
Invited Paper

EDWARD C. VAN DER MEULEN, MEMBER, IEEE

Abstract—The advances in the area of multi-way channels during the period 1961–1976 are described. Shannon’s two-way channel, the multiple-access channel, the interference channel, the broadcast channel, and the relay channel are treated successively. Only channel coding aspects are discussed. Thirty-nine open problems are mentioned.

I. INTRODUCTION

This survey paper describes results obtained on multiple-user communication channels during the period 1961–1975 and the first half of 1976. The paper starts with Shannon’s two-way channel, and then treats successively the multiple-access, interference, broadcast, and relay channels. An attempt has been made to include all papers in this area which have come to the author’s attention prior to the writing of this article. A relatively large portion of the paper is devoted to the discussion of new results which have not appeared yet in the form of a publication. In particular, recent results on the two-way, multiple-access, interference, and broadcast channels, obtained by various researchers around the world during the past few years, have been incorporated.

Most thoroughly covered is the broadcast channel which has received the greatest attention among information theorists. The abundance of results on the degraded broadcast channel serves as an example of how to attack and analyze other channels for which the theory is not yet quite so developed. We have restricted ourselves to the pure Shannon theory of multiple-user communication channels, and then only to the channel coding aspects of it. The theory of source coding for multiple-user networks is entirely left out, although this is also a very active area at present.

In this survey we have listed thirty-nine unsolved problems, indicated by indented Roman numerals in the text. It is hoped that both old and new workers in the field of information theory can be attracted to work on them, and that new ideas and theories may develop from tackling these problems. Naturally, the emphasis in this article is placed on those aspects of the theory of multiple-user communication channels which seem the most interesting to the author and which are the closest to his own work.

II. TWO-WAY CHANNELS

Shannon [1], in one of his pioneering articles, investigated the two-way channel (TWC), shown in Fig. 1. One of the main features of a TWC is that the transmission in one direction interferes with the transmission in the opposite direction. The channel has two terminals. Each terminal attempts to get across a message to the other terminal. The sources which generate the messages are assumed to be independent. At terminal 1, an input $x_1$ is sent which is received as corresponding output $y_2$ at terminal 2. At the same time, an input $x_2$ is sent at terminal 2 and a corresponding output $y_1$ is received at terminal 1. This implies that the encoding at each terminal must depend on both the message to be transmitted and the sequence of symbols received at that terminal. Similarly the decoding at each terminal depends on the sequences of symbols received and sent at that terminal. Shannon investigated in detail the discrete memoryless (d.m.) TWC and, to some extent, TWC’s with memory. Although he obtained many deep results on the TWC, the precise characterization of the so-called capacity region of this channel remains an open problem. This problem is one of the most difficult ones in multiple-user information theory, and has been open for fifteen years. As the results on the d.m. TWC are far from complete yet, no further models of TWC’s have been investigated. Thus, one could conceive of continuous, compound, nonstationary TWC’s, etc., and could investigate many other properties of the TWC, as has been done for the one-way channel (OWC). However, one of the steps required before proceeding is to determine the capacity region of the d.m. TWC.

A d.m. TWC is defined by a set of transition probabilities $P(y_1,y_2|x_1,x_2)$ where $x_1,x_2,y_1,y_2$ range over finite alphabets $A_1,A_2,B_1,B_2$. In the general setup, a block code pair of length $n$ for a TWC, with $M_1$ messages in the for-
ward direction and $M_2$ in the reverse direction, consists of two sets of $n$ functions which specify how the next input letter at each terminal should be chosen on the basis of the message to be transmitted and the observed outputs at that terminal up to the current time. A decoding system for a block code pair of length $n$ consists of a pair of functions $\phi(m_1,y_{12},\ldots,y_{1n})$ and $\psi(m_2,y_{21},\ldots,y_{2n})$ which take values from 1 to $M_2$ and from 1 to $M_1$, respectively. Here $m_1$ (or $m_2$) denotes the message transmitted at terminal 1 (or 2).

Somewhat easier to treat than the general TWC just described is a restricted version of the TWC where no side information is used at the encoder within each terminal (Fig. 2). This restricted TWC is often referred to as a TWC “without feedback.” However, we avoid this terminology as the term “with feedback” suggests the existence of a link from the receiver at one terminal to the sender at the other terminal, which is not the case in the general TWC of Fig. 1. On the other hand, a TWC with equal outputs $y_1 = y_2$ can be regarded as a TWC with feedback, since here each receiver can observe the output at the other terminal. This is the case in Blackwell’s example to be described below.

Henceforth, we shall refer to the special TWC of Fig. 2 as the restricted TWC.

In the restricted TWC one artificially limits the codes to those where the transmitted sequence at each terminal depends only on the message and not on the received sequence at that terminal. Physically, this can be accomplished if transmission and reception points at each terminal are at different places with cross communication only from transmission to reception but not from reception to transmission. We denote the general d.m. TWC by $K_{22}$ and the restricted d.m. TWC by $K_{22}'$.

A point $(R_1,R_2)$ belongs to the capacity region $C(K_{22})$ if, for every $\epsilon > 0$, there exists a block code and decoding system for $K_{22}$ with $M_1$ and $M_2$ messages in the opposite direction such that both arithmetic average error probabilities are less than $\epsilon$, and $\log M_i \geq n(R_i - \epsilon)$, $i = 1, 2$. Clearly, $C(K_{22}') \subset C(K_{22})$.

Using a random coding argument, Shannon [1] found a simple characterization for $C(K_{22})$. This expression serves as an inner bound for $C(K_{22})$. Shannon also found an outer bound $g_0(K_{22})$ which is relatively simple to evaluate, and showed that $C(K_{22})$ and $g_0(K_{22})$ are both convex.

Let $P_{12} = P(x_1,x_2)$ be an arbitrary probability distribution (PD) over the inputs of $K_{22}$. Based on the joint distribution

$$P(x_1,x_2,y_1,y_2) = P(x_1,x_2)P(y_1|y_2,x_1,x_2),$$

(2)

define $I(X_1;Y_2|X_2)$ and $I(X_2;Y_1|X_1)$ in the usual way (see [2, p. 211]). If $P_{12} = P_1(x_1)P_2(x_2)$ is a product PD, these expressions reduce to $I(X_1;X_2,Y_2)$ and $I(X_2;X_1,Y_1)$ respectively. Let

$$g_0(K_{22}) = \{(I(X_1;Y_2|X_2),I(X_2;Y_1|X_1)) : P_{12} \text{ arbitrary PD}\}$$

(3a)

$$g_1(K_{22}) = \{(I(X_1;Y_2|X_2),I(X_2;Y_1|X_1)) : P_{12} \text{ product PD}\}.$$  

(3b)

Thus, $g_0(K_{22})$ is the set of all rate pairs $(I(X_1;Y_2|X_2), I(X_2;Y_1|X_1))$ obtained as $P_{12}$ varies over all dependent assignments, whereas $g_1(K_{22})$ is the corresponding set when $P_{12}$ is restricted to vary over all independent probability distributions. We denote the convex hull of a set $A$ by $\co(A)$. Shannon [1] proved the following theorems.
Theorem 1: For the general discrete memoryless two-way channel, the convex hull of the region $\mathcal{S}_f(K_{22})$ defined in (3b) is an inner bound on the capacity region, whereas the region $\mathcal{S}_0(K_{22})$ defined in (3a) is an outer bound. Thus

$$\text{co } (\mathcal{S}_f(K_{22})) \subseteq \mathcal{C}(K_{22}) \subseteq \mathcal{S}_0(K_{22}).$$

Theorem 2: For the restricted discrete memoryless two-way channel, the capacity region is precisely equal to the inner bound, that is,

$$\mathcal{C}(K_{22}) = \text{co } (\mathcal{S}_f(K_{22})).$$

For a large class of TWC’s with sufficient symmetric structure, Shannon showed that the inner and outer bound coincide, in which case $\mathcal{C}(K_{22}) = \mathcal{C}(K_{22})$. However, Blackwell (see [1]) provided an example, the binary multiplying channel, for which the inner and outer bound differ. The calculation of the inner bound for the capacity region of the binary multiplying channel amounts to finding the envelope of points $(R_1, R_2)$ such that

$$R_1 = p_1 h(p_1) \quad \text{and} \quad R_2 = p_2 h(p_2)$$

as $0 \leq p_1, p_2 \leq 1$. Here

$$h(p) = -p \log_2 p - (1 - p) \log_2 (1 - p)$$

denotes the usual entropy function. Conversely, the calculation of the outer bound amounts in this case to finding the envelope of points $(R_1, R_2)$ such that

$$R_1 = (1 - p_1) h \left( \frac{p_2}{1 - p_1} \right)$$

and

$$R_2 = (1 - p_2) h \left( \frac{p_1}{1 - p_2} \right)$$

as $0 \leq p_1, p_2 \leq 1$, and $0 \leq p_1 + p_2 \leq 1$. The two bounds are depicted in Fig. 3 and are seen to differ considerably. This difference comes about because for the restricted TWC the inputs $x_1$ and $x_2$ are unrelated at each transmission period, whereas in a code for the general TWC the inputs at a particular time may be related to each other through previous inputs and outputs and hence may be dependent.

Shannon [1] also described $\mathcal{C}(K_{22})$ by a limiting expression. Given d.m. TWC $K_{22}$, one can define a series of derived channels $K_{22}(n)$. The input letters of $K_{22}(n)$ are coding strategies corresponding to $n$ uses of $K_{22}$ with successive input letters being functions of previous inputs and outputs at each terminal. For each channel $K_{22}(n)$, one could in principle calculate the inner bound on its capacity region. This bound is given by $\text{co } (\mathcal{S}_f(K_{22}(n)))$ and equals $\mathcal{C}(K_{22}(n))$, the capacity region of the restricted TWC $K_{22}(n)$. If one multiplies this inner bound by a factor $1/n$, one obtains a region $B_n$ which is comparable in size with the capacity region of the original TWC $K_{22}$. Shannon [1] proved the following theorem.

Theorem 3: Let $B_n$ be the inner bound of the capacity region for the derived channel $K_{22}(n)$, reduced in scale by a factor $1/n$. Then, as $n \to \infty$, the regions $B_n$ approach a limit $B$ which is the capacity region of $K_{22}$. Thus,

$$\mathcal{C}(K_{22}) = \lim_{n \to \infty} \frac{1}{n} \text{co } (\mathcal{S}_f(K_{22}(n))).$$

Unfortunately, this limiting expression is of no practical use as it cannot be evaluated on any computer. However, it is the only theoretical expression presently available for $\mathcal{C}(K_{22})$.

Shannon’s original article [1] remains a source for many open problems. Here we mention only a few.

I. First, what is the precise capacity region of the d.m. TWC $K_{22}$? Can one give a simple characterization of it? To phrase it another way, can one prove a converse showing that Shannon’s inner bound is indeed the capacity region? Alternatively, do there exist encoding techniques based on side information which yield rates exceeding the present inner bound? So far, no examples have been found of achievable rates outside the inner bound $\text{co } (\mathcal{S}_f(K_{22}))$ for any TWC $K_{22}$.

II. Second, can one prove a Feinstein-type of coding theorem for the TWC based on codes with maximal rather than average error?

III. Third, what is the extension of the theory of the d.m. TWC to other models of TWC’s?

The first two problems are considered to be extremely difficult by workers in the field. The presence of side information at the encoder makes the first problem particularly hard. This side information is not present with the multiple-access and broadcast channels to be discussed later. For those multi-way channels, the theory is much further developed. As to the second problem, all attempts have failed up to now [3]. It should be mentioned here that
the capacity region of the TWC depends on the error concept used. This fact was shown by Ahlswede [4] and Dueck [5] and will be further discussed below.

Jelinek [6] investigated whether the inner region \( \mathcal{C}(K_{22}) \) can be exceeded if the past is allowed to influence the selection of input signals. He analyzed the structure of sources for TWC's which may be used to obtain larger inner regions. However, as pointed out earlier, examples of TWC's with achievable rate points exceeding the inner bound have not been found. Jelinek [6] also investigated the decomposition of binary TWC's into an interconnection of pairs of oppositely oriented memoryless binary one-way channels, and gave a thorough classification of such TWC's. Jelinek's canonical decomposition has been effectively applied by Carleial [7] in his study of binary interference channels (see Section IV).

In another paper Jelinek [8] introduced the concept of coding information loss and defined a class of lossless TWC's. See [9] for a survey of Jelinek's results.

The late Russian mathematician Libkind [10] investigated whether \( \mathcal{C}(K_{22}) \) equals \( \mathcal{C}(K_{22}) \), but his proof of equality is generally considered to be incorrect. Wolowitz [11] gave a perspicuous random coding proof of Shannon's result that (1) can be replaced by the weaker conditions

\[
\frac{1}{M_2} \sum_i P^n(A_{ij}|u_{i1},u_{i2}) \geq 1 - \mu_1, \quad i = 1, \ldots, M_1 \tag{7a}
\]

\[
\frac{1}{M_1} \sum_i P^n(B_{ij}|u_{i1},u_{i2}) \geq 1 - \mu_2, \quad j = 1, \ldots, M_2. \tag{7b}
\]

Ahlswede [4] investigated further the restricted TWC and proved the following strong converse: given a code with maximal errors for \( K_{22} \), which signals at a rate pair outside \( \mathcal{C}(K_{22}) \), then, for sufficiently large blocklength, at least one of the maximal error probabilities tends to one.

IV. It would be desirable to obtain a similar result for average error and hence obtain a strong converse of Theorem 2.

Using results from combinatorics, Ahlswede [4] showed that, for TWC's, one can in general not reduce a code with average errors to a code with maximal errors without an essential loss in code length or error probability. This is in contrast with the coding theory for the OWC [12, p. 16].

In [4] Ahlswede pointed out that, for multi-way channels, there is a definite distinction between the theory of coding for average error and the theory of coding for maximal error. Recently, Dueck [5] has substantiated this claim and strengthened Ahlswede's results in the following way. Introducing a so-called "contraction" TWC, Dueck [5] has shown that i) the capacity region of the restricted TWC based on maximal error is smaller than the corresponding capacity region using average error, and ii) the maximal error capacity region of the general TWC \( K_{22} \) is smaller than the average error capacity region of the restricted TWC \( K_{22} \). Dueck [5] has obtained a similar result for the multiple-access channel (see Section III).

V. These results indicate that the determination of the maximal error capacity region of the TWC, both for the general case \( K_{22} \) and the restricted version \( K_{22} \), is another challenging problem.

III. THE MULTIPLE-ACCESS CHANNEL

Shannon wrote [1, p. 64], "In another paper we will discuss the case of a channel with two or more terminals having inputs only and one terminal with an output only, a case for which a complete and simple solution of the capacity region has been found." This channel has later been called the multiple-access channel (MAC). It is not known which solution Shannon had in mind when he wrote this, but it took ten years before other researchers solved this problem. A simple characterization of the capacity region of the d.m. MAC with two (and also three) input users was first presented by Ahlswede at the Tsahkadsor Symposium in 1971 [13]. His paper can be regarded as one of the first major breakthroughs in the field of multiple-user channels since Shannon's paper [1]. At the same symposium, the author [14] put forward a limiting expression for the capacity region of the d.m. MAC and provided simple inner and outer bounds on it. Later, other simple characterizations were obtained by Ahlswede [17], by Liao [15], and by Slepian and Wolf [16]. We shall consider several communication situations.

A. Two Input Users: Separate Encoder Inputs

A d.m. MAC with two input users consists of a set of transition probabilities \( P(y|x_1,x_2) \) where \( x_1,x_2,y \) range over finite alphabets \( A_1,A_2,B \). The channel has two input terminals with no cross communication between them and one output terminal. The inputs are \( x_1 \) and \( x_2 \), and the output is \( y \). Each input terminal attempts to get across a message to the output terminal. The sources which generate the messages are assumed to be independent. Moreover, the messages are connected separately to the encoders, i.e., message \( m_1 \) to input \( x_1 \) and message \( m_2 \) to input \( x_2 \). We call this "separate encoder inputs," as opposed to "correlated encoder inputs" to be described in subsection III-B. We denote a d.m. MAC with two input users, two independent message sources, and separate encoder inputs by \( K_{21} \). The communication configuration for \( K_{21} \) is shown in Fig. 4.

A code \( (n,M_1,M_2,\mu) \) for the d.m. MAC \( K_{21} \) consists of \( M_1 \) codewords \( u_{i1} \in A_1^n \); \( M_2 \) codewords \( u_{i2} \in A_2^n \); and \( M_1M_2 \) decoding sets \( D_{ij} \subseteq B^n \); such that

\[
\frac{1}{M_1M_2} \sum_{ij} \sum_i P^n(D_{ij}|u_{i1},u_{i2}) \geq 1 - \mu \tag{8}
\]

A rate pair \( (R_1,R_2) \) belongs to the capacity region \( \mathcal{C}(K_{21}) \) if, for any \( \epsilon > 0 \) and any \( 0 < \mu < 1 \), for \( n \) sufficiently large, there exists a code \( (n,M_1,M_2,\mu) \) so that \( \log M_i \geq n(R_i - \epsilon), \quad i = 1,2 \).

We remark here that \( \mathcal{C}(K_{21}) \) is the average error capacity region. There exist several equivalent characterizations for \( \mathcal{C}(K_{21}) \) which are presented below. Recently, Dueck [5] has shown that the maximal error capacity re-
VAN DER MEULEN: SURVEY OF MULTI-WAY CHANNELS

Fig. 4. Regular multiple-access channel.

region of $K_{21}$ is smaller than the corresponding average error capacity region.

VI. The determination of the maximal error capacity region of the MAC in various communication situations remains an open problem.

With $P_{12} = P(x_1, x_2)$ an arbitrary PD on the inputs, the joint distribution of inputs and outputs is given by

$$P(x_1, x_2, y) = P(x_1, x_2)P(y|x_1, x_2).$$

(9)

The following identity always holds:

$$i(X_1, X_2; Y) = i(X_1; Y) + i(X_2; Y | X_1).$$

(10)

In this subsection, we shall only be concerned with the case where $P_{12}$ is a product PD, i.e., $P(x_1, x_2) = P_1(x_1)P_2(x_2)$. However, in subsections III-C, III-D, and III-E, we will also allow dependent assignments. We now proceed with the various characterizations of the average error capacity region $\mathcal{E}(K_{21})$.

Characterization 1: Define the regions

$$\mathcal{E}_1 = \{(i(X_1; Y), i(X_2; Y | X_1)) : P_{12} \text{ product PD}\}.$$  

(11a)

$$\mathcal{E}_2 = \{(i(X_1; Y), i(X_2; Y | X_1)) : P_{12} \text{ product PD}\}.$$  

(11b)

Ahlswede [13] proved the following theorem.

Theorem 4: For the discrete memoryless multiple-access channel $K_{21}$, whereby separate independent messages are transmitted over two different input terminals to one common receiver, the capacity region is the convex hull of the union of the two regions $\mathcal{E}_1$ and $\mathcal{E}_2$ described in (11). That is,

$$\mathcal{E}(K_{21}) = co (\mathcal{E}_1 \cup \mathcal{E}_2).$$

The proof of Theorem 4 is based on an application of Shannon’s random coding theorem for the OWC [18] and TWC [1].

