\ldots,n$, and capacity region of $p(y_1,y_2,\ldots,y_n|x)$ under which this transmission can be achieved? Shannon gives the answer for one process and one receiver; namely, if the entropy of the process $H(V)$ is less than the capacity of the channel C, then asymptotically error free transmission of V to the output can be achieved. The general answer awaits more detailed knowledge of the decomposition of the stochastic processes as well as the solution of the broadcast channel.

REFERENCES

1. T. M. Cover, Broadcast channels, IEEE Trans. Inform. Theory, IT-18, 2-14 (Jan., 1972). Reprinted in Key Papers in the Development of Information Theory, D. Slepian, (ed.), pp. 437-449. IEEE Press, N.Y., N.Y. 1974.

2. P. P. Bergmans, Random coding theorem for broadcast channels with degraded components, IEEE Trans. Inform. Theory, 19, 197-207 (March, 1974).

3. R. G. Gallager, "Coding for degraded broadcast channels," Peredachi Informatsii. To appear.

4. C. van der Meulen, "General Coding Theorems for the General Discrete Memoryless Broadcast Channel", IEEE Trans. Inform. Theory. To appear.

5. T. M. Cover, An achievable rate region for the broadcast channel. To appear IEEE Trans. Inform. Theory.

6. A Wyner, Recent results in the Shannon theory, IEEE Trans. Inform. Theory, IT-20 (1), 2-10 (Jan. 1974). (See also Ref. 30.)

7. P. P. Bergmans, A simple converse for broadcast channels with additive white Gaussian noise, IEEE Trans. Inform. Theory, IT-20, 279-280 (March, 1974).

8. P. P. Bergmans, and T. M. Cover, Cooperative broadcasting, IEEE Trans. Inform. Theory, IT-20 (3), 317-324 (May, 1974).

9. A. Wyner, and J. Ziv, A theorem on the entropy of certain binary sequences and applications: Part I, IEEE Trans. Inform. Theory, (1973).

10. A. Wyner, A theorem on the entropy of certain binary sequences and applications: Part II, IEEE Trans. Inform. Theory, (1973).

11. R. Ahlswede, and J. Körner, "Source coding with side information and a converse for degraded broadcast channels." In preparation.

12. G. D. Forney, "Information theory." Unpublished course notes, 1972.

13. T. M. Cover, A proof of the data compression theorem of Slepian and Wolf for ergodic sources. IEEE Trans. Inform. Theory. (See Ref. 21.)

14. C. E. Shannon, Two-way communication channels, Proc. 4th Berkeley Symp. Prob. Stat., Vol. 1. Univ. of California Press, Berkeley, 1961, pp. 611-644.

15. C. E. Shannon, A mathematical theory of communication, (BSTJ, 379-423 (1948); A mathematical theory of communication, BSTJ, 623-656, October (1948); Communication in the presence of noise, Proc. IRE, Jan. 1949; all reprinted in Key Papers in the Development of Information Theory, D. Slepian (ed.) pp. 5-46, N.Y., N.Y. IEEE Press, 1974.

16. R. Ahlswede, Multi-way communication channels, Proc. 2nd Inter. Sym. Inform. Transmission. Hungarian Press. Tsahkadsor, Armenia, U.S.S.R., 1971.

17. H. Liao, A coding theorem for multiple access communication, presented at Interna. Symp. Inform. Theory, Asilomar, 1972. Also Ph.D. dissertation, "Multiple access channels," Dept. of Electrical Engineering, University of Hawaii, 1972.

18. M. Ulrey, A coding theorem for a channel with several senders and receivers. Submitted to Information and Control. Also, Ph.D. dissertation, Ohio State University, 1973.

19. R. G. Gallager, "Information Theory and Reliable Communication", Wiley, New York, 1968.

20. D. Slepian and J. Wolf, A coding theorem for multiple access channels with correlated sources, B.S.T.J. (September 1973).

21. D. Slepian and J. K. Wolf, Noiseless coding of correlated information sources, IEEE Trans. Inform. Theory, 19, 471-480 (July, 1973).

22. R. M. Gray and A. D. Wyner, Source coding for simple networks. In preparation.

23. A. D. Wyner, Another look at the coding theorem of information theory, Proc. IEEE, 58, 894-913 (June, 1970).

24. D. J. Sakrison, A geometric treatment of the problem of source encoding a Gaussian random variable, IEEE Trans. Inform. Theory, IT-14, 481-486 (May, 1968).

25. A. D. Wyner, On the transmission of correlated Gaussian data over a noisy channel with finite encoding block length, Information and Control, 20, 193-215 (April, 1972).

26. T. T. Kadota, M. Zakai, and J. Ziv, Mutual information of the white Gaussian channel with and without feedback, IEEE Trans. Inform. Theory, IT-17, 368-371 (July, 1971).

27. T. T. Kadota, M. Zakai, and J. Ziv, Capacity of a continuous memoryless channel with feedback, IEEE Trans. Inform. Theory, IT-17, 372-378, (July, 1971).

28. J. K. Wolf, and J. Ziv, Transmission of noisy information to a noisy receiver with minimum distortion, IEEE Trans. Inform. Theory, IT-16, 406-411 (July, 1970).

29. H. S. Witsenhausen, On sequences of pairs of dependent random variables. To appear.

30. P. Gacs, and J. Körner, Common information is much less than mutual information. To appear in Probs. of Control and Inform. Theory.

31. R. L. Dobrushin, Survey of Soviet research in information theory, IEEE Trans. Inform. Theory, IT-18, 703-724 (November, 1972).