In [14], the author provided five different examples of d.m. MAC’s, among which were the binary multiplying MAC and the noiseless binary erasure MAC. For the first channel the inputs and outputs are binary, and the channel operation is defined by $y_1 = y_2 = x_1 x_2$. The capacity region of this MAC is just the set of positive rates for which $R_1 + R_2 \leq 1$. It can be shown [14] that, for a large class of MAC’s with binary inputs and outputs, the capacity region $\mathcal{E}(K_{21})$ reduces to the same triangular region. The noiseless binary erasure MAC will be treated in more detail in subsection III-F below.

Characterization 2: Let $P_{12} = P_1 \times P_2$ be a product PD on inputs. Denote by $\mathcal{G}(P_1 \times P_2)$ the set of all pairs $(R_1, R_2)$ such that

$$0 \leq R_1 \leq i(X_1; Y | X_2),$$

(12a)

$$0 \leq R_2 \leq i(X_2; Y | X_1),$$

(12b)

$$R_1 + R_2 \leq i(X_1, X_2; Y).$$

(12c)


Theorem 5: The capacity region of the discrete memoryless multiple-access channel $K_{21}$ is the convex hull of the union over all independent probability distributions $P(x_1, x_2) = P_1(x_1)P_2(x_2)$ of the sets of rates $(R_1, R_2)$ satisfying (12). That is,

$$\mathcal{E}(K_{21}) = co (\bigcup P_1 \times P_2)$$

where the union is taken over all possible product distributions $P_1 \times P_2$.

The continuous MAC was first discussed by Slepian and Wolf [16] who outlined an approach. The above characterizations carry immediately over to the continuous case if discrete probabilities are replaced by densities and constraints are imposed at the inputs. For a clear presentation of the solution to the capacity region of the Gaussian MAC in the case $K_{21}$ see [19]. The result for the Gaussian MAC was first given at a Communication Theory Workshop in California in 1973, in adjacent presentations by T. Cover [20] and A. Wyner [19].

The results for the discrete case go back to the Symposium in Tsahkadsor in 1971. We shall return to the Gaussian MAC and its capacity region in subsection III-F where it will be discussed in conjunction with other results.

We next turn to a third characterization in terms of a limiting expression.

Characterization 3: Given the d.m. MAC $K_{21}$, define, for each $n \geq 1$ a derived channel $K_{21}(n)$ with transition probabilities given by

$$P^n(x_1, x_2) = \prod_{i=1}^{n} P(y_i | x_1, x_2).$$

(13)

For each product PD, $P^n((x_1, x_2) = P_1^n(x_1)P_2^n(x_2)$, on $\mathcal{A}_1^n \times \mathcal{A}_2^n$, define the mutual informations $I^n(X_1; Y)$ and $I^n(X_2; Y)$ based on (13) in the usual way. Let

$$\mathcal{G}^n(K_{21}) = \{(i^n(X_1; Y), i^n(X_2; Y)) : P_{12}^n = P_1^n \times P_2^n\}.$$  

(14)

The author [14] proved the following theorem.

Theorem 6: Let $D_n$ be the convex hull of the region $\mathcal{G}^n(K_{21})$ defined in (14) but reduced in scale by a factor $1/n$. Then, as $n \to \infty$, the regions $D_n$ approach a limit $D$ which is the capacity region of $K_{21}$. That is,

$$\mathcal{E}(K_{21}) = \lim_{n \to \infty} \frac{1}{n} \mathcal{E}(\mathcal{G}(K_{21})).$$

We discuss a fourth and final characterization in terms of superposition. Notice that time-sharing plays an important role when the capacity region is characterized according to Theorem 4 or 5. Many points are added to the achievable region by taking the convex hull. The original
proofs [19], [20] which established the capacity region of the Gaussian MAC required a time-sharing argument to achieve points on a straight line part of the boundary. (See subsection III-F.) Carleial [7] has recently shown that all points on the boundary of the capacity region of the Gaussian MAC can also be achieved by a superposition random coding scheme. (Superposition originated with the broadcast channel, to be discussed in Sections V, VI, VII, and VIII.) For the discrete case, Carleial's observation leads to the following analog. For every PD of the form \( X | Y \), define\( I(U;Y), I(X_1;Y|U), \) and\( I(X_2;Y|X_1,U) \). Denote by \( g_3(Q_0,P_1,P_2) \) the set of rates \( (R_1,R_2) \) such that
\[
0 \leq R_1 \leq I(U;Y) + I(X_1;Y|X_1,U) (16a)
\]
\[
0 \leq R_2 \leq I(X_2;Y|U). (16b)
\]
It follows from later results on superposition (Theorem 22) that every point in \( g_3(Q_0,P_1,P_2) \) is achievable for \( K_{12} \). Moreover, the union over all triples \( (Q_0,P_0,P_2) \) satisfying (15) of the regions \( g_3(Q_0,P_0,P_2) \) contains \( g_1 \cup g_2 \) as defined in (11). Thus as a fourth characterization, we have the following.

**Theorem 7:** The capacity region of the discrete memoryless MAC \( K_{12} \) is the convex hull of the union over all triples \( (Q_0,P_0,P_2) \) satisfying (15) of the regions \( g_3(Q_0,P_0,P_2) \) defined in (16). That is,
\[
\mathcal{C}(K_{12}) = \text{co} \left( \bigcup_{(Q_0,P_0,P_2)} g_3(Q_0,P_0,P_2) \right). (17)
\]

VII. It is conjectured that it is not necessary to take the convex hull in (17), but that it suffices to take the union. In other words, there is indication that superposition will exhaust the set of rates which were previously found by time-sharing.

Superposition is further taken up in Section VI.

**B. Two Input Users: Correlated Encoder Inputs**

Slepian and Wolf [16] studied a communication situation for the d.m. MAC where the information to be transmitted by the two input users is correlated in a special way. The general configuration is shown in Fig. 5. Here, three sources emit statistically independent messages \( i, j, \) and \( k \) at rates \( R_1, R_2, \) and \( R_0 \). Message pair \( (i,k) \) is encoded at terminal 1 via an encoding function \( f(i,k) = u_{ik} \), and message pair \( (j,k) \) is encoded at terminal 2 via an encoding function \( g(j,k) = v_{jk} \). The receiving terminal must estimate the three source messages \( i, j, \) and \( k \). Thus each input terminal attempts to get across one separate and one common message to the output terminal. This extra feature adds a totally new dimension to the problem and literally so to the rate region (Fig. 6).

We may regard the inputs to the encoders as correlated because the mappings \( f \) and \( g \) are related to each other. However, this form of correlation is only a special case of the more general situation of two dependent source outputs. One could think of the three independent source outputs as being the result of a factorization of two dependent source outputs \( i' = (i,k) \) and \( j' = (j,k) \) into a common part \( k \) and two private messages \( i \) and \( j \). If \( i, j, \) and \( k \) are independent, then \( k \) can indeed be regarded as the common part of \( i' \) and \( j' \). However, in the case when \( i' \) and \( j' \) are dependent, Gács and Körner [92] have shown that the extraction of a common part \( k \) is generally no longer possible.2

We denote a d.m. MAC with two input terminals over which three independent messages are sent in the way just described by \( K_{12} \).

A code \( (n,M_1,M_2,M_0,\mu) \) for \( K_{12} \) consists of \( M_1 M_0 \) codewords \( u_{ik} \in \mathcal{A}_1 \); \( M_0 M_0 \) codewords \( v_{jk} \in \mathcal{A}_2 \); and \( M_1 M_2 M_0 \) pairwise disjoint decoding sets \( D_{ijk} \subset \mathcal{B}_0 \); such that
\[
\frac{1}{M_1 M_2 M_0} \sum_i \sum_j \sum_k P^m(D_{ijk}|u_{ik},v_{jk}) \geq 1 - \mu. (18)
\]

A rate triple \( (R_1,R_2,R_0) \) belongs to the capacity region \( \mathcal{C}(K_{12}) \) if, for any \( \epsilon > 0 \) and any \( 0 < \mu < 1 \), for \( n \) sufficiently large, there exists a code \( (n,M_1,M_2,M_0,\mu) \) such that \( \log M_i \geq n(R_i - \epsilon), i = 1,2,3 \). For the characterization of \( \mathcal{C}(K_{12}) \), Slepian and Wolf introduced an artificial random variable \( (RV) X_0 \) which takes values in an arbitrary finite set \( \mathcal{A}_0 \) of size \( M_0 = |\mathcal{A}_0| \) according to a PD \( P_0(x_0) \). Let \( P(x_1,x_2|x_0) = P_0(x_1|x_0)P_0(x_2|x_0) \) be a PD on \( \mathcal{A}_1 \times \mathcal{A}_2 \)

\[
^2 \text{This result was later sharpened by Witsenhausen [93]. For further contributions, see [95] and [96]. For an exposition of the intricate concept of "common information," see the articles by Ahlswede and Körner [94] and Wyner [21]. The general problem of transmission of correlated messages over multi-way channels is far from being solved.}

Fig. 5. Multiple-access channel with correlated encoder inputs

Fig. 6. Three-dimensional rate region for multiple-access channel with correlated sources. Reprinted from [16].
such that $X_1$ and $X_2$ are conditionally independent given $X_0$. Then the joint distribution $P(x_1,x_2,x_0,y)$ is given by

$$P_0(x_0)P_{01}(x_1|x_0)P_{02}(x_2|x_0)P(y|x_1,x_2).$$  \hfill (19)

Based on (19), let $\mathcal{G}^*(P_0,P_{01},P_{02})$ denote the set of all triples $(R_1,R_2,R_0)$ such that

$$0 \leq R_1 \leq I(X_1;Y|X_2,X_0)$$ \hfill (20a)

$$0 \leq R_2 \leq I(X_2;Y|X_1,X_0)$$ \hfill (20b)

$$R_1 + R_2 \leq I(X_1,X_2;Y|X_0)$$ \hfill (20c)

$$0 \leq R_0 + R_1 + R_2 \leq I(X_1,X_2;Y).$$ \hfill (20d)

This three-dimensional rate region is a polyhedron like the one depicted in Fig. 6. Slepian and Wolf [16] proved the following.

**Theorem 8:** The capacity region of a discrete memoryless multiple-access channel with two input terminals, over which three independent messages $i$, $j$, and $k$ are sent with the aid of two encoding functions $f(i,k)$ and $g(j,k)$, is the convex hull of the union, over all distributions $P_0(x_0)$, $P_{01}(x_1|x_0)$, $P_{02}(x_2|x_0)$ satisfying (19), of the regions $\mathcal{G}^*(P_0,P_{01},P_{02})$ described in (20). That is,

$$\mathcal{C}(K_{21}) = \text{co} \left( \bigcup \mathcal{G}^*(P_0,P_{01},P_{02}) \right)$$

where the union is taken over all possible choices of $|\mathcal{A}_0|,|\mathcal{A}_1|,|\mathcal{A}_2|$.

The region $\mathcal{C}(K_{21})$ lies in the first octant and is bounded by the planes $R_0 = 0$, $R_1 = 0$, $R_2 = 0$ and a convex surface. Theorem 8 implies Theorem 5 if one puts $R_0 = 0$.

In [16] Slepian and Wolf have also provided a reliability function upper bound on the error probability for $K_{21}$. Moreover, these authors formulated several open problems related to the MAC. We summarize these problems here and describe some of the progress which has been made on them since.

**VIII.** Slepian and Wolf conjectured that, in determining $\mathcal{C}(K_{21})$, the size of $|\mathcal{A}_0|$, $|\mathcal{A}_1|$, and $|\mathcal{A}_2|$ can be bounded. Recently, Cover [22] has shown that it suffices to take $|\mathcal{A}_0| \leq \min(|\mathcal{A}_1|,|\mathcal{A}_2|,|\mathcal{B}|)$. The question remains whether $|\mathcal{A}_0|$ can be further bounded, possibly by $\exp_2(R_0)$ as suggested in [16].

**IX.** The explicit construction of good codes for the MAC is an almost unexplored field that, even for noiseless MAC’s, leads to problems not encountered with OWC’s. Recently this problem has been attacked by Kasami and Lin [23] for simple MAC’s.

**X.** What is the zero-error capacity region of the MAC?

**XI.** Gaarder and Wolf [24] have shown that feedback can enlarge $\mathcal{C}(K_{21})$. What is the capacity region of $K_{21}$ with feedback? Recently, Cover and Leung Yan Cheong [25a], [25b] have made progress on this problem and found an achievable rate region which includes the Gaarder-Wolf result as a special case. We shall describe their result in subsection III-E.

**XII.** Another area to be explored is the MAC with memory.

**XIII.** The rate distortion theory for MAC’s is another area.

**C. Two Input Users: Correlated Encoder Outputs**

The author [26] has investigated the d.m. MAC with two input users, two independent message sources, separate encoder inputs, and correlated encoder outputs. Here, a special kind of cooperation is allowed between the encoders. The situation is depicted in Fig. 7. There are again two sources which put out independent messages $i$ and $j$ which are delivered to two separate encoders 1 and 2. The first encoder encodes his message $i$ in the usual way by an encoding function $f(i) = u_i$. However, the output of the second encoder is based on the message delivered to him plus the codeword put out by the first encoder. Thus, the second encoder makes up his own codeword only after learning the entire codeword produced by the first encoder. The resulting codeword for transmitter 2 is $u_{ij} = g(j,u_i)$. The role of the two encoders can of course be interchanged. We denote this MAC by $K_{21}$.

A code $(n,M_1,M_2,\mu)$ for $K_{21}$ consists of $M_1$ codewords $u_i \in \mathcal{A}_1$; $M_1M_2$ codewords $v_{ij} \in \mathcal{A}_2$; and $M_1M_2$ disjoint decoding sets $D_{ij} \subset B^n$ such that

$$\frac{1}{M_1M_2} \sum_{i,j} P^n(D_{ij}|u_i,v_{ij}) \geq 1 - \mu. \hfill (21a)$$

Alternatively, with the role of the two encoders interchanged, such a code may consists of $M_2$ codewords $v_j \in \mathcal{A}_2$ and $M_1M_2$ codewords $u_{ji} \in \mathcal{A}_1$ such that

$$\frac{1}{M_1M_2} \sum_{i,j} P^n(D_{ij}|u_{ji},v_j) \geq 1 - \mu. \hfill (21b)$$

Let $P_{12}$ be an arbitrary dependent PD on $\mathcal{A}_1 \times \mathcal{A}_2$. Denote by $\mathcal{G}(P_{12})$ the set of all pairs $(R_1,R_2)$ such that

$$0 \leq R_1 \leq I(X_1;Y|X_2)$$ \hfill (22a)

$$0 \leq R_2 \leq I(X_2;Y|X_1)$$ \hfill (22b)

$$R_1 + R_2 \leq I(X_1,X_2;Y).$$ \hfill (22c)

Thus, $\mathcal{G}(P_{12})$ differs from $\mathcal{G}(P_1 \times P_2)$ defined in (12) by the fact that dependent PD’s are now allowed. The author [26] proved the following.

**Theorem 9:** For a discrete memoryless multiple-access channel with two input users, two independent message sources, and correlated encoder outputs, the capacity re-
Fig. 8. Multiple-access channel with total cooperation between transmitters.

region is the convex hull of the union over all dependent probability distributions $P_{12} = P(x_1,x_2)$ of the regions $\mathcal{E}(P_{12})$ defined in (22). That is,

$$\mathcal{E}(\tilde{K}_{21}) = \co \left( \bigcup \mathcal{E}(P_{12}) \right)$$

where the union is taken over all arbitrary PD's $P_{12}$.

D. Two Input Users: Total Cooperation.

Consider again a d.m. MAC with two input users and two independent source outputs. Suppose that both source outputs are available to both encoders before the code-words are made up. This configuration is shown in Fig. 8. The transmitters can now fully cooperate and share their knowledge about the source messages $i$ and $j$.

The d.m. MAC $K_{21}$ can be viewed as a d.m. OWC with capacity

$$C'_{21} = \max_{P(x_1,x_2)} I(X_1,X_2;Y),$$

(23)

where the maximum is taken over all dependent assignments $P(x_1,x_2)$. The full cooperation between the encoders enables the transmitters to send jointly at total rate $C'_{21}$. Two alternative procedures can now be followed. Either $R_1$ is kept at zero and the second source output is sent at rate $R_2 = C'_{21}$, or the opposite procedure is adopted. This yields the rate points $(0,C'_{21})$ and $(C'_{21},0)$. Time-sharing, and also superposition, will yield all rate pairs on the line connecting these two points (total cooperation line). Thus we have the following theorem.

**Theorem 10:** For a discrete memoryless multiple-access channel with two input users, when total cooperation is allowed between the two transmitters, the capacity region is the set of all rates $(R_1,R_2)$ such that $R_1 \geq 0$, $R_2 \geq 0$, and $R_1 + R_2 \leq C'_{21}$.

In general, the total cooperation rate region will include the region $\mathcal{E}(\tilde{K}_{21})$ established in subsection III-C. In turn, the latter region contains the region $\mathcal{E}(K_{21})$ found in subsection III-A.

E. Feedback

Consider again a d.m. MAC with two input terminals and two independent source outputs, but now with a noiseless feedback link available from the output to the two encoders, as shown in Fig. 9. Although the source outputs are connected to separate encoders, the feedback enables the encoders to cooperate to some extent. The feedback allows each encoder to let the next input letter at his terminal depend on the previous outputs of the channel as well as on the message to be transmitted at that terminal.

Thus, the $(t)$th input pair to the channel has components $x_{1t} = f(m_1,y_1,\ldots,y_{t-1})$ and $x_{2t} = g(m_2,y_1,\ldots,y_{t-1})$.

Gaarder and Wolf [24] have shown, by example, the surprising fact that noiseless feedback can enlarge the capacity region of a d.m. MAC. As an example, they used a noiseless binary erasure MAC which will be discussed in subsection III-F below. The Gaarder–Wolf feedback scheme uses the fact that when a codeword is transmitted by each encoder, and the corresponding outputs are observed via the feedback link, the encoders know precisely the position where the erasures occurred. The encoders can then cooperate to retransmit those symbols which correspond to received erasures. The Gaarder–Wolf scheme is restricted to the special erasure structure of the channel.

Recently, Cover and Leung-Yan-Cheong [25a, 25b] have found a feedback scheme for enlarging the capacity region of a general d.m. MAC. Their scheme consists of two stages which can be superimposed on each other. During the first stage, the two transmitters send information reliably to each other at the maximum possible rate. Because of the feedback, the communication between the transmitters can be viewed as taking place over a TWC. Let $P(x_1,x_2) = P_{1}(x_1)P_{2}(x_2)$ be a product PD on the inputs of the d.m. MAC $P(y|x_1,x_2)$ using feedback. It follows from Theorem 1 that the two transmitters can send to each other at rates $(R_1,R_2)$ arbitrarily close to $(I(X_1;Y|X_2), I(X_2;Y|X_1))$. This rate of transmission will generally be too high for reliable transmission to the actual receiver. However, recall from Theorem 4 and (10) that the total amount of information received by the receiver during the first stage is about $R_3 = I(X_1,X_2;Y)$. This makes intuitively clear what Cover and Leung-Yan-Cheong have proved precisely, namely that the stage 1 transmissions will enable the receiver to narrow down his set of possible transmitted messages to a considerably smaller set of "typical messages." (The concept of jointly typical sequences was introduced by Forney [27] and further developed by Cover [28].)

Cover and Leung-Yan-Cheong [25a] have shown that the number of pairs $(uii,uzj)$ of codewords which are jointly typical with a particular output sequence of length $n$ is about equal to $\exp nR_3$, whereby

$$R_3 = I(X_1,X_2;Y).$$

With probability one, the set of codewords which are jointly typical with the received output sequence will contain the actual transmitted message pair. Because of the feedback, both transmitters know the set of typical codewords corresponding to the received sequence. Then they can cooperate (as in subsection III-D) to send the
missing information, thereby informing the receiver which message pair on the list of typical sequences was sent. This is done at stage 2, at a rate \( R_0 \) arbitrarily close to \( C_{21}^* \) where \( C_{21}^* \) is defined in (23).

Using time-sharing, Cover and Leung-Yan-Cheong [25a] have shown that it is possible to transmit over the d.m. MAC with feedback at rates \( R_1, R_2 \) such that

\[
R_1 + R_2 = \left( I(X_1;Y|X_2) + I(X_2;Y|X_1) \right) + \frac{C_{21}^*}{C_{21} + I(X_1,X_2;Y)}.
\]

(24)

The above feedback scheme includes the Gaarder–Wolf scheme as a special case.

Recent research done on the same problem at Stanford during the summer of 1976 has revealed that superposition of the first stage on the second in the above scheme leads to a still larger region of achievable rates. More precisely, while missing information is sent to the receiver during the second stage of transmission of a certain pair codewords, the two transmitters can already embark on the first stage of sending to each other a new pair of codewords. This leads to the following description of a region of achievable rates for the d.m. MAC with feedback.

Let \( X_0 \) be an artificial RV taking values in an arbitrary finite set \( \mathcal{A}_0 \) of size \( M_0 = |\mathcal{A}_0| \) according to a PD \( P_{x_0}(x_0) \). Let \( P(x_1,x_2|x_0) \) be a PD on \( \mathcal{A}_1 \times \mathcal{A}_2 \) such that \( X_1 \) and \( X_2 \) are conditionally independent given \( X_0 \). The joint distribution \( P(x_0,x_1,x_2,y) \) is given by

\[
P_{x_0}(x_0)P_{x_1}(x_1|x_0)P_{x_2}(x_2|x_0)P(y|x_1,x_2). \tag{25}
\]

Based on (25), let \( g^F(P_0,P_{01},P_{02}) \) denote the set of all pairs \( (R_1, R_2) \) such that

\[
0 \leq R_1 \leq I(X_1;Y|X_2,X_0) \tag{26a}
\]
\[
0 \leq R_2 \leq I(X_2;Y|X_1,X_0) \tag{26b}
\]
\[
R_1 + R_2 - I(X_1,X_2;Y|X_0) \leq I(X_0;Y). \tag{26c}
\]

Notice that (26c) can also be written as

\[
R_1 + R_2 \leq I(X_1,X_2;Y). \tag{26c'}
\]

After the foregoing discussion and in anticipation of later results on superposition, the following theorem, proved in [25b], should be clear.

**Theorem 11:** For the discrete memoryless multiple-access channel with feedback, having two independent message sources and two input terminals, an achievable rate region is given by the union, over all distributions of type (25), of the regions \( g^F(P_0,P_{01},P_{02}) \) defined in (26). That is, every pair \( (R_1,R_2) \) belonging to

\[
g^F = \cup g^F(P_0,P_{01},P_{02}) \tag{27}
\]

is achievable.

The similarity of (26) with (20) is immediately apparent. Whereas in the Slepian–Wolf setup (subsection III-B)

there is a common information rate \( R_0 \), in the present problem \( R_0 \) represents the amount of missing information resulting from the fact that \( R_1 + R_2 \) may exceed the bound \( I(X_1,X_2;Y|X_0) \) in (20c). In the Slepian–Wolf problem, the three source outputs \( i, j, \) and \( k \) are independent. In the present problem, the two source outputs \( m_1 \) and \( m_2 \), and the missing information sent during the second stage are generally dependent, as they are related through the channel. At the time of the writing of this survey article, the optimality of the region \( g^F \) in (27) has not yet been established.

**F. Examples**

1) **Noiseless Binary Erasure MAC**: The noiseless binary erasure MAC was introduced in [14]. Here \( \mathcal{A}_1 = \mathcal{A}_2 = [0,1] \) and \( \mathcal{B} = \{0,1,2\} \). The transition probabilities are given by

\[
P(0,0) = P(2,1) = P(1,0) = P(1,1) = 1. \]

Output 1 may be called an erasure.

An upper bound for \( \mathcal{E}(K_{21}) \) is, in this example, given by the total cooperation line \( R_1 + R_2 = \log_2 3 = 1.5849 \). The region \( \mathcal{E}(K_{21}) \), characterized according to Theorem 5, was computed in [24]. The mutual informations in (12) are, under variation of product PD’s, simultaneously maximized by \( P_0(0) = P_0(0) = 0.5 \) as \( I(X_1;Y|X_2) = I(X_2;Y|X_1) = 1 \) and \( I(X_1,X_2;Y) = 1.5 \). Therefore, for this example, \( \mathcal{E}(K_{21}) \) is the set of all \( (R_1,R_2) \) such that

\[
0 \leq R_1 \leq 0.76, 0 \leq R_2 \leq 0.76, R_1 + R_2 \leq 1. \]

The points \( A = (0.5, 0.5) \) and \( B = (1, 0.5) \) are on the boundary of \( \mathcal{E}(K_{21}) \) as shown in Fig. 10.

Gaarder and Wolf [24] have shown that when a noiseless feedback link is available from the output to both encoders, one can achieve for this example the rate point \( C = (0.76,0.76) \) which lies outside \( \mathcal{E}(K_{21}) \). This is in contrast with the OWC where feedback cannot increase capacity [29], [30]. The Gaarder–Wolf example was the first one to suggest that the capacity region of the MAC can be larger with feedback than without.

![Fig. 10. Achievable rate points for noiseless binary erasure multiple-access channel.](image-url)
Using the superposition feedback scheme described in subsection III-E, one can further improve on the rate point $C$ found by Gaarder and Wolf and obtain a point very close to the total cooperation line. With the aid of Theorem 11, the author has found that the rate pair $D = (0.7910, 0.7910)$ can be achieved for the present example when feedback is used. Our method is as follows.

The problem of finding boundary points of the region $\mathcal{S}_F$ described in (27) reduces in this example to finding a PD of the type (25) which maximizes $H(Y|X_2, X_0) + H(Y|X_1, X_0)$ subject to the condition

$$H(Y(X_2, X_0) + H(Y|X_1, X_0) \leq H(Y).$$

In our search for achievable rate points, we have restricted ourselves to the case of a binary auxiliary RV $X_0$ and to equal rates $R_1 = R_2$.

Choose $A_0 = [0,1]$, $P_0(0) = P_0(1) = 0.5$, and let the transition probability matrices $[P_{01}(x_1|x_0)]$ and $[P_{02}(x_2|x_0)]$ be the same and equal to

$$
\begin{pmatrix}
0.7624 & 0.2376 \\
0.2376 & 0.7624
\end{pmatrix}.
$$

With this choice of parameters, one obtains

$$H(Y|X_2, X_0) = H(Y|X_1, X_0) = 0.7910$$

so that (28) is satisfied, and the point $D = (0.7910, 0.7910)$ is achievable. Notice how close $D$ is to the total cooperation line. The coordinates of $D$ satisfy $R_1 + R_2 = 1.5820$, whereas on the total cooperation line one has $R_1 + R_2 = 1.5849$. It is conceivable that a more extensive search will yield rate points still closer to, or on, the boundary line.

In [26], the author has shown that one can achieve for the present example the rate points $E = (0.92, 0.66)$ and $F = (0.66, 0.92)$ in the case $K_{21}$ (subsection III-C). These points lie outside $\mathcal{C}(K_{21})$ but on the boundary of $\tilde{\mathcal{C}}(K_{21})$. Indeed, the dependent PD $P(0,0) = P(1,1) = \frac{1}{2}$, $P(0,1) = P(1,0) = \frac{1}{2}$, yields $I(X_1; Y) = I(X_2; Y) = \frac{1}{2}$ and $I(X_1, X_2; Y) = \log_2 3 = 1.5849$ bits.

Kasami and Lin [23] have shown that in situation $K_{21}$, it is possible to construct for the present example a uniquely decodable code pair of length $2$ with rate pair $G = (0.5, 0.5)$ which lies inside $\mathcal{C}(K_{21})$. The results are summarized in Fig. 10.

2) Gaussian Multiple-Access Channel: The additive white Gaussian noise (AWGN) multiple-access channel is a continuous alphabet channel. The previous theory can be easily modified so as to apply to this case. We assume again that the MAC has two input terminals and that two independent sources deliver separate messages $m_1$ and $m_2$ to the two encoders. For this channel, $A_1 = A_2 = B = \mathcal{T}$, and the channel operation is defined by $Y = X_1 + X_2 + Z$. Here $Z$ is a zero-mean Gaussian noise RV, independent of the inputs $X_1$ and $X_2$, with variance $N$, and $Y$ is the output signal. There is an encoding constraint on the inputs which requires that the encoded messages $x_1 = (x_{11}, \ldots, x_{1n})$, $x_2 = (x_{21}, \ldots, x_{2n})$ satisfy

$$\frac{1}{n} \sum_{i=1}^{n} x_{ii}^2 \leq E_i, \quad i = 1, 2$$

for any blocklength $n$.

Cover [20] and Wyner [19] have determined the capacity region of the AWGN-MAC with two input users for the case $K_{21}$. They showed that $\mathcal{C}(K_{21})$ is the set of all pairs $(R_1, R_2)$ such that

$$0 \leq R_1 \leq \frac{1}{2} \log \left(1 + \frac{E_1}{N}\right) \leq \bar{C}_1$$

$$0 \leq R_2 \leq \frac{1}{2} \log \left(1 + \frac{E_2}{N}\right) \leq \bar{C}_2$$

$$R_1 + R_2 \leq \frac{1}{2} \log \left(\frac{1 + E_1 + E_2}{N}\right) \leq \bar{C}_{21}.$$  

Both also gave an encoding method to achieve the corner points $A = (\bar{C}_1, \bar{C}_{21} - \bar{C}_1)$ and $B = (\bar{C}_{21} - \bar{C}_1, \bar{C}_2)$ shown in Fig. 11. Any point on the line connecting $A$ and $B$ is achievable by time-sharing, and also by superposition coding, as was shown by Carleial [7].

The case of full cooperation was worked out by Carleial [7]. A point on the total cooperation line is achieved by using encoding functions

$$X_1 = f_1(m_1, m_2)$$

$$X_2 = X_1(E_2/E_1)^{1/2}$$

with $X_1 \sim \mathcal{N}(0, E_1)$, $t = 1, \ldots, n$. Notice that $X_1$ and $X_2$ are perfectly correlated so that each component $Y_t$ of the output sequence $Y$ has power var $Y_t = E_1 + E_2 + 2(E_1E_2)^{1/2} + N$. It follows that the total cooperation capacity (23) is equal to

$$C_{21}^* = \frac{1}{2} \log \left(1 + \frac{E_1 + E_2 + 2(E_1E_2)^{1/2} + N}{N}\right).$$

In the case of equal input powers $E_1 = E_2 = E$, (32) reduces to

$$C_{21}^* = \frac{1}{2} \log \left(1 + \frac{4E}{N}\right).$$
Now consider feedback. Bergmans and Cover [31] have shown that the achievable rate region $\mathcal{F}$ established in Theorem 11 reduces in the case of the AWGN-MAC to the set of rates $(R_1, R_2, R_3)$ such that

$$0 \leq R_1 \leq \frac{1}{2} \log \left( 1 + \frac{\alpha_1 E_1}{N} \right)$$

$$0 \leq R_2 \leq \frac{1}{2} \log \left( 1 + \frac{\alpha_2 E_2}{N} \right)$$

$$R_1 + R_2 \leq \frac{1}{2} \log \left( 1 + \frac{E_1 + E_2 + 2(\alpha_1 \alpha_2 E_1 E_2)^2}{N} \right)$$

where $0 \leq \alpha_i \leq 1$, $\alpha_i = 1 - \alpha_i$, $i = 1, 2$. They showed that equality in (34) can be achieved throughout by the following scheme. Let $X_0 \sim \mathcal{N}(0,1)$, $X_1 \sim \mathcal{N}(0, \alpha_1 E_1)$, $X_2 \sim \mathcal{N}(0, \alpha_2 E_2)$, and $X_0$, $X_1$, $X_2$ mutually independent. Moreover, define $X_1 = X_1 + X_0 \alpha_1 E_1^{1/2}$, and $X_2 = X_2 + X_0 \alpha_2 E_2^{1/2}$. Then it is easily verified that $I(X_1; Y|X_2, X_0)$, $I(X_2; Y|X_1, X_0)$, and $I(X_1, X_2; Y)$ are equal to the right sides of (34a), (34b), and (34c), respectively.

We now specialize to the symmetric case $E_1 = E_2 = E$ and investigate how to achieve a point on the boundary of $\mathcal{F}$. The problem reduces to finding $R_1 = R_2 = R^*$ so that equality obtains in (34). Cover and Leung-Yan-Cheong [25b] showed that the solution to this problem is given by

$$R^* = \frac{1}{2} \log \left( 2 \left( 1 + \frac{E}{N} \right)^{1/2} - 1 \right).$$

In the symmetric case, one easily verifies that

$$C_{21} < 2R^* < C_{21}^*.$$ 

This shows that for the AWGN-MAC with feedback, one can, at least in the symmetric case, always transmit at rate pairs $(R^*, R^*)$ exceeding the classical capacities region $\mathcal{C}(K_{31})$, but one may not be able to reach the total cooperation line $C_{21}^*$. A similar statement is true in the asymmetric case. Cover and Leung-Yan-Cheong [25b] have calculated $R^*$ for the case $E/N = 10$ and found $R^* = 0.8643$ nats, whereas $C_{21} = 1.5223$ and $C_{21}^* = 1.8568$. The rate pair (24), achieved by the feedback scheme in [25a], yields $R_1 = R_2 = 0.8147$. The results are summarized in Fig. 11.

G. Three Input Users: Separate Encoder Inputs

The foregoing can be extended to the case of a d.m. MAC with three input terminals and a single output. Then the channel is specified by a set of transition probabilities $P(y|x_1, x_2, x_3, y)$ where $x_1, x_2, x_3, y$ range over finite alphabets $\mathcal{A}_{1,2,3}$. We consider first the case where three independent sources deliver three separate messages to three different input terminals. These messages are encoded individually and the resulting codewords are sent simultaneously over the channel. The decoder at the output must reproduce separately the messages from the three sources. We denote this case by $K_{31}$. A code $(n, M_1, M_2, M_3, \mu)$ for $K_{31}$ consists of $M_1$ codewords $u_{1i} \in \mathcal{A}_{1i}$, $M_2$ codewords $u_{2j} \in \mathcal{A}_{2j}$, $M_3$ codewords $u_{3k} \in \mathcal{A}_{3k}$, and $M_1 M_2 M_3$ disjoint decoding sets $D_{ijk} \subset \mathcal{B}_{3i}$ such that

$$\frac{1}{M_1 M_2 M_3} \sum_{i} \sum_{j} \sum_{k} \sum_{\mu} I(U_{ij}, U_{2j}, U_{3k}) \geq 1 - \mu.$$ 

A rate point $(R_1, R_2, R_3)$ belongs to the capacity region $\mathcal{C}(K_{31})$ if, for sufficiently large blocklength, there exist codes with signaling rates arbitrarily close to $(R_1, R_2, R_3)$ and error probability arbitrarily small. Two simple characterizations of $\mathcal{C}(K_{31})$ have been found. Let

$$P_{123} = P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_2)$$

be a product PD on $\mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{A}_3$. The identity

$$I(X_1, X_2, X_3; Y) = I(X_1; Y) + I(X_2; Y|X_1)$$

$$+ I(X_3; Y|X_1, X_2)$$

holds for all six permutations $(i, j, k)$ of the triple $(1, 2, 3)$.

Characterization 4: For each permutation $(i, j, k)$ of the triple $(1, 2, 3)$, define the region

$$\mathcal{C}(i, j, k) = \{(R_1, R_2, R_3) : R_1 \leq R_{ij}, R_2 \leq R_{ik}, R_3 \leq R_{jk}\}.$$ 

Ahlsweede [13] proved the following theorem.

**Theorem 12:** The capacity region of the discrete memoryless multiple-access channel with three input terminals and one output, when three independent sources deliver separate messages to the three encoders, is equal to the convex hull of the union of the six regions $\mathcal{C}(i, j, k)$ described in (40). That is,

$$\mathcal{C}(K_{31}) = \bigcup (i, j, k) \mathcal{C}(i, j, k)$$

where the union is taken over all permutations $(i, j, k)$ of $(1, 2, 3)$.

The proof of Theorem 12 is surprisingly complicated compared to that of Theorem 4, especially the proof of the converse part.

XIV. Ahlsweede [13] proposed finding a simpler proof of the converse part as a problem. In [32], the author provided a simplification of Ahlsweede's arguments. Presented there was a simple proof of the inequalities which bound from above the codelengths and their products in the case of a d.m. MAC with three or more input users. There might still be a much simpler proof of the converse part of Theorem 12.

**Characterization 5:**

For every product PD $P_{123} = P_1 \times P_2 \times P_3$ on $\mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{A}_3$, denote by $\mathcal{C}(P_1 \times P_2 \times P_3)$ the set of triples $(R_1, R_2, R_3)$ such that (i) the inequalities

$$0 \leq R_i \leq I(X_i; Y|X_j, X_k)$$

and

$$R_i + R_j \leq I(X_i, X_j; Y|X_k)$$

are satisfied.
hold for all permutations \((i,j,k)\) of \((1,2,3)\), and ii) 
\[
R_1 + R_2 + R_3 \leq I(X_1,X_2,X_3;Y). 
\]

(Slepian and Wolf [16] and Ulrey [33] independently proved the following.

**Theorem 13:** The capacity region of the discrete memoryless multiple-access channel with three input users and one output, and with three separate independent encoder inputs, is equal to the convex hull of the union over all product PD's of the form \((38)\) of the regions \(\mathcal{E}(P_1 \times P_2 \times P_3)\) defined by \((41)\). That is,
\[
\mathcal{E}(K_{31}) = \{\mathcal{E}(P_1 \times P_2 \times P_3)\}
\]
where the union is taken over all product PD's \((38)\).

**H. Three Input Users: Correlated Encoder Inputs**

Slepian and Wolf [16] treated also the d.m. MAC with three input users and a single output when the information supplied to the input users is correlated in a special way. Suppose there are seven independent sources \(S_1, S_2, S_3, S_{12}, S_{13}, S_{23}, S_{123}\) producing information at rates \(R_1, R_2, R_3, R_{12}, R_{13}, R_{23}, R_{123}\), respectively. There are only three encoders. Encoder 1 sees the outputs of only \(S_1, S_{12}, S_{13}\), \(S_{23}\), and similarly for the other two encoders. Now the decoder at the channel output must reproduce the messages from these seven sources. We denote this case by \(K_{31}\).

Slepian and Wolf [16] showed that the seven-dimensional capacity region \(\mathcal{E}(K_{31})\) is described by fifteen inequalities. As their solution is too complicated to describe here, we refer to their article [16] for a full description.

**I. More Than Three Input Users**

The generalization of the foregoing to \(s > 3\) input users is simple in concept, but requires a tremendous formal setup to describe. The capacity region \(\mathcal{E}(K_s)\) for the case of \(s\) input users and separate encoder inputs has been characterized by Ulrey [33]. The region \(\mathcal{E}(K_{s1})\) for the case of \(s\) input users and correlated encoder inputs has been conjectured in [16].

**XV.** This conjecture still needs to be verified.

**XVI.** It would also be interesting to know how to handle correlations more general than the special form between the encoded messages assumed in [16].

**J. Convolutional Codes**

Recently, Ohkubo [34] has extended the Viterbi decoding algorithm [35] to the case of a d.m. MAC. He showed that, for rate pairs \((R_1,R_2)\) in the capacity region \(\mathcal{E}(K_{21})\), the probability of error in decoding a convolutional tree code transmitted over a d.m. MAC is bounded from above as a function of the constraint length of the code.

**IV. INTERFERENCE CHANNELS**

**A. Two Input Users; Two Output Users**

On page 636 of [1], Shannon considered a TWC for which "the transmission and reception points at each end were at different places with no direct cross communication." This channel has later been termed an interference channel (IFC), as the transmission of information from one sender to its corresponding receiver may interfere with the transmission of information from the other sender to its receiver. An IFC with two input and two output terminals is depicted in Fig. 12. This IFC differs from the TWC in two respects. First, the senders at each terminal do not observe the outputs at that terminal. This reduces the TWC to a restricted TWC (Fig. 2). Second, the receivers do not observe directly any inputs, i.e., there is no side information at the receivers. This last feature distinguishes the IFC from a restricted TWC.

A d.m. IFC with two input and two output users consists of a set of transition probabilities \(P(y_1,y_2|x_1,x_2)\) where \(x_1,x_2,y_1,y_2\) range over finite alphabets \(\mathcal{A}_1,\mathcal{A}_2,\mathcal{B}_1,\mathcal{B}_2\). Here \(x_1(x_2)\) is an input at terminal 1(2), whereas \(y_1(y_2)\) is an output at terminal 3(4). There is no cross communication between any of the four terminals. In particular, this implies that the receivers cannot collaborate in their decoding effort, and thus that IFC's with the same marginal conditional PD's are equivalent as far as their communication performance is concerned. Therefore, a d.m. IFC with two inputs and two outputs is completely characterized by its marginal conditional probabilities \(P(y_1|x_1,x_2)\) and \(P(y_2|x_1,x_2)\).

We consider here only the case where two sources deliver two independent separate messages to the two encoders.

**XVII.** However, one could possibly also conceive of IFC's where the encoded messages are correlated in a special way, as in [16], but this situation has not been studied yet.

In this survey we mainly restrict ourselves to the case of an IFC with two input and two output terminals. However, many of the results which follow carry over directly to any IFC with \(s \geq 3\) senders and \(r \geq 2\) receivers. For a brief discussion of the general case, see subsection IV-G below.

**B. Single Interference Channel: Preliminaries**

We first consider the situation where one transmitter \((x_1)\) wishes to communicate exclusively with the first receiver \((y_1)\), while the other transmitter \((x_2)\) wishes to communicate exclusively with the second receiver \((y_2)\). The channel can be thought of as being made up of two principal links (between corresponding terminals) and two interference links (between noncorresponding sender-receiver pairs). We call this case the single interference channel, as opposed to the compound IFC discussed in subsection IV-C.

A single d.m. IFC with two inputs and two outputs is denoted by \(K_{22}\). A code \((n,M_1,M_2,\mu_1,\mu_2)\) for \(K_{22}\) consists of \(M_1\) codewords \(u_{1i} \in \mathcal{A}_1^n\); \(M_2\) codewords \(u_{2j} \in \mathcal{A}_2^n\); \(M_1\) decoding sets \(D_{ji} \subset \mathcal{B}_1^n\); \(M_2\) decoding sets \(D_{2j} \subset \mathcal{B}_2^n\), such that
\[
\frac{1}{M_1M_2} \sum_i \sum_j P^n(D_{ji}|u_{1i},u_{2j}) \geq 1 - \mu_1 
\]
\[
\frac{1}{M_1M_2} \sum_i \sum_j P^n(D_{2j}|u_{1i},u_{2j}) > 1 - \mu_2 
\]
A rate point \((R_1,R_2)\) belongs to the capacity region \(\mathcal{C}(\mathcal{K}_{22})\) if, for any \(\epsilon > 0\) and any \(0 < \mu_1, \mu_2 < 1\), for \(n\) sufficiently large, there exists a code \((n,M_1,M_2,\mu_1,\mu_2)\) such that \(\log M_i \geq n(R_i - \epsilon), i = 1,2\).

XVIII. The determination of the capacity region of the single d.m. IFC \(\mathcal{K}_{22}\) is still an open problem.

Using techniques similar to the ones employed by Shannon [1] for the TWC, one can obtain simple inner and outer bounds on \(\mathcal{C}(\mathcal{K}_{22})\) and also establish a limiting expression for it. For IFC’s with a special degenerate structure, the determination of the capacity region is straightforward. We shall dispense with this latter case first.

Notice that the IFC \(\mathcal{K}_{22}\) can be viewed as a combination of two MAC’s with transition probabilities \(P(y_1|x_1,x_2)\) and \(P(y_2|x_1,x_2)\), respectively. We say that the IFC has statistically equivalent output signals if the two comprising MAC’s are the same, i.e., \(P(y_1|x_1,x_2) = P(y_2|x_1,x_2)\) for all \(x_1,x_2,y_1,\) and \(y_2\). (This terminology is due to Carleial [7]; Sato [36] refers to an IFC of this kind as a twin two-user channel.) The capacity region of such an IFC equals the capacity region of the common component MAC, and thus can be derived from Theorem 5. Ahlswede [17], Carleial [7], and the author [14] independently proved the following.

**Theorem 14:** The capacity region of a discrete memoryless single interference channel with two input and two output users, and with statistically equivalent output signals, is equal to the convex hull of the union over all product PD’s \(P(x_1,x_2) = P_1(x_1)P_2(x_2)\) of the sets of rates \((R_1,R_2)\) such that

\[
0 \leq R_1 \leq I(X_1;Y_1|X_2) \tag{43a}
\]
\[
0 \leq R_2 \leq I(X_2;Y_2|X_1) \tag{43b}
\]
\[
R_1 + R_2 \leq I(X_1,X_2;Y_1,Y_2) \tag{43c}
\]

We next relate Shannon’s outer bound for the TWC [1] to a corresponding outer bound for the d.m. IFC \(\mathcal{K}_{22}\), and obtain an analogous inner bound. Since a code for the IFC \(\mathcal{K}_{22}\) is a code for the restricted TWC \(\mathcal{K}_{22}\), it follows that \(\mathcal{C}(\mathcal{K}_{22})\), the capacity region of the restricted TWC, is an outer bound on \(\mathcal{C}(\mathcal{K}_{22})\), the capacity region of the IFC. Let \(P_{12} = P_1(x_1)P_2(x_2)\) be a product PD on inputs. Define

\[
\mathcal{G}_1(\mathcal{K}_{22}) = \{(I(X_1;Y_1), I(X_2;Y_2)) : P_{12} \text{ product PD}\} \tag{44a}
\]
\[
\mathcal{G}_0(\mathcal{K}_{22}) = \{(I(X_1;Y_1|X_2), I(X_2;Y_2|X_1)) : P_{12} \text{ product PD}\}. \tag{44b}
\]

Notice that \(\mathcal{G}_0(\mathcal{K}_{22})\) equals the region \(\mathcal{G}_1(\mathcal{K}_{22})\) defined in (43b), provided \(Y_1\) and \(Y_2\) are interchanged. Ahlswede [17], Carleial [7], and the author [14] proved the following.

**Theorem 15:** For the general discrete memoryless single interference channel with two inputs and two outputs, an inner bound \(\mathcal{G}_1(\mathcal{K}_{22})\) on the capacity region is given by the convex hull of the region \(\mathcal{G}_1(\mathcal{K}_{22})\) defined in (44a), whereas the convex hull of the region \(\mathcal{G}_0(\mathcal{K}_{22})\) defined in (44b) is an outer bound. That is,

\[
\mathcal{R}_1(\mathcal{K}_{22}) = \text{co} (\mathcal{G}_1(\mathcal{K}_{22})) \subset \mathcal{C}(\mathcal{K}_{22}) \subset \text{co} (\mathcal{G}_0(\mathcal{K}_{22})). \tag{45}
\]

Just as in the case of the TWC, one may characterize the capacity region of the d.m. IFC by a limiting expression. Let \(\mathcal{K}_{22}(n)\) be the product channel of length \(n\) derived from the d.m. IFC \(\mathcal{K}_{22}\). Ahlswede [13] proved the following theorem.

**Theorem 16:** Let \(E_n\) be the inner bound of the capacity region for the derived discrete memoryless single interference channel \(\mathcal{K}_{22}(n)\), reduced in scale by a factor \(1/n\). Then, as \(n \to \infty\), the regions \(E_n\) approach a limit \(E\) which is the capacity region of the original interference channel \(\mathcal{K}_{22}\). That is,

\[
\mathcal{C}(\mathcal{K}_{22}) = \lim_{n \to \infty} \frac{1}{n} \text{co} (\mathcal{G}_1(\mathcal{K}_{22}(n))).
\]

Theorems 15 and 16 were anticipated by Shannon [1]. The question arises whether or not the inner bound \(\text{co} (\mathcal{G}_1(\mathcal{K}_{22}))\) equals the capacity region \(\mathcal{C}(\mathcal{K}_{22})\) in (45). Ahlswede [17] showed, by example, that generally \(\text{co} (\mathcal{G}_1(\mathcal{K}_{22}))\) is smaller than \(\mathcal{C}(\mathcal{K}_{22})\). As an example, he considered an
IFC with both component MAC's equal to the noiseless binary erasure MAC discussed in subsection III-F. By Theorem 14, the capacity region $\mathcal{C}(\mathcal{K}_{22})$ for this IFC equals the capacity region $\mathcal{C}(\mathcal{K}_{21})$ of the component MAC. Ahlswede [17] showed that the inner bound $\mathcal{C}_{22}$ computed for this example is strictly smaller than $\mathcal{C}(\mathcal{K}_{22})$, which completes the argument.

Sato [36] has recently shown that the outer bound $\mathcal{G}_{22}$ as the union over all product PD's $P(x_1,x_2) = P_1(x_1)P_2(x_2)$ of the sets of rates $(R_1,R_2)$ which satisfy (43). Quite clearly

$$\text{co} \{ \mathcal{G}_{22} \} \subset \{ \mathcal{G}_{22} \}.$$  

Sato [36] showed

$$\mathcal{C}(\mathcal{K}_{22}) \subset \{ \mathcal{G}_{22} \}.$$  

Moreover, this outer bound can be further strengthened by taking the intersection of all such regions over all IFC's with the same marginals. For an IFC with statistically equivalent outputs, this outer bound coincides with the capacity region, as follows directly from Theorem 14. Carleial [7] and Sato [36] have also considered degraded IFC's and obtained simple inner and outer bounds on the capacity region for this class of channels. For more advanced results on the d.m. single IFC, in particular for the description of other achievable rate regions, we turn now to the study of the compound IFC.

C. Compound Interference Channel

Consider the situation where each transmitter attempts to communicate simultaneously with both receivers by sending the same message to the two output users as in the case of the compound OWC [37]. This situation is equivalent to the problem of sending two messages simultaneously over the two component MAC's $P(y_1|x_1,x_2)$ and $P(y_2|x_1,x_2)$. We refer to this situation as the compound interference channel. A compound d.m. IFC with two input and two output users and separate encoded messages is denoted by $\mathcal{K}_{22}$. The mathematical description of $\mathcal{K}_{22}$ in terms of transition probabilities and terminals is the same as for $\mathcal{K}_{21}$, only the communication problem has changed. Ahlswede [17] obtained a simple characterization of the capacity region of the compound d.m. IFC $\mathcal{K}_{22}$. A code $(n,M_1,M_2,\nu_1,\mu_2)$ for $\mathcal{K}_{22}$ consists of $M_1$ codewords $u_1 \in \mathcal{A}_1^n; M_2$ codewords $u_2 \in \mathcal{A}_2^n; M_1M_2$ decoding sets $B_{ij} \subset \mathcal{B}_1^n; M_1M_2$ decoding sets $D_{ij} \subset \mathcal{B}_2^n$ such that

$$\frac{1}{M_1M_2} \sum_i \sum_j P^n(B_{ij}|u_i,u_2) \geq 1 - \mu_1$$  

$$\frac{1}{M_1M_2} \sum_i \sum_j P^n(D_{ij}|u_1,u_2) \geq 1 - \mu_2.$$  

The capacity region of $\mathcal{K}_{22}$, denoted by $\mathcal{C}(\mathcal{K}_{22})$, is defined as usual and can be characterized as follows. Let $P_1 \times P_2$ be a product PD on $\mathcal{A}_1 \times \mathcal{A}_2$, and evaluate the six-tuples

$$I(P_1 \times P_2) = (I(X_1;Y_1|X_2),I(X_1;Y_2|X_2),I(X_2;Y_1|X_1),I(X_2;Y_2|X_1),I(X_1;X_2;Y_1),I(X_1;X_2;Y_2)).$$  

Do this for every product PD $P_1 \times P_2$, and take the convex hull of the set of all six-tuples thus obtained. We denote this region by $\mathcal{F}_{22}$. A typical six-tuple in $\mathcal{F}_{22}$ is denoted by

$$J = (J(1;1|2),J(1;2|2),J(2;1|1),J(2;2|1),J(1,2;1),J(1,2;2)).$$  

For each $J \in \mathcal{F}_{22}$, let $\mathcal{G}_{22}(J)$ be the set of all pairs $(R_1,R_2)$ such that

$$0 \leq R_1 \leq \min (J(1;1|2),J(1;2|2))$$  

$$0 \leq R_2 \leq \min (J(2;1|1),J(2;2|1))$$  

$$R_1 + R_2 \leq \min (J(1,2;1),J(1,2;2)).$$  

Ahlswede [17] proved the following theorem.

**Theorem 17:** The capacity region of the compound discrete memoryless interference channel with two input and two output users is equal to the union over all vectors $J$ defined in (50) of the regions $\mathcal{G}_{22}(J)$ given by (51). That is,

$$\mathcal{C}(\mathcal{K}_{22}) = \bigcup_{J \in \mathcal{F}_{22}} \mathcal{G}_{22}(J).$$

Cover [28], in the context of the broadcast channel, observed that in the expression for $\mathcal{C}(\mathcal{K}_{22})$ in Theorem 17, it suffices to take the union over those points $J \in \mathcal{F}_{22}$ which are convex mixtures of at most six six-tuples $I(P_1 \times P_2)$, each one of the type (49).

We conclude that for the compound d.m. IFC $\mathcal{K}_{22}$, there exists a concise and neat expression for the capacity region. Such an expression has not been found yet for the single d.m. IFC $\mathcal{K}_{21}$. However, Theorem 17 enables us to derive another achievable rate region for the single IFC in addition to the inner bound already established in Theorem 15. This region, which will be described next, can be further improved by superposition.

D. Single Interference Channel (Continued)

We now reconsider the single d.m. IFC $\mathcal{K}_{22}$ discussed in subsection IV-B. Clearly, every rate pair achievable for the compound IFC $\mathcal{K}_{22}$ is a fortiori achievable for the single IFC $\mathcal{K}_{22}$. We may rewrite the inequalities in (51) in the following way. Let $P_1 \times P_2$ be a product PD on $\mathcal{A}_1 \times \mathcal{A}_2$. Denote by $\mathcal{F}_i(P_1 \times P_2)$, $i = 1,2,3,4$, the four sets of pairs $(R_1,R_2)$ which are specified, respectively, by the following four inequalities:

$$\mathcal{F}_1(P_1 \times P_2): 0 \leq R_1 \leq I(X_1;Y_1), 0 \leq R_2 \leq I(X_2;Y_1|X_1),$$

$$\mathcal{F}_2(P_1 \times P_2): 0 \leq R_1 \leq I(X_1;Y_1|X_2), 0 \leq R_2 \leq I(X_2;Y_1),$$

$$\mathcal{F}_3(P_1 \times P_2): 0 \leq R_1 \leq I(X_1;Y_2), 0 \leq R_2 \leq I(X_2;Y_2|X_1),$$

$$\mathcal{F}_4(P_1 \times P_2): 0 \leq R_1 \leq I(X_1;Y_2|X_3), 0 \leq R_2 \leq I(X_2;Y_2).$$
Observe that, by Theorem 4, every rate pair \((R_1, R_2)\) which satisfies (52a) or (52b) can be attained for sending information over the d.m. MAC \(P(y_1|x_1, x_2)\). Similarly, every rate pair \((R_1, R_2)\) satisfying (52c) or (52d) can be attained for sending information over the MAC \(P(y_2|x_1, x_2)\). Therefore, every rate pair \((R_1, R_2)\) which satisfies either (52a) or (52b) and at the same time either (52c) or (52d), is achievable for \(K_{22}\) and a fortiori achievable for \(R_{zz}\). Consequently, define

\[
\mathcal{F}_{12}(P_1 \times P_2) = \mathcal{F}_1(P_1 \times P_2) \cup \mathcal{F}_2(P_1 \times P_2) \quad (53a)
\]

\[
\mathcal{F}_{34}(P_1 \times P_2) = \mathcal{F}_3(P_1 \times P_2) \cup \mathcal{F}_4(P_1 \times P_2) \quad (53b)
\]

and

\[
\mathcal{F}(P_1 \times P_2) = \mathcal{F}_{12}(P_1 \times P_2) \cap \mathcal{F}_{34}(P_1 \times P_2). \quad (54)
\]

Then, by Theorem 17 every \((R_1, R_2) \in \mathcal{F}(P_1 \times P_2)\) is achievable for \(K_{22}\) and hence for \(K_{zz}\). Thus, we obtain the following.

**Theorem 18:** For the general discrete memoryless single interference channel with two inputs and two outputs, an inner bound \(\mathcal{R}_{zz}(K_{22})\) on the capacity region is given by the convex hull of the union over all product PD's \(P(x_1, x_2) = P_1(x_1)P_2(x_2)\) of the regions \(\mathcal{F}(P_1 \times P_2)\) described in (54).

That is,

\[
\mathcal{R}_{zz}(K_{22}) \triangleq \text{co} \left( \bigcup_{P_1 \times P_2} \mathcal{F}(P_1 \times P_2) \right) \subset \mathcal{C}(K_{22}). \quad (55)
\]

Although Theorem 18 is nowhere explicitly stated in the literature, it is of course a direct consequence of Theorem 17. The region \(\mathcal{R}_{zz}(K_{22})\) described in Theorem 18 is closely related to the achievable rate region for the broadcast channel put forward in [28] and [38]. Theorem 18 occurs implicitly in the work of Bergmans [40] and Carleial [39] in their exploration of the Gaussian IFC. More recently, Carleial [7] has improved Theorem 18 by introducing superposition codes. We shall describe his result next.

**E. Single Interference Channel: Superposition Coding**

Consider again a single d.m. IFC \(K_{22}\) with transition probabilities \(P(y_1, y_2|x_1, x_2)\). Let \(Z_1\) and \(Z_2\) be two auxiliary RV's, taking values in arbitrary finite sets \(D_1\) and \(D_2\) of size \(M_1 = |D_1|\) and \(M_2 = |D_2|\), respectively. \(Z_1\) and \(Z_2\) are assumed independent with arbitrary PD's \(Q_1(z_1)\) and \(Q_2(z_2)\), respectively. Let \(P(x_1, x_2|z_1, z_2)\) be a PD on \(\mathcal{A}_1 \times \mathcal{A}_2\) so that

\[
P(x_1, x_2|z_1, z_2) = P_1(x_1|z_1)P_2(x_2|z_2). \quad (56)
\]

Then the joint distribution of \(Z_1, Z_2, X_1, X_2\) is given by

\[
P(z_1, z_2, x_1, x_2) = Q_1(z_1)Q_2(z_2)P_1(x_1|z_1)P_2(x_2|z_2), \quad (57)
\]

whereas the joint distribution \(P(z_1, z_2, x_1, x_2, y_1, y_2)\) equals

\[
Q_1(z_1)Q_2(z_2)P_1(x_1|z_1)P_2(x_2|z_2)P(y_1, y_2|x_1, x_2). \quad (58)
\]

Based on (58), we define four rate regions similar to the ones described by (52). For each choice of \(P = (M_1, M_2, Q_1, Q_2, P_1, P_2)\), denote by \(T_i(P), i = 1, 2, 3, 4\), four sets of pairs \((R_{10}, R_{20})\) which are specified, respectively, by the following four inequalities:

\[
T_1(P) : 0 \leq R_{10} \leq I(Z_1; Y_1), \quad 0 \leq R_{20} \leq I(Z_2; Y_1), \quad (59a)
\]

\[
T_2(P) : 0 \leq R_{10} \leq I(Z_1; Y_2), \quad 0 \leq R_{20} \leq I(Z_2; Y_2), \quad (59b)
\]

\[
T_3(P) : 0 \leq R_{10} \leq I(Z_1; Y_2), \quad 0 \leq R_{20} \leq I(Z_2; Y_2), \quad (59c)
\]

\[
T_4(P) : 0 \leq R_{10} \leq I(Z_1; Y_2), \quad 0 \leq R_{20} \leq I(Z_2; Y_2). \quad (59d)
\]

In addition, define the regions

\[
T_{12}(P) = T_1(P) \cup T_2(P), \quad (60a)
\]

\[
T_{34}(P) = T_3(P) \cup T_4(P), \quad (60b)
\]

and

\[
T(P) = T_{12}(P) \cap T_{34}(P). \quad (61)
\]

According to Theorem 18, every pair \((R_{10}, R_{20}) \in T(P)\) is achievable for \(K_{22}\) for every choice of \(P\).

Carleial [7] has shown that by superposition one can add to every rate pair \((R_{10}, R_{20}) \in T(P)\) another rate pair \((R_{11}, R_{22})\) so that, in total, one can transmit at rates \(R_1 = R_{10} + R_{11}\) and \(R_2 = R_{20} + R_{22}\) over the single IFC \(K_{22}\).

For every choice of positive integers \(M_1 \geq 1, M_2 \geq 1,\) and PD's \(Q_1, Q_2, P_1, P_2\) satisfying (57), denote by \(\mathcal{S}_{22}(P)\) the set of rates \((R_1, R_2)\) such that

\[
R_1 = R_{10} + R_{11} \quad \text{and} \quad R_2 = R_{20} + R_{22}, \quad (62a)
\]

\[
(R_{10}, R_{20}) \in T(P) \quad (62b)
\]

\[
0 \leq R_{11} \leq I(X_1; Y_1 | Z_1, Z_2), \quad (62c)
\]

\[
0 \leq R_{22} \leq I(X_2; Y_2 | Z_1, Z_2). \quad (62d)
\]

whereby \(T(P)\) is defined as in (61) and \(P = (M_1, M_2, Q_1, Q_2, P_1, P_2)\). Carleial [7] proved the following theorem.

**Theorem 19:** An achievable rate region \(\mathcal{R}_C(K_{22})\) for the single discrete memoryless interference channel \(K_{22}\) is given by the convex hull of the union over all vectors \(P\) of the regions \(\mathcal{S}_{22}(P)\) defined in (62). That is,

\[
\mathcal{R}_C(K_{22}) \triangleq \text{co} \left( \bigcup_{P \in \mathcal{S}_{22}(P)} \mathcal{S}_{22}(P) \right) \subset \mathcal{C}(K_{22}). \quad (63)
\]

Carleial's proof is based on a random coding superposition scheme. The RV's \(Z_1\) and \(Z_2\) represent the common part of the input information that is to be decoded by both receivers. The RV's \(X_1\) and \(X_2\) represent the private information that each sender sends to its corresponding receiver. \(X_1\) and \(X_2\) are superimposed on \(Z_1\) and \(Z_2\).

It is noteworthy to observe that the rate region \(\mathcal{R}_C(K_{22})\) defined in (63) includes both Shannon's inner bound \(R_S(K_{22})\) defined in (45) and Ahlswede's rate region \(\mathcal{R}_A(K_{22})\) defined in (55) as special cases. This can be seen by taking either the \(Z\)-variables or the \(X\)-variables to be degenerate in (62).
As examples, Carleial [7] studied the binary IFC and the Gaussian IFC. The analysis of the binary IFC turns out to be surprisingly difficult in the general case. In this survey, we shall discuss only the Gaussian IFC.

F. The Gaussian Interference Channel

The previous theory can be modified in an obvious way so as to apply to IFC's with continuous input and output alphabets and constraints at the inputs. Bergmans [40] and Carleial [7], [39] have investigated the Gaussian IFC. For this channel, all four alphabets are the real line. The inputs $X_1$ and $X_2$ are two independent transmitted signals with power constraints $E_1$ and $E_2$, whereas the output signals are given by

$$Y_1 = \sqrt{a_{11}} X_1 + \sqrt{a_{21}} X_2 + Z_1$$

$$Y_2 = \sqrt{a_{12}} X_1 + \sqrt{a_{22}} X_2 + Z_2.$$  

Here $Z_1$ and $Z_2$ are zero-mean Gaussian noise RV's, independent of $X_1$ and $X_2$, with variances $N_1$ and $N_2$, respectively. The coefficients $a_{ij}$ are known positive constants, also called the power-transmission coefficients.

Absence of interference is represented by $Y_1 = X_1 + Z_1$ and $Y_2 = X_2 + Z_2$. In this case, one can transmit simultaneously in both directions at rates

$$0 \leq R_1 \leq C_1 \triangleq \frac{1}{2} \log \left(1 + \frac{E_1}{N_1}\right)$$

$$0 \leq R_2 \leq C_2 \triangleq \frac{1}{2} \log \left(1 + \frac{E_2}{N_2}\right).$$

At this point, Carleial [39] imposed the strong interference condition

$$a_{12} N_1 E_2 + N_2 = a_{21} E_1 + N_1.$$  

This condition implies $C_1 \leq C_{12}$ and $C_2 \leq C_{21}$, so that (66) reduces to (66). This shows that in the case of strong interference, the capacity region of the Gaussian IFC is the same as in the absence of interference. This phenomenon can also be explained by observing that under (69), $I(X_1; Y_2)$ and $I(X_2; Y_1)$ are so large that the interfering signals can be subtracted from the output signals, after which the channel can be viewed as if no interference existed at all.

In this case, the region $\mathcal{F}_2(P_1 \times P_2) \cap \mathcal{F}_2(P_1 \times P_2)$ becomes the set of rates $(R_1, R_2)$ such that

$$0 \leq R_1 \leq \min (C_1, C_{12})$$

$$0 \leq R_2 \leq \min (C_2, C_{21}).$$

In addition to the quantities $C_1$, $C_2$, $C_{12}$, and $C_{21}$ defined in (66) and (67), define

$$D_{11} \triangleq \frac{1}{2} \log \left(1 + \frac{E_1}{a_{21} E_2 + N_1}\right)$$

$$D_{22} \triangleq \frac{1}{2} \log \left(1 + \frac{E_2}{a_{12} E_1 + N_2}\right).$$

Clearly, $D_{11} = I(X_1; Y_1)$ and $D_{22} = I(X_2; Y_2)$. Next, define the following rate pairs

$$R^{(11)} = (D_{11}, D_{22})$$

$$R^{(10)} = (C_1, \min (C_{21}, D_{22}))$$

$$R^{(01)} = (\min (C_{12}, D_{11}), C_2)$$

$$R^{(00)} = (\min (C_1, C_{12}), \min (C_2, C_{21})).$$
Fig. 14. Achievable rate region for Gaussian interference channel with $E_1 = E_2 = 6$ and $\alpha_{12} = \alpha_{21} = \frac{1}{2}$ where $C_1 = C_2 = 0.9730$, $D_{11} = D_{22} = 0.5493$, $C_{12} = C_{21} = 0.1257$ nats/transmission. Reprinted from [7].

Bergmans [40] and Carleial [7] showed that the four rate pairs (71) are in the capacity region of AWGN-IFC. The achievability of $R^{(11)}$ follows from Theorem 15, the achievability of $R^{(10)}$ from (52b), the achievability of $R^{(01)}$ from (52c), and the achievability of $R^{(00)}$ from (52b) in conjunction with (52c).

Bergmans [40] has determined which of the four points in (71) are not dominated for varying degrees of interference. His results are summarized as follows. Weak interference in both directions is given by the conditions $C_{12} \leq D_{11}$ and $C_{21} \leq D_{22}$. Moderate interference in both directions is described by $D_{11} \leq C_{12} < C_1$ and $D_{22} \leq C_{21} < C_2$. Strong interference in both directions is defined by $C_1 \leq C_{12}$ and $C_2 \leq C_{21}$ (cf. (69)). Moreover, Bergmans [40] characterized the situation where interference in one direction is stronger than in the other. Bergmans [40] showed that in each case there are several undominated modes of operation. These can be derived from (71). In the case of weak interference, (71) reduces to

$$R^{(11)} = (D_{11}, D_{22})$$

$$R^{(10)} = (C_1, C_{21})$$

$$R^{(01)} = (C_{12}, C_2)$$

This situation is represented in Fig. 14. In the case of moderate interference, the undominated rate points are

$$R^{(10)} = (C_1, D_{22})$$

$$R^{(01)} = (D_{11}, C_2)$$

$$R^{(00)} = (C_{12}, C_{21})$$

whereas, in the case of strong interference, the best rate point is

$$R^{(00)} = (C_1, C_2)$$

(See Fig. 13). Bergmans [40] also found that if the interference in the 2-1 direction is stronger than in the 1-2 direction, the rate points

$$R^{(01)} = (C_1, D_2)$$

$$R^{(10)} = (C_{12}, C_2)$$

are undominated modes of operation.

The rate points (71) serve mostly to motivate more general results. Carleial [7] has determined the rate region $\mathcal{R}_C(K_{22})$, defined in (63), for the AWGN-IFC. This region is based on superposition coding, and it turns out that the rate points (71) are included in it. The constraint (62b) corresponds to several decoding strategies which may be chosen by each receiver, leading to four possibilities, represented by (60) and (61). Each configuration imposes six upper bounds on the component rates $R_{11}, R_{22}, R_{12}, R_{21}$.

Theorem 20: For a standard-form AWGN-IFC with two sender-receiver pairs as defined by (65), with average-power constraints $E_1, E_2$, noise powers $N_1, N_2$ and interference coefficients $\alpha_{12}, \alpha_{21}$, the achievable rate region $\mathcal{R}_C(K_{22})$ defined in (63) is equal to the set of rate pairs $R = (R_1, R_2)$ such that

$$R_1 \leq \min \left\{ \gamma_1 E_1 / (N_1 + (1 + \alpha_1) E_1 + \alpha_2 E_2) \right\}^*$$

$$\quad + \left\{ (1 + \alpha_1) E_1 / (N_1 + (1 + \alpha_2) E_2) \right\}^*$$

$$\quad + \left\{ \alpha_1 E_1 / (N_1 + (1 + \alpha_2) E_2) \right\}^*$$

$$\quad + \left\{ \alpha_2 E_2 / (N_2 + (1 + \alpha_1) E_1) \right\}^* \right\}$$

$$R_2 \leq \min \left\{ \gamma_2 E_2 / (N_2 + (1 + \alpha_2) E_2 + \alpha_2 E_2) \right\}^*$$

$$\quad + \left\{ (1 + \alpha_1) E_1 / (N_2 + (1 + \alpha_1) E_1) \right\}^*$$

$$\quad + \left\{ \alpha_2 E_2 / (N_2 + (1 + \alpha_1) E_1) \right\}^* \right\}$$

for some $0 \leq \alpha_1, \alpha_2 \leq 1$, with $\gamma_1 = 1 - \alpha_1 - \beta_1 \geq 0$, $\gamma_2 = 1 - \alpha_2 - \beta_2 \geq 0$, and $Q_* \leq 1/2 \log(1 + Q)$.

The boundary of the achievable rate region defined in Theorem 20 is plotted as a solid line in Fig. 14 for a particular example. The rate points (71) are also sketched. Although the region $\mathcal{R}_C(K_{22})$ could be considered as a good candidate for the capacity region of the Gaussian IFC, it turns out that $\mathcal{R}_C(K_{22})$ is generally a proper subset of $\Theta(K_{22})$. In fact, Carleial [7] has shown that it is sometimes possible to extend the direct-superposition region described in Theorem 20 by means of frequency or time division. This is shown in Fig. 14, where the boundary of the set of rates that can be achieved by frequency or time division is represented by a broken line. This is in contrast with the situation for the Gaussian MAC and Gaussian broadcast channel where the rate regions based on frequency or time division are suboptimum [72].

Sato [41], independently, established two achievable rate regions for the Gaussian IFC. One region is similar to the region $\mathcal{R}_A(K_{22})$ defined in (55) and involves suppression of the interfering source as in [40]. The other region is based on frequency division. Sato [41] also established an outer bound on the capacity region of the Gaussian IFC based on the capacity region of the Gaussian broadcast channel, which will be discussed in Section VIII.
XIX. We conclude that, in the absence of a satisfactory converse, the characterization of the capacity region of the Gaussian interference channel in terms of a single-letter expression remains an unsolved problem.

G. Three Input Users; Two Output Users

Ulrey [33] has characterized the capacity region $\mathcal{C}(K_{sr})$ of the d.m. IFC with $s \geq 3$ transmitters and $r \geq 2$ receivers when all senders send messages simultaneously to all receivers. The characterization of $\mathcal{C}(K_{sr})$ generalizes Theorem 17, but is much more complicated to describe.

V. Broadcast Channels. Preliminaries

The broadcast channel (BC) was introduced by Cover [42]. Here the problem is how to send different pieces of information from a single source to several receivers. We may think of the single source being composed of several subsources, each one producing different information. The messages might be interrelated. In order to model the various degrees of dependency between the subsources, it is convenient to consider several communication situations for the BC, including the cases of sending separate and common information. For the two-receiver BC, three communication situations are sketched in Fig. 15.

The BC has only one input terminal so that the various kinds of information must be coded into one single codeword. The BC itself is specified by a set of OWC’s, one for each receiver, with common input alphabet. The BC, as conceived by Cover [42], attempts to model the situation of a TV broadcaster with multiple receivers, or of a lecturer with many listeners.

A d.m. BC with $r$ receivers consists of a finite input alphabet $\mathcal{A}$, $r$ finite output alphabets $\mathcal{Y}_1, \ldots, \mathcal{Y}_r$, and a set of transition probabilities $P(y_1, \ldots, y_r|x)$, one for each $x \in \mathcal{A}$. The receivers cannot collaborate so that it suffices to consider the marginal transition probabilities $P(y_i|x)$, $i = 1, \ldots, r$. In the broadcast situation, all BC’s with the same marginals are equivalent. In the sequel, we shall assume that the BC has only two receivers.

Classical approaches which can be applied to the BC are time-sharing and maximin, but Cover [42], introducing a superposition random coding scheme, has shown that one can generally transmit at higher rates. Cover [42] originally designed his coding scheme for the binary symmetric BC (BSBC) and the Gaussian BC. The optimality of his scheme for the BSBC was later shown by Wyner [43], and for the Gaussian BC by Bergmans [44]. The BSBC and Gaussian BC are special cases of the degraded BC. Bergmans [45] reformulated Cover’s superposition scheme for the case of the degraded BC and proved a rigorous random coding theorem for it. Later, his rate region was shown to be optimal by Gallager [46] and Ahlswede and Körner [47].

A prime example here is the one given by Blackwell [48].

A d.m. BC with two output users is specified by two stochastic matrices $[P_1(y_1|x), P_2(y_2|x)]$ and thus can be written as $\mathcal{BC}(P_1, P_2)$. We will further denote this channel by $K_{12}$. We distinguish between three different communication situations for $K_{12}$, described as follows. ($K_{12}$I): a separate message is sent to each receiver at rates $R_1$ and $R_2$, respectively, and no common information is sent at all (Fig. 15a). ($K_{12}$II): a common message is sent to both receivers at rate $R_0$, and in addition a private message is sent to receiver one at rate $R_1$ (Fig. 15b). ($K_{12}$III): a common message is sent to both receivers at rate $R_0$, and in addition, each receiver gets separate information at rates $R_1$ and $R_2$, respectively (Fig. 15c).

XX. As of now, it is not known whether the achievable rate region set forth in Theorem 33 below is the true capacity region of the general nondegraded broadcast channel.

A prime example here is the one given by Blackwell [48].

A d.m. BC with two output users is specified by two stochastic matrices $[P_1(y_1|x), P_2(y_2|x)]$ and thus can be written as $\mathcal{BC}(P_1, P_2)$. We will further denote this channel by $K_{12}$. We distinguish between three different communication situations for $K_{12}$, described as follows. ($K_{12}$I): a separate message is sent to each receiver at rates $R_1$ and $R_2$, respectively, and no common information is sent at all (Fig. 15a). ($K_{12}$II): a common message is sent to both receivers at rate $R_0$, and in addition a private message is sent to receiver one at rate $R_1$ (Fig. 15b). ($K_{12}$III): a common message is sent to both receivers at rate $R_0$, and in addition, each receiver gets separate information at rates $R_1$ and $R_2$, respectively (Fig. 15c).

The most general communication situation is ($K_{12}$III). An $(n,M_1,M_2,h_1,h_2)$-code for ($K_{12}$III) consists of $M_1M_2$ codewords $w_{ijk} \in \mathcal{A}^n$; $M_1M_2$ decoding sets $B_{jk} \subseteq \mathcal{B}_1^n$; $M_2M_0$ decoding sets $D_{jk} \subseteq \mathcal{B}_2^n$ such that

$$\frac{1}{M_1M_2M_0} \sum_{i} \sum_{j} \sum_{k} P_i^1(B_{jk} | w_{ijk}) \geq 1 - \mu_1$$

$$\frac{1}{M_1M_2M_0} \sum_{i} \sum_{j} \sum_{k} P_i^2(D_{jk} | w_{ijk}) \geq 1 - \mu_2.$$
VI. The Degraded Broadcast Channel

Definition 1:
A d.m. OWC $A_2 = (A, P_2(z | x), B_2)$ is said to be a degraded version of the d.m. OWC $A_1 = (A, P_1(y | x), B_1)$ if there exists a third channel $D = (B_3, P_3(z | y), B_2)$ such that

$$P_2(z | x) = \sum_y P_3(z | y) P_1(y | x).$$

We write $A_2 = A_1 D$ to denote (76). We express the fact that $A_2$ is a degraded version of $A_1$ by writing $A_2 \subseteq A_1$. The relation $A_2 \subset A_1$ occurs if $A_2$ is the physical cascade of $A_1$ followed by $D$, and also if the input variable $X$ and the outputs $Z$ and $Y$ form a Markov chain $(X, Y, Z)$. The concept of degradedness is a special case of channel inclusion as defined by Shannon [49].

A d.m. BC with two components $A_1, A_2$ is said to be degraded if one of the component channels, $A_2$, is a degraded version of the other channel, i.e., $A_2 \subseteq A_1$. We denote the d.m. degraded BC (DBC) with two components by $(K_{12}, \Pi)$. This DBC can be described by $(P_1, P_2)$ as well as $(P_1, P_3)$, but we prefer the notation $DBC(P_1, P_2)$. A DBC is shown in Fig. 17 in cascade representation.

A special type of $(K_{12}, \Xi)$ occurs when both $A_1$ and $A_2$ are binary symmetric channels. If $A_1$ is a BSC($p_1$) and $A_2$ is a BSC($p_2$), $0 \leq p_1 \leq p_2 \leq 0.5$, then $A_2 = A_1 D$ where $D$ is a BSC($p_3$) with $p_3 = (p_2 - p_1)/(1 - 2p_1)$. Therefore, in this case, $A_1$ and $A_2$ constitute together a DBC which is called the binary symmetric BC and is denoted by BSB($p_1, p_2$).

The DBC has the property that the receiver at the end of the least noisy channel $A_1$ can always see or simulate the output of the noisier channel $A_2$. Therefore, to determine for the DBC the region $\mathcal{E}(K_{12}, \Pi)$, it suffices to characterize either $\mathcal{E}(K_{12}, \Xi)$ or $\mathcal{E}(K_{12}, \Pi)$. Following tradition, we define the capacity region $\mathcal{E}(K_{12}, \Xi)$ of the DBC via $\mathcal{E}(K_{12}, \Pi)$.

An $(n, M_1, M_2, \lambda_1, \lambda_2)$-code for the DBC $(P_1, P_2)$ consists of $M_1 M_2$ codewords $w_{ij} \in \mathbb{A}^n$; $M_1$ decoding sets $B_i \subset \mathbb{B}^n_1$; $M_2$ decoding sets $D_j \subset \mathbb{B}^n_2$ such that

$$\max_i \max_j P^1_i (R_i | w_{ij}) < \lambda_1$$

and

$$\max_i \max_j P^2_i (D_j | w_{ij}) < \lambda_2.$$

Cover [42] conjectured, and Bergmans [45] proved in full generality, the achievability of a certain rate region for $(K_{12}, \Xi)$. This region can be characterized as follows.
$U$ be an artificial RV taking values $u$ in a finite set $\mathcal{U}$ according to a PD $Q_2(u)$. Let $Q_1(x|u)$ be the conditional PD of the channel input $X$ given $U$. Thus, $Q_1(x|u)$ can be thought of as the transition probability of an artificial channel $\mathcal{E} = (\mathcal{U},Q_1(x|u),\mathcal{A})$. The joint distribution of $U,X,Y,Z$ is defined by

$$P(u,x,y,z) = Q_2(u)Q_1(x|u)P_1(y|x)P_3(z|y)$$

with $P_1$ and $P_3$ as in (76). $(U,X,Y,Z)$ forms a Markov chain with prescribed $P_1(y|x)$, $P_3(z|y)$. The probabilities $Q_2(u)$, $Q_1(x|u)$ are parameters in Bergmans’ random coding scheme [45]. Notice that since $A_2 \subseteq A_1$, one has $\overline{E}A_2 \subseteq \overline{E}A_1$ for every premultiplying channel $\mathcal{E}$. Thus the DBC($P_1,P_3$) has the property that

$$I(U;Z) \leq I(U;Y)$$

for all pairs $(Q_2(u),Q_1(x|u))$. Based on (79), we define the region $\mathcal{G}(Q_1,Q_2)$ as the set of all pairs $(R_1,R_2)$ such that:

$$0 \leq R_1 \leq I(X;Y|U) \quad (80a)$$
$$0 \leq R_2 \leq I(U;Z). \quad (80b)$$

Let

$$\mathcal{G}(K_{12},\mathcal{D}) = \bigcup_{Q_1,Q_2} \mathcal{G}(Q_1,Q_2). \quad (81)$$

Bergmans [45] proved the following.

**Theorem 22:** For the discrete memoryless degraded broadcast channel whose component channels have transition probabilities $P_1(y|x)$, $P_2(z|x)$ related by (76), an achievable rate region is given by the set of all $(R_1,R_2)$ which satisfy (80) for some $Q_2(u)$ and $Q_1(x|u)$ giving rise to (78). Thus, with $\mathcal{G}(K_{12},\mathcal{D})$ defined as in (81), one has $\mathcal{G}(K_{12},\mathcal{D}) \subseteq \mathcal{C}(K_{12},\mathcal{D})$.

The proof of Bergmans is based on a random coding scheme in which $M_2 = \exp(nR_2)$ cloud centers $u_j \in \mathcal{U}$ are selected at random according to $Q_2(u)$, and, corresponding to each $u_j$, $M_1 = \exp(nR_1)$ satellite codewords $w_{ij} \in \mathcal{A}$ are selected according to $Q_1(x|u)$. It can be said that the novel idea of superposition coding, initiated in [42] and made precise in [45] through Theorem 22, has had a strong impact on the development of multiple-user information theory.

In Bergmans’ description of $\mathcal{G}(K_{12},\mathcal{D})$, the size of the input alphabet of $\mathcal{E}$ is taken equal to the size of the input alphabet of the DBC, i.e., $|\mathcal{U}| = |\mathcal{A}|$. Conceivably, one could allow arbitrary finite values for $|\mathcal{U}|$, but Gallager [46] has shown that all points on the boundary of $\mathcal{G}(K_{12},\mathcal{D})$ are already obtained if $|\mathcal{U}| = \min (|\mathcal{A}|, |B_1|, |B_2|)$. Gallager [46] also established a reliability function upper bound on the error probabilities associated with Bergmans’ random coding scheme. We next discuss some geometric properties of the region $\mathcal{G}(K_{12},\mathcal{D})$.

Let $C_1 = \max I(X;Y)$ be the capacity of the component channel $A_1$. For the quantities figuring in (80), one has

$$I(U;Z) + I(X;Y|U) \leq I(X;Y) \leq C_1$$

so that $\mathcal{G}(K_{12},\mathcal{D})$ lies entirely underneath the line of slope minus one through $(C_1,0)$. The boundary curve of $\mathcal{G}(K_{12},\mathcal{D})$ can be conveniently described as a function of $R_1$, $0 \leq R_1 \leq C_1$, as follows. As in [46], we define, for all $\lambda \geq 0$,

$$C'(\lambda) = \sup_{Q_1,Q_2} [I(U;Z) + \lambda I(X;Y|U)]. \quad (83)$$

Next we define, for $0 \leq R_1 \leq C_1$, the capacity function

$$C_2(R_1) = \inf_{\lambda \geq 0} [C'(\lambda) - \lambda R_1]. \quad (84)$$

Clearly, $C_2(\cdot)$ is concave and constitutes the boundary of $\mathcal{G}(K_{12},\mathcal{D})$. Alternatively, one could define

$$t_2(R_1) = \sup_{I(X;Y|U) \geq R_1} I(U;Z). \quad (85)$$

Ahlsvede and Körner [47] showed that $t_2(\cdot)$ is concave so that $C_2(\cdot) = t_2(\cdot)$. They actually worked with the inverse function

$$t_1(R_2) = \sup_{I(U;Z) \geq R_2} I(X;Y|U). \quad (86)$$

Gallager [46] showed that in (83) it suffices to take the supremum over those $(Q_1,Q_2)$ for which $|\mathcal{U}| \leq \min (|\mathcal{A}|, |B_1|, |B_2|)$. This applies of course also to (85) and (86). Hence, the supremum in (83), (85), and (86) can be replaced by a maximum. For the BSBC($P_1,P_2$), Cover [42] and Bergmans [45] have shown that $\mathcal{G}(K_{12},\mathcal{D})$ can be
characterized as the set of all \((R_1,R_2)\) such that
\[
0 \leq R_1 \leq R_1(\alpha) \triangleq h(\alpha \ast p_1) - h(p_1) \tag{87a}
\]
\[
0 \leq R_2 \leq R_2(\alpha) \triangleq \log 2 - h(\alpha \ast p_2), \tag{87b}
\]
for \(0 \leq \alpha \leq 0.5\). Here \(h(p)\) is as defined in \((5)\), and \(\alpha \ast p = \alpha(1 - p) + p(1 - \alpha)\). In this case, the capacity function \(C_0(\cdot)\) is completely described by the family of pairs \((R_1(\alpha),R_2(\alpha))\), for \(0 \leq \alpha \leq 0.5\).

Witsenhausen [51] has shown that the set of rate pairs given by \((87)\) also constitutes the capacity region \(C(P_1,P_2)\) of the BSC\((p_1,p_2)\). His result is based on an interesting theorem, due to Wyner and Ziv [50], which relates the entropy of the input and output distribution of a BSC. We shall make a brief digression to describe this result and its generalizations.

Let \(h(p)\) denote the usual entropy function \((\cdot)\), but now restricted to the interval \([0,0.5]\) so that its inverse \(h^{-1}\) is defined unambiguously. For \(0 < p_0 \leq 0.5, p_0\) fixed, consider the function
\[
w_z(u;p_0) = h(p_0 \ast h^{-1}(u)) \tag{88}\]
defined for \(0 \leq u \leq \log 2\). In a remarkable paper, Wyner and Ziv [50] showed the following.

**Theorem 23:** The function \(w_z(u;p_0)\) defined in \((88)\) is strictly convex \(\cup\) and increasing in \(u\) for \(0 < p_0 < 0.5\).

As an application, consider a BSC\((p_0)\) with input \(X\) and output \(Y\), and set \(Pr(X = 0) = p\) and \(p_1 = \min(p_1' - p)\). Then
\[
H(Y) = w_z(h(p_1);p_0) = w_z(h(X);p_0). \tag{89}\]
Thus, if a BSC\((p_0)\) is operated a single time and its input distribution has entropy \(H(X) = u\), then its output distribution has entropy precisely equal to \(w_z(u;p_0)\).

Ahlswede and Körner [52] and, independently, Witsenhausen [51] have generalized the function \(w_z(\cdot;p_0)\) to the case of an arbitrary d.m. channel (DMC). Following [51] and [52], we define, for a DMC \(T\) with input \(X\) and output \(Y\), the characteristic function
\[
g_T(u) = \min \{H(Y):H(X) \geq u\}. \tag{90}\]
This function establishes the sharp lower bound on the output entropy for a given input entropy and a single use of the channel. Ahlswede and Körner [52] and Witsenhausen [51] showed that \(g_T(\cdot)\) is not necessarily convex. Therefore, consider \(f_T\), the convex envelope of \(g_T\). For a BSC\((p_0)\) one has
\[
f_T(u) = g_T(u) = w_z(u;p_0).\]
The main result obtained by Wyner and Ziv [50] is that, for the BSC\((p_0)\), the function \(w_z(\cdot;p_0)\) establishes the sharp lower bound on the output entropy per symbol for a given input entropy per symbol and blocks of arbitrary length and arbitrary joint input distributions. Witsenhausen [51] and Ahlswede and Körner [52] proved the same property for the function \(f_T\) in the case of an arbitrary DMC \(T\). This result can be formulated as follows.

Let \(X\) be an arbitrary input vector of length \(n\) to the DMC \(T\) and let \(Y\) be the corresponding output vector. Let \(P_X\) be an arbitrary PD for \(X\), giving rise to a corresponding PD for \(Y\). Denote the entropies of \(X\) and \(Y\) by \(H(X)\) and \(H(Y)\), respectively. Wyner and Ziv [50] proved the following theorem which is also referred to as Mrs. Gerber’s lemma.

**Theorem 24:** For any binary symmetric channel with crossover probability \(p_0\) and any \(n \geq 1\), if \(H(X) \geq nu\), then \(H(Y) \geq nw_z(u;p_0)\), where \(X\) and \(Y\) are as defined above and \(w_z(\cdot;p_0)\) is as defined in \((88)\).

Ahlswede and Körner [52] and Witsenhausen [51] extended Theorem 24 to Theorem 25.

**Theorem 25:** For any discrete memoryless channel \(T\), and any \(n \geq 1\), if \(H(X) \geq nu\), then \(H(Y) \geq nf_T(u)\), where \(X\) and \(Y\) are as defined above and \(f_T(\cdot)\) is the convex envelope of the function \(g_T(\cdot)\) defined in \((90)\).

From Theorem 25, it follows that
\[
f_T(u) = \inf_{n \geq 1} \frac{1}{n} \min \{H(Y):H(X) \geq nu\}. \tag{91}\]
This establishes \(f_T\) as the sharp lower bound on the output entropy per symbol for a given per-symbol input entropy for a given DMC \(T\), blocks of arbitrary length \(n\), and arbitrary joint input PD’s \(P_X\). This concludes our digression. 4 We now turn to an application of Theorem 24 to BSBC’s.

Using a conditional version of Theorem 24, Wyner [43] proved the following theorem which is the weak converse to Theorem 22 in the case of the BSBC.

**Theorem 26:** If \((R_1,R_2)\) is an achievable rate pair for the binary symmetric broadcast channel with crossover probabilities \(p_1\) and \(p_2\), \(0 \leq p_1 \leq p_2 \leq 0.5\), and \(R_1 \leq R_1(\alpha)\) for some \(0 \leq \alpha \leq 0.5\), then \(R_2 \leq R_2(\alpha)\), with \(R_1(\alpha)\) and \(R_2(\alpha)\) as defined in \((87)\).

Gallager [46] proved the general weak converse to Theorem 22, thus extending Theorem 26 from the case of the BSBC to the arbitrary DBC. Gallager’s setup involves the source rates \((R_1,R_2)\) rather than the signaling rates \((R_1,R_2)\). For two uniformly and independently distributed source outputs, the two formulations coincide.

Let \(W_1\) and \(W_2\) denote the random outputs of two discrete, possibly correlated, sources. Suppose that the pair \((W_1,W_2)\) is connected to two output users through a sequence of \(n\) uses of the DBC\((P_1,P_2)\), \(0 \leq p_1 \leq p_2 \leq 0.5\), and \(R_1 \leq R_1(\alpha)\) for some \(0 \leq \alpha \leq 0.5\). Then \(R_2 \leq R_2(\alpha)\), with \(R_1(\alpha)\) and \(R_2(\alpha)\) as defined in \((87)\).

The main result obtained by Wyner and Ziv [50] is that, for the BSC\((p_0)\), the function \(w_z(\cdot;p_0)\) establishes the sharp lower bound on the output entropy per symbol for a given input entropy per symbol and blocks of arbitrary length and arbitrary joint input distributions. Witsenhausen [51] and Ahlswede and Körner [52] proved the same property for the function \(f_T\) in the case of an arbitrary DMC \(T\). This result can be formulated as follows.

\[
\tilde{R}_1 = \frac{1}{n} H(W_1|W_2) \tag{92a}\]

\(^4\) For a further analysis of the function \(f_T\), see [96].
Let $P_{e,i}$, $i = 1,2$, denote the probability that the outputs of source $i$ and decoder $i$ differ. Gallager [46] proved the following.

**Theorem 27:** If $(R_1,R_2)$ is a pair of source rates outside the achievable rate region $\mathcal{S}(K_{12};D)$ described in (81), then the error probabilities $P_{e,1}$ and $P_{e,2}$ defined above cannot both be arbitrarily small.

Gallager’s proof involves several steps which make use of the crucial assumption that the BC is memoryless and degraded. We highlight a few of these steps. Let $Y_k = (Y_1, \ldots, Y_{k-1})$, $Z_k = (Z_1, \ldots, Z_{k-1})$, and $U_k = (W_2, Y_k)$ for $k = 2, \ldots, n$. Set $U_1 = W_2$. Then

\[
I(W_2; Z) \leq \sum_k I(U_k; Z_k) \tag{93a}
\]

\[
I(W_1; Y | W_2) \leq \sum_k I(X_k; Y_k | U_k). \tag{93b}
\]

Moreover, $(U_k, X_k, Y_k, Z_k)$ is a Markov chain. From this it follows that

\[
I(W_2; Z) + \lambda I(W_1; Y | W_2) \leq n C'(\lambda) \tag{94}
\]

for $\lambda \geq 0$, where $C'(\lambda)$ is defined as in (83). Using (94) and Fano’s inequality, Gallager [46] then proceeds to prove Theorem 27.

Ahlswede and Körner [47] gave another proof of the weak converse for the DBC. Their method of proof combines aspects of both Gallager’s and Wyner’s proof, and is based on the concavity of the function $t_1(\cdot)$ defined in (86). The characterization obtained in [47] is somewhat stronger than the one in [46]. Gallager [46] proved that no rate point dominating the set of points $(R_1, C_2(R_1))$ is attainable, thereby leaving open the question whether or not $C'_2(\cdot)$ is concave. Ahlswede and Körner [47] proved first that $t_1(\cdot)$ is concave and then that no rate point dominating the graph of $t_1(\cdot)$ can be achieved. Making use of (93) and the concavity of $t_1(\cdot)$, they showed that

\[
\sup_{I(W_2; Z) \geq \mu} I(X, Y | W_2) \leq nt_1 \left( \frac{\mu}{n} \right) \tag{95}
\]

for all $\mu \geq 0$. Using this fact, Ahlswede and Körner [47] proved the following.

**Theorem 28:** If $(R_1, R_2)$ is an achievable rate pair for the degraded broadcast channel with transition probabilities $P_1(y|x)$ and $P_2(z|x)$ related by (76), then $R_1 \leq t_1(R_2)$ where $t_1(\cdot)$ is as defined in (86).

Theorem 22 on the one hand, and Theorems 27 and 28 on the other, establish that the region $\mathcal{S}(K_{12};D)$ defined in (81) is the true capacity region of the DBC. Combining the three theorems into one, we obtain the following characterization of the capacity region $\mathcal{S}(K_{12};D)$ of the DBC, given also in [63].

**Theorem 29:** The rate pair $(R_1, R_2)$ is achievable for the degraded broadcast channel $(P_1, P_2)$ if and only if there exists a joint probability assignment $P(u,x,y,z)$ satisfying (78) with $|U| \leq \min (|A|, |B_1|, |B_2|)$ such that

\[
0 \leq R_1 \leq I(X; Y | U) \quad \text{and} \quad 0 \leq R_2 \leq I(U; Z). \tag{96}
\]

Here $|U|$ denotes the cardinality of $U$.

Theorems 27 and 28 are weak converses rather than strong ones, because they give an asymptotic upper bound on the number of messages in direction 1 (as a function of the number of messages in the other direction) only when the error probabilities tend to zero.

Theorem 28 states that, for arbitrary $\epsilon > 0$ and $n$ sufficiently large, an $(n, M_1, M_2, \lambda_1, \lambda_2)$-code for the DBC must satisfy

\[
M_1 \leq \exp \left[ nt_1 \left( \frac{\log M_2}{n} - \epsilon \right) + \epsilon n \right] \tag{96}
\]

only when $\lambda_1$ and $\lambda_2$ are sufficiently small. Notice that (96) closely resembles the corresponding upper bound on the length of the longest code for an OWC [12]. Now the capacity function $t_1(\cdot)$ defined in (86) plays the role of the one-way channel capacity $C$. Since $t_1(\cdot)$ is monotonically decreasing, there is a trade-off between $M_1$ and $M_2$.

A strong converse theorem for the DBC would state that (96) also holds for values of $\lambda_1$ and $\lambda_2$ arbitrarily close to one. Ahlswede, Gács, and Körner [53] have proven such a strong converse by a new method which is partly based on a combinatorial lemma by Margulis [54]. Their result can be described as follows.

Let $(\lambda_1, \lambda_2)$, $0 < \lambda_1, \lambda_2 \leq 1$ be given. A rate pair $(R_1, R_2)$ is said to be $(\lambda_1, \lambda_2)$-achievable for $(K_{12};D)$ if, for any $\epsilon > 0$ and sufficiently large $n$, there exists a code $(n, M_1, M_2, \lambda_1, \lambda_2)$ such that $\log M_i \geq n(R_i - \epsilon)$; $t = 1,2$. Here $\lambda_1$ and $\lambda_2$ are fixed. A rate pair $(R_1, R_2)$ is achievable in the ordinary sense if it is $(\lambda_1, \lambda_2)$-achievable for all $(\lambda_1, \lambda_2)$, $0 < \lambda_1, \lambda_2 \leq 1$. Denote by $\mathcal{C}(\lambda_1, \lambda_2)$ the set of all rate pairs $(R_1, R_2)$ which are $(\lambda_1, \lambda_2)$-achievable. Then

\[
\mathcal{C}(K_{12};D) = \bigcap_{\lambda_1, \lambda_2} \mathcal{C}(\lambda_1, \lambda_2). \tag{97}
\]

By Theorem 29, $\mathcal{C}(K_{12};D) = \mathcal{S}(K_{12};D)$. Ahlswede, Gács, and Körner [63] showed that

\[
\mathcal{C}(\lambda_1, \lambda_2) \subset \mathcal{S}(K_{12};D) \tag{98}
\]

regardless of the choice for $(\lambda_1, \lambda_2)$. In particular, they proved the following.

**Theorem 30:** If for a degraded broadcast channel the rate pair $(R_1, R_2)$ is $(\lambda_1, \lambda_2)$-achievable for a fixed pair $(\lambda_1, \lambda_2)$, $0 < \lambda_1, \lambda_2 \leq 1$, then $R_1 \leq t_1(R_2)$ where $t_1(\cdot)$ is as defined in (86).

The main idea of the proof of this theorem is that the error probability of a code for the DBC can be substantially decreased by enlarging the decoding sets in a particular way. The resulting effect is that the original code becomes a list code, but with so small a list size that Fano’s lemma can still be applied to give a strong converse. This method of proof was first used by Ahlswede and Dueck [55] in the case of the DMC.
It follows from Theorem 30 that
\[ e(\lambda_1, \lambda_2) = G(K_1, n) = e(\lambda_1, \lambda_2), \]  
(99)
so that if \((R_1, R_2)\) is \((\lambda_1, \lambda_2)\)-achievable for some pair \((\lambda_1, \lambda_2)\), then it is achievable for any pair \((\lambda_1, \lambda_2)\).

This establishes the strong converse for the DBC. As a consequence, we obtain the following statement which is a strengthening of (96). Every code \((n, M_1, M_2, \lambda_1, \lambda_2)\) for the DBC must satisfy
\[ M_1 \leq \exp \left[ n \left( \frac{\log M_2}{n} - o(1) \right) + o(n) \right] \]  
(100)
for any \(0 < \lambda_1, \lambda_2 \leq 1\), and any \(n\).

XXII. It would be of interest to investigate whether (100) also holds with \(o(n)\) replaced by \(o(\sqrt{n})\) in the exponent, as in [12, theorem 3.3.1].

Besides proving a strong converse for the DBC, Ahlswede, Gács, and Körner [53] have developed a new and general method to prove strong converses of coding theorems which seems to be applicable to many other problems in information theory. Their method involves the extension of the concept of a decoding set \(B\) corresponding to a sequence \(x\) of letters to that of a decoding set \(B\) corresponding to a set \(A\) of sequences of letters.

Let \(V = (A, P(y \mid x), B)\) be a DMC and \(Q(x)\) a PD on \(A\). Let \(R(y)\) denote the corresponding output PD. For any subset \(B \subseteq B^n\) and any \(0 < \epsilon < 1\), define
\[ \psi_{(c,t)}(B) = \{ x : a^n_1 P^n(B \mid x) \geq 1 - \epsilon \}. \]  
(101)
Thus, \(\psi_{(c,t)}(B)\) is the set of all sequences \(x\) which are \(c\)-decoded by \(B\). Ahlswede, Gács, and Körner [53] were interested in the minimum size of a set \(B\) which satisfies a prescribed lower bound on the size of \(\psi_{(c,t)}(B)\). For this end, consider, for any \(\epsilon \leq 0\),
\[ A_{\psi}(c, \epsilon) \triangleq \min \left\{ R^n(B) : P^n(\psi(B)) \geq \exp \frac{2}{n} n c \right\} \]  
(102)
and
\[ S_{n}(c, \epsilon) = \frac{1}{n} \log_2 A_{\psi}(c, \epsilon). \]  
(103)

Ahlswede, Gács, and Körner [53] have shown that the limit of \(S_{n}(c, \epsilon)\) is independent of \(\epsilon\) for any fixed value of \(c\), and have given a computable formula for this limit. This result enabled them to prove a strong converse for a source coding problem and also aided in the above-mentioned proof of the strong converse for the DBC.

Körner and Marton [56, 57, 58] obtained a variety of strong results on the BC. Here we shall discuss only those results which are related to the DBC.

In [56], Körner and Marton generalized Feinstein's fundamental lemma [59] and determined the asymptotically-exact size of a maximal block code that can be selected from a prescribed subset \(S\) of the input space of a DMC. Denote again this DMC by \(V = (A, P(y \mid x), B)\) and let \(Q(x)\) be a PD on \(A\). Let \(T_n(Q)\) be the set of typical \(n\)-sequences defined in the sense of [12], but in a more specialized way. For any subset \(S \subseteq A^n\) and any \(0 < \epsilon < 1\), define
\[ G_{\psi}(S, \epsilon) = \min \left\{ P^n(B) : B \subseteq B^n, P^n(B \mid w) \geq \epsilon, \quad \text{for all} \ w \in S \right\}. \]  
(104)
The quantity \(G_{\psi}(S, \epsilon)\), which is clearly related to the quantity \(A_{\psi}(c, \epsilon)\) defined in (102), is the minimum probability of a set \(B\) of output sequences which decodes every \(w \in S\) correctly with a chance of at least \(\epsilon\). We now state the generalization of Feinstein's lemma as obtained by Körner and Marton [56].

**Theorem 31:** Let \(0 < \lambda, \delta, \epsilon < 1\) be given.

a) Direct part: For \(n\) sufficiently large, there exists for every subset \(S \subseteq T_n(Q)\), a code \((n, M, \lambda)\) for \(V\) such that all codewords belong to \(S\) and
\[ \frac{1}{n} \log_2 M \geq I(X;Y) - \delta + \frac{1}{n} \log_2 G_{\psi}(S, \epsilon). \]

b) Converse part: For \(n\) sufficiently large, every code \((n, M, \lambda)\) for \(V\) which has the property that all codewords belong to \(T_n(Q)\) must satisfy
\[ \frac{1}{n} \log_2 M \leq I(X;Y) + \delta + \frac{1}{n} \log_2 G_{\psi}(S, \epsilon). \]

Theorem 31 can be regarded as an asymptotic version of [12, Theorem 3.2.1 and Lemma 3.3.1]. Körner and Marton [56] showed that the quantity \(\frac{2}{n} \log_2 G_{\psi}(S, \epsilon)\) is asymptotically independent of \(\epsilon\), much in the same way as Ahlswede, Gács, and Körner [53] have shown that the quantity \(S_{n}(c, \epsilon)\) defined in (103) is independent of \(\epsilon\) for any choice of \(c\). Theorem 31 gives a useful generalization of the local converse theorem of Ahlswede and Dueck [55].

Theorem 31 plays an important role in the work of Körner and Marton [56, 57, 58]. Using the direct part of Theorem 31, Körner and Marton [56, 58] gave a Feinstein-type proof of Bergmans' random coding theorem for the DBC (Theorem 22).

Recall that Theorem 22 is based on a random coding argument using average probability of error. Theorem 21 guarantees that the rate region of Theorem 22 can also be achieved in the sense of maximum probability of error. Körner and Marton [56, 58] proved the same fact with a direct and constructive argument in the sense of Feinstein [59] and Wolfowitz [12], using maximum probability of error throughout. Their proof can be summarized in four steps as follows.

Let a DBC\((P_1, P_2)\) and parameters \(P_1(x \mid u)\) and \(P_2(u)\) yielding the joint distribution (78) be given. Consider two d.m. OWC's, \(V\) and \(W\), with common input alphabet \(\mathcal{U}\), respective output alphabets \(\mathcal{B}_1\), \(\mathcal{B}_2\), and respective transition probabilities \(P_1(y \mid x)\) and \(P_2(x \mid u)\) derived from (78). Fix \(0 < \lambda, \delta < 1\). Let \(n\) be sufficiently large.

1) Select a code \((n, M, \lambda/2)\) for simultaneous use on \(V\) and \(W\). This code consists of \(M_2\) codewords \(u_j \in \mathcal{U}^n; M_2\)
decoding sets \( B_j \subset B_i^1; M_2 \) decoding sets \( D_j \subset B_i^2 \) such that
\[
P_i^1(B_j|w_j) \geq 1 - \lambda/2 \quad (105a)
\]
\[
P_i^2(D_j|w_j) \geq 1 - \lambda/2 \quad (105b)
\]
for \( j = 1, \ldots, M_2 \), and
\[
\frac{1}{n} \log M_2 \geq I(U;Z) - \delta. \quad (106)
\]

2) By the reverse Markov inequality [60], each \( u_j \) can be replaced by a "cloud" \( A_j \subset A^n \), such that \( Q_i^1(A_j|u_j) \geq 1 - 3\lambda/2 \) and
\[
P_i^1(B_j|w) \geq 1 - \lambda/2 \quad (107a)
\]
\[
P_i^2(D_j|w) \geq 1 - \lambda/2 \quad (107b)
\]
for all \( w \in A_j \) and \( j = 1, \ldots, M_2 \).

3) As \( n \) is sufficiently large, one can, for each \( j, 1 \leq j \leq M_2 \), select within \( A_j \) a code \((n,M_1,\lambda/2)\) for use on channel \( P_i(y|x) \). The \( j \)th such code consists of \( M_1 \) codewords \( w_{ij} \subset A_j \subset A^n; M_1 \) decoding sets \( B_{ij} \subset B_i^1 \) such that
\[
P_i^1(B_{ij}|w_{ij}) \geq 1 - \lambda \quad (108)
\]
for \( i = 1, \ldots, M_1 \) and
\[
\frac{1}{n} \log M_1 \geq I(X;Y|U) - \delta. \quad (109)
\]

4) Now set \( B_{ij} = B_j \cap B_{ij} \). Combining (107) and (108), we obtain
\[
P_i^1(B_{ij}|w_{ij}) \geq 1 - \lambda
\]
\[
P_i^2(D_{ij}|w_{ij}) \geq 1 - \lambda
\]
for \( i = 1, \ldots, M_1 \) and \( j = 1, \ldots, M_2 \). Thus the system \( \{w_{ij}, B_j, D_j\} \) is an \((n,M_1,M_2,\lambda,\lambda)\)-code for the DBC \( \{P_i^1,P_i^2\} \) with \( M_1 \) and \( M_2 \) satisfying (109) and (106). This completes our sketch of the proof by Körner and Marton [58] of the direct half of Theorem 29. Ahlswede [61] has informed the author that the same result can be proved by straightforward classical techniques without making use of Theorem 31.

In [57], Körner and Marton also proved a strong converse part, thus providing an alternative proof of Theorem 30. Their proof is more direct than the one given in [53] and follows the style of [12]. It uses the converse part of Theorem 31 as well as another major theorem contained in [56].

Körner and Marton [57] also showed that Theorem 29 holds for a class of BC’s larger than the class of DBC’s. We shall next describe this result. Consider a BC \( K_{12} \) with component channels \( A_1 = (A,P_1(y|x),B_1) \) and \( A_2 = (A,P_2(x|y),B_2) \). Recall that the BC is said to be degraded if there exists a third channel \( D \) such that \( A_2 \equiv A_1|^D \). (Cf. Definition 1.) We denoted this by \( A_2 \subset A_1 \). This relationship defines a partial ordering on the set of component channels. In [57], Körner and Marton considered other relations between component channels \( A_1 \) and \( A_2 \), besides degradedness, which can be interpreted as “channel \( A_1 \) is better for communication than channel \( A_2 \).”

**Definition 2:**

\( A_1 \) is said to be less noisy than \( A_2 \) if
\[
I(U;Z) \leq I(U;Y) \quad (110)
\]
for every pair of parameters \((Q_2(u),Q_1(x|u))\) giving rise to the joint distributions
\[
P(u,x,y) \neq Q_2(u)Q_1(x|u)p_1(y|x) \quad (111a)
\]
\[
P(u,x,y) \neq Q_2(u)Q_1(x|u)p_2(z|x). \quad (111b)
\]
In this case we write \( A_2 \prec A_1 \).

We denote a BC with two components satisfying Definition 2 by \((K_{12},B)\).

**Definition 3:**

\( A_1 \) is said to be more capable than \( A_2 \) if
\[
I(X;Z) \leq I(X;Y) \quad (112)
\]
for all input PD’s \( Q(x) \) on \( A \). We denote this by \( A_2 \prec\prec A_1 \).

Körner and Marton [57] gave several different characterizations of the partial orderings induced by the last two definitions. In view of (79), one clearly has
\[
A_2 \subset A_1 \Rightarrow A_2 \prec\prec A_1 \Rightarrow A_2 \prec A_1 \Rightarrow A_2 \prec A_1. \quad (113)
\]

Thus each definition is more restrictive than the following one. Körner and Marton [57] gave an example of a BC \( K_{12} \) that satisfies Definition 2 but is not of degraded type. Similarly, they [57] conjectured, and Ahlswede [61] proved that there are BC’s which satisfy Definition 3 and not Definition 2. Thus all three notions are different.

Körner and Marton conclude [57] by showing that Theorem 29 carries over from the class of all DBC’s to the class of all BC’s of type \((K_{12},B)\). In particular, they [57] prove the following.

**Theorem 32:** A rate pair \((R_1,R_2)\) is attainable for a broadcast channel \((K_{12},B)\) in situation \((K_{12},I)\) if and only if there exist parameters \((Q_2(u),Q_1(x|u))\) giving rise to the distributions (111) with \(|U| \leq \min (|A|, |B_1|, |B_2|) \) such that
\[
0 \leq R_1 \leq I(X;Y|U) \text{ and } 0 \leq R_2 \leq I(U;Z). \quad (114)
\]

XXIII. The characterization of the capacity region of a BC for which one component is more capable than the other remains an open problem.

We conclude this section with a remark on feedback. Recently, Abbas El Gamal, a doctoral student at Stanford, has shown that feedback cannot increase the capacity region of a D.M. DBC [62]. By “feedback” here is meant that...
the transmitter observes both the received $Y$ and the received $Z$ sequence through a noiseless link. El Gamal's proof is along the lines of [46] except that the variable $U_k$ in (93) is defined as $(W_k, Y_k, Z_k)$.

VII. THE GENERAL BROADCAST CHANNEL

A. Communication Situation $(K_{12,III})$

For the general BC, one is interested in the capacity region $\mathcal{C}(K_{12,III})$ defined in Section V. At present, a simple characterization of $\mathcal{C}(K_{12,III})$ is not known. Both Cover [28] and van der Meulen [38] have independently put forth an achievable rate region which has a relatively simple form and exceeds time-sharing in some important cases. However, all attempts to prove a converse have failed. The author [38] has also characterized $\mathcal{C}(K_{12,III})$ by a limiting expression. In our description of the achievable rate region, we shall follow the approach taken in [38].

Let a BC $K_{12}$ with input alphabet $\mathcal{A}$, transition probabilities $P(y,z|x)$, and marginals $P_1(y|x), P_2(z|x)$ be given. Consider an artificial d.m. test channel

$$\tilde{P} = (\mathcal{U}_1 \times \mathcal{U}_0 \times \mathcal{U}_2, Q(x|u_1,u_0,u_2), \mathcal{A})$$

placed in front of $K_{12}$. The cascade of $\tilde{P}$ followed by $K_{12}$ is a multiway channel with three inputs and two outputs, denoted by $K_{32}$, and with transition probabilities given by

$$P(y,z|u_1,u_0,u_2) = \sum_x P(y,z|x)Q(x|u_1,u_0,u_2).$$

Consider now the modified problem of sending private information from $u_1$ to $y$ and from $u_2$ to $z$, and common information from $u_0$ to $y$ and $z$, all over $K_{32}$. We denote this communication situation by $K_{32}$. Notice that $K_{32}$ contains certain aspects of $K_{22}$ and $K_{32}$ as defined in Section IV.

Clearly, every rate triple attainable for $K_{32}$ is attainable for $(K_{12,III})$. The author [38] has shown that in this context it suffices to restrict attention to pure test channels only. The channel $\tilde{P}$ is said to be pure if its transition probability $Q(x|u_1,u_0,u_2)$ corresponds to a strict mapping $h_3: \mathcal{U}_1 \times \mathcal{U}_0 \times \mathcal{U}_2 \to \mathcal{A}$. (116)

An achievable rate region for $(K_{12,III})$ can now be established by slightly modifying Ahlswede's Theorem 17 and the corresponding Theorem by Ulrey [33] for $K_{32}$ so as to apply to the case $K_{32}$.

With every mapping function $h_3(\cdot)$ defined as in (116), and every triple of PD's $(P_1(u_1), P_0(u_0), P_2(u_2))$, associate the joint PD

$$P(u_1,u_0,u_2, y,z) = P_1(u_1)P_0(u_0)P_2(u_2)P(y,z|h_3(u_1,u_0,u_2)).$$

Based on (117), evaluate the six-tuple

$$I' = (I(U_1;Y|U_0), I(U_2;Z|U_0), I(U_0;Y|U_1), I(U_0;Z|U_2), I(U_1,U_0;Y), I(U_0;U_2;Z)).$$

Do this for every mapping function $h_3(\cdot)$ and every triple $(P_1(u_1), P_0(u_0), P_2(u_2))$, and take the convex hull of the set of all six-tuples thus obtained. Denote this region by $\mathcal{I}_{III}$. A typical six-tuple in $\mathcal{I}_{III}$ is denoted by $J = (J(1;1|0), J(2;2|0), J(0;1|1), J(0;2|2), J(1,0;1), J(0,2;2)).$ (119)

For each $J' \in \mathcal{I}_{III}$, let $\mathcal{S}_{III}(J')$ be the set of all triples $(R_1, R_2, R_0)$ such that

$$0 \leq R_1 \leq J(1;1|0),$$

$$0 \leq R_2 \leq J(2;2|0),$$

$$0 \leq R_0 \leq \min(J(1;0;1), J(0;2;2)),$$

$$R_1 + R_0 \leq J(1,0;1),$$

$$R_0 + R_2 \leq J(0,2;2).$$

Define

$$\mathcal{S}(K_{12,III}) = \bigcup_{J' \in \mathcal{I}_{III}} \mathcal{S}_{III}(J').$$

The region $\mathcal{S}(K_{12,III})$ is the union, over all vectors $J'$ defined as in (119), of the regions $\mathcal{S}_{III}(J')$ described by (120). (Notice the similarity between expressions (118) through (120) on the one hand, and (49) through (51) on the other.) Cover [28] and the author [38] have independently proved the following.

Theorem 33: For the general discrete memoryless broadcast channel with two receivers, the region $\mathcal{S}(K_{12,III})$ defined in (121) is an achievable rate region. Thus,

$$\mathcal{S}(K_{12,III}) \subset \mathcal{C}(K_{12,III}).$$

Cover [28] has shown that in (121) it suffices to take the union over those points $J' \in \mathcal{I}_{III}$ which are convex mixtures of at most six six-tuples $I'$, each one of the type (118).

Whereas the author’s proof is an adaptation of the Ahlswede–Ulrey theorem [17], [33], the one given by Cover [28] is quite different. It involves the idea of jointly typical sequences introduced in [27], and uses a simultaneous asymptotic equipartition property. For a clear exposition of his method, see [28] and [20].

Using Theorem 33, the author [38] has characterized $\mathcal{C}(K_{12,III})$ by a limiting expression in the following way. For each $n \geq 1$, denote by $K_{12}(n)$ the product channel of length $n$ derived from $K_{12}$. In [38], the author proved the following.

Theorem 34: Let $F_n$ be the achievable rate region evaluated according to Theorem 33, but reduced in scale by a factor $1/n$, of the general discrete memoryless broadcast channel $K_{12}(n)$ of length $n$ described above. Then, as $n \to \infty$, the regions $F_n$ approach a limit $F$ which is the capacity region of the original broadcast channel $K_{12}$. Thus,

$$\mathcal{C}(K_{12,III}) = \lim_{n \to \infty} \frac{1}{n} \mathcal{S}(K_{12}(n), III).$$
As was mentioned at the beginning of this section, it is not known whether the region \( \mathcal{J}(K_{12};II) \) defined in (121) is the true capacity region, so that a simple characterization of \( \mathcal{C}(K_{12};II) \) still needs to be found (Problem XX).  

XXIV. It is also unknown whether the sizes of the input alphabets \( \mathcal{U}_1, \mathcal{U}_0, \mathcal{U}_2 \) can be restricted to a finite number without changing the region \( \mathcal{J}(K_{12};III) \). A similar fact was shown to be true for the DBC by Gallager [46]. It has been conjectured [38] that for a general d.m. BC, it suffices to take \( |\mathcal{U}_1| = \min (|\mathcal{A}|, |B_1|), |\mathcal{U}_0| = \min (|\mathcal{A}|, |B_1|, |B_2|), \) and \( |\mathcal{U}_2| = \min (|\mathcal{A}|, |B_2|) \).

B. Communication Situation \( (K_{12},II) \)

Based on Theorem 33, the author [38] derived an attainable rate region for \( (K_{12},II) \). Subsequently, Körner and Marton [58] found another equivalent characterization and showed, by proving a strong converse, that this region is the true capacity region \( \mathcal{C}(K_{12};II) \). Körner and Marton [58] referred to communication situation \( (K_{12},II) \) as a BC with degraded messages.

If we intersect both expressions in Theorem 33 with the plane \( R_2 = 0 \), we obtain a coding theorem for \( (K_{12},II) \). The projection of \( \mathcal{C}(K_{12},III) \) into the \( (R_1,R_0) \)-plane is

\[
\mathcal{C}(K_{12},II) = \{ (R_1,R_0): (R_1,0,R_0) \in \mathcal{C}(K_{12},III) \}. \tag{122}
\]

The projection of \( \mathcal{J}(K_{12},III) \) into the \( (R_1,R_0) \)-plane is defined similarly and denoted by \( \mathcal{J}(K_{12},II) \). By Theorem 33, one has

\[
\mathcal{J}(K_{12},II) \subset \mathcal{C}(K_{12},II). \tag{123}
\]

The two-dimensional rate region \( \mathcal{J}(K_{12},II) \) can be characterized by various simplified expressions in the following way. First notice that if \( R_2 = 0 \), then in (120) inequality (120d) disappears, and since

\[
I(U_0;Z|U_2) \leq I(U_0;U_3;Z), \tag{124}
\]

inequality (120e) becomes redundant because of (120c). Moreover, the conditioning variable can be taken out of \( J(0;Z|2) \) in (120c). This leads to the following characterization for \( \mathcal{J}(K_{12},II) \).

Characterization 8: With every mapping function \( h_2: \mathcal{U}_1 \times \mathcal{U}_0 \rightarrow \mathcal{A} \) and every pair of PD’s \( (P_1(u_1),P_0(u_0)) \), associate the joint probability distribution

\[
P(u_1,u_0,y,z) = P_1(u_1)P_0(u_0)P(y,z|h_2(u_1,u_0)). \tag{125}
\]

Based on (125), evaluate the quadruple

\[
I^* = (I(U_1;Y|U_0),I(U_0;Y|U_1),I(U_0;Z),I(U_1;U_0;Y)). \tag{126}
\]

Do this for every mapping \( h_2(\cdot) \) and every pair \( (P_1(u_1),P_0(u_0)) \). Take the convex hull of the set of all quadruples thus obtained and denote this region by \( \mathcal{J}_{II} \). A typical quadruple in \( \mathcal{J}_{II} \) is denoted by

\[
I^* = (J(1;1|0),J(0;1|1),J(0;2),J(1,0;1)). \tag{127}
\]

For each \( I^* \in \mathcal{J}_{II} \), let \( \mathcal{J}_{II}(I^*) \) be the set of all pairs \( (R_1,R_0) \) such that

\[
\begin{align*}
0 &\leq R_1 \leq J(1;1|0) \tag{128a} \\
0 &\leq R_0 \leq \min (J(0;0|1),J(0;2)) \tag{128b} \\
R_1 + R_0 &\leq J(1,0;1). \tag{128c}
\end{align*}
\]

Define

\[
\mathcal{J}(K_{12},II) = \bigcup_{I^* \in \mathcal{J}_{II}} \mathcal{J}_{II}(I^*). \tag{129}
\]

Thus the region \( \mathcal{J}(K_{12},II) \) is the union over all vectors \( I^* \) defined in (127) of the regions \( \mathcal{J}_{II}(I^*) \) defined in (128). In [38], the author proved the following.

Theorem 35: For the discrete memoryless broadcast channel in communication situation \( (K_{12},II) \) (also referred to as a broadcast channel with degraded messages), the rate region \( \mathcal{J}(K_{12},III) \) established in (121) reduces to the rate region \( \mathcal{J}(K_{12},II) \) defined in (129). Thus,

\[
\mathcal{J}(K_{12},II) = \mathcal{J}(K_{12},II) \subset \mathcal{C}(K_{12},II).
\]

The author [38], [63] has specialized Theorem 35 to the case of the DBC and obtained an alternative expression for the capacity region \( \mathcal{C}(K_{12},II) \) defined in Section VI; details of this are presented in subsection VII-C. The basic idea (see [63]) is that every triple \( (h_2(u_1,u_0),P_0(u_0),P_1(u_1)) \) defined as above corresponds to a pair \( (Q_3(u_0),Q_1(x|u_0)) \) occurring in (18), and vice versa. On the one hand, we can choose \( Q_2 \equiv P_0 \) and set

\[
Q_1(x|u_0) = \sum_{u_1:h_2(u_1,u_0)=x} P_1(u_1). \tag{130}
\]

Conversely, every \( Q_1(x|u_0) \) can be written as a finite probability mixture of pure channels \( h_2: \mathcal{U}_1 \times \mathcal{U}_0 \rightarrow \mathcal{A} \), as was noted in [49]. We denote this correspondence by

\[
(h_2(u_1,u_0),P_0(u_0),P_1(u_1)) \rightarrow (Q_3(u_0),Q_1(x|u_0)). \tag{131}
\]

Whereas the author [38], [63] used this correspondence to establish Theorem 40 below, Körner and Marton [58] applied it to prove that the region \( \mathcal{J}(K_{12},II) \) defined in (129) is equal to another rate region \( \mathcal{M}(K_{12},II) \), to be described now.

Characterization 7: Let a general d.m. BC \( K_{12} \) with marginals \( P_1(y|x) \) and \( P_0(x|u) \) be given. For every pair of parameters \( (Q_2(u),Q_1(x|u)) \), define the two joint PD's \( P(u,x,y) \) and \( P(u,x,z) \) as in (111). Based on (111), derive the mutual informations \( I(X;Y|U),I(U;Z),I(X;Y) \). Let \( \mathcal{M}_{II}(Q_1,Q_2) \) be the set of all pairs \( (R_1,R_0) \) such that

\[
\begin{align*}
0 &\leq R_1 \leq I(X;Y|U) \tag{132a} \\
0 &\leq R_0 \leq I(U;Z) \tag{132b} \\
R_1 + R_0 &\leq I(X;Y). \tag{132c}
\end{align*}
\]

Let

\[
\mathcal{M}(K_{12},II) = \bigcup_{Q_1,Q_2} \mathcal{M}_{II}(Q_1,Q_2). \tag{133}
\]
The region $\mathcal{M}(K_{12}, II)$ can be described as the union, over all pairs $(Q_1(x), Q_2|x|)$ giving rise to the distributions (111), of the sets of rates $(R_1, R_0)$ which satisfy (132). Körner and Marton [58] stated that in (133) it suffices to take the union over those parameter channels $\mathcal{E} = (\mathcal{U}, Q_1|x|, \mathcal{A})$ for which $|\mathcal{U}|$ is finite, as is the case for the DBC [46]. Körner and Marton [58] proved the following.

**Theorem 36:** For the discrete memoryless broadcast channel with degraded messages, the region $\mathcal{V}(K_{12}, II)$ established in (129) equals the region $\mathcal{M}(K_{12}, II)$ defined in (133). Moreover, the region $\mathcal{M}(K_{12}, II)$ is the capacity region of this channel. Thus,

$$\mathcal{V}(K_{12}, II) = \mathcal{M}(K_{12}, II)$$

(a)

$$\mathcal{E}(K_{12}, II) = \mathcal{M}(K_{12}, II).$$

(b)

The proof of part (a) uses, besides correspondence (131), the fact that if $R_1 = I(U_1; Y|U_0)$, then

$$R_0 + R_1 \leq I(U_1, U_0; Y)$$

if and only if $R_0 \leq I(U_0; Y)$,

$$\text{(134)}$$

whereas $I(U_0; Y) \leq I(U_0; Y|U_1)$ always.

With part (a) proved, the direct half of part (b) becomes equivalent to Theorem 35. However, Körner and Marton [58] gave an independent proof of the direct half of part (b) similar to the proof of the coding theorem for the DBC which we discussed above following Theorem 30. The most interesting part of Theorem 36 is the strong converse part of (b) which Körner and Marton [58] proved in the same fashion as the strong converse part of Theorem 32 established in [57].

We now reformulate Theorem 36(b) in the same way as we combined Theorems 22, 27, and 29 into Theorem 29. (See [58].)

**Theorem 37:** A rate pair $(R_1, R_0)$ is $(\lambda_1, \lambda_2)$-achievable for the broadcast channel with degraded messages, denoted by $(K_{12}, II)$, for a fixed pair $(\lambda_1, \lambda_2)$, $0 < \lambda_1, \lambda_2 \leq 1$, if and only if there exist parameters $(Q_1, Q_2)$ which yield the joint distributions (111) with $|\mathcal{U}| \leq \min (|\mathcal{A}|, |\mathcal{B}_1|, |\mathcal{B}_2|)$ such that (132) holds. Here $|\mathcal{U}|$ denotes the cardinality of the range of the variable $U$.

**Characterization 8:** Based on (134) and Theorem 37, we present here yet another characterization of the capacity region $\mathcal{E}(K_{12}, II)$, which is new but straightforward in view of the preceding theorem.

**Theorem 38:** A rate pair $(R_1, R_0)$ is $(\lambda_1, \lambda_2)$-achievable for the broadcast channel $(K_{12}, II)$ for a fixed pair $(\lambda_1, \lambda_2)$, $0 < \lambda_1, \lambda_2 \leq 1$, if and only if there exist parameters $(Q_1, Q_2)$ which yield the joint distributions (111) with $|\mathcal{U}| \leq \min (|\mathcal{A}|, |\mathcal{B}_1|, |\mathcal{B}_2|)$ such that

$$0 \leq R_1 \leq I(X; Y|U)$$

(135a)

$$0 \leq R_0 \leq I(U; Z)$$

(135b)

$$0 \leq R_0 \leq I(U; Y).$$

(135c)

Because of (134) and (131), one sees directly that (135) follows from (128) and is equivalent to (132), whereas in the case of the DBC (135) implies (80). This last property makes Characterization 8 perhaps the most perspicuous one among the three characterizations given in this subsection.

For a BC $K_{12}$ for which one component is less noisy than the other component (see Definition 2, Section VI), we notice that Theorem 38 reduces to the following.

**Theorem 39:** A rate pair $(R_1, R_0)$ is attainable for a broadcast channel for which one component is less noisy than the other component in situation $(K_{12}, II)$ if and only if there exist parameters $(Q_1, Q_2)$ giving rise to the joint distributions (111) with $|\mathcal{U}| \leq \min (|\mathcal{A}|, |\mathcal{B}_1|, |\mathcal{B}_2|)$ such that (114) holds.

Theorems 32 and 39 imply that, for channels of type $(K_{12}, B)$, the rate regions $\mathcal{E}(K_{12}, I)$ and $\mathcal{E}(K_{12}, II)$ coincide. In this case, the three-dimensional region $\mathcal{E}(K_{12}, III)$ is completely determined by either of the above regions and can be constructed by time-sharing between the boundaries of $\mathcal{E}(K_{12}, I)$ and $\mathcal{E}(K_{12}, II)$. (Compare Fig. 16 in this regard.) Like the DBC, channels of type $(K_{12}, B)$ have the property that $(R_1, R_2, R_0) \in \mathcal{E}(K_{12}, III)$ if and only if $(R_1, R_0 + R_2) \in \mathcal{E}(K_{12}, II)$ and in turn if and only if $(R_1, R_0 + R_2) \in \mathcal{E}(K_{12}, I)$. Poltyrev [64] has investigated the general class of BC's for which this property holds.

**C. Comparison with DBC**

In [38] and [63], the author gave an alternative expression for the capacity region of the DBC based on Theorem 35 and the correspondence (131). In the case of the DBC, the mutual informations figuring in (126) satisfy

$$I(U_0; Z) \leq I(U_0; Y) \leq I(U_0; Y|U_1)$$

(136a)

$$I(U_0; Z) + I(U_1; Y|U_0) \leq I(U_1, U_0; Y)$$

(136b)

with the result that in (128) the inequality (128c) disappears and (128b) reduces to $0 \leq R_0 \leq J(0, 2)$. This leads to the following.

**Alternative Characterization of $\mathcal{E}(K_{12}, D)$:** Let a d.m. DBC $(P_1, P_2)$ with $P_1$ and $P_2$ satisfying (76) be given. Associate with every mapping function $h_1: \mathcal{U}_1 \times \mathcal{U}_2 \rightarrow \mathcal{A}$ and every pair $(P_1(u_1), P_2(u_2))$ of PD's the joint PD $P(u_1, u_2, y, z) = P_1(u_1)P_2(u_2)P_1(y|h_1(u_1, u_2))P_2(z|y)$.

(137)

Denote by $\mathcal{R}_D(h_1, P_1, P_2)$ the set of all pairs $(R_1, R_2)$ such that

$$0 \leq R_1 \leq I(U_1; Y)$$

(138a)

$$0 \leq R_2 \leq I(U_2; Z).$$

(138b)

Let

$$\mathcal{V}(K_{12}, D) = \co \left( \bigcup_{h_1} \mathcal{R}_D(h_1, P_1, P_2) \right),$$

(139)

where the union is taken over all pure test channels $h_1$ and product PD's $P_1 \times P_2$. In [38], [63], the author proved the following.
Theorem 40: An alternative description of the capacity region of the degraded broadcast channel is given by the rate region $\mathcal{V}(K_{12},D)$ defined in (139). More in particular,

$$\mathcal{V}(K_{12},D) = \mathcal{S}(K_{12},D) = \mathcal{C}(K_{12},D)$$

where $\mathcal{S}(K_{12},D)$ is the rate region defined in (81).

D. Communication Situation $(K_{12},I)$

If we intersect both expressions in Theorem 33 with the plane $R_o = 0$, we obtain a coding theorem for $(K_{12},I)$. The projection of $\mathcal{C}(K_{12},III)$ into the $(R_1,R_2)$ plane is

$$\mathcal{C}(K_{12},II) \triangleq \{(R_1,R_2) : (R_1,R_2,0) \in \mathcal{C}(K_{12},III)\}.$$  

(140)

The intersection of $\mathcal{S}(K_{12},III)$ with $R_o = 0$ is defined similarly and denoted by $\mathcal{S}(K_{12},I)$. By Theorem 33, we have

$$\mathcal{S}(K_{12},I) \subseteq \mathcal{C}(K_{12},I).$$  

(141)

In [38], the author obtained the following simple characterization for $\mathcal{S}(K_{12},I)$. With every strict mapping $h_o: \mathcal{U}_1 \times \mathcal{U}_2 \rightarrow \mathcal{A}$ and every pair of PD's $(P_1(u_1), P_2(u_2))$, we associate the joint PD

$$P(u_1,u_2,y,z) = P_1(u_1)P_2(u_2)P(y,z|h_o(u_1,u_2)).$$  

(142)

Denote by $\mathcal{R}_1(h_0,P_1,P_2)$ the set of all pairs $(R_1,R_2)$ such that

$$0 \leq R_1 \leq I(U_1;Y)$$  

(143a)

$$0 \leq R_2 \leq I(U_2;Z).$$  

(143b)

Let

$$\mathcal{V}(K_{12},I) = \bigcup_{h_0, P_1,P_2} \mathcal{R}_1(h_0,P_1,P_2).$$  

(144)

Thus, $\mathcal{V}(K_{12},I)$ is the convex hull of the union, over all mappings $h_o$ and PD's $P_1(u_1)$ and $P_2(u_2)$, of the sets of rates $(R_1,R_2)$ satisfying (143). In [38], the author proved the following.

Theorem 41: For the discrete memoryless broadcast channel in communication situation $(K_{12},I)$, an achievable rate region is given by the region $\mathcal{V}(K_{12},I)$ defined in (144).

Notice that for each $h_o$, the channel $P(y,z|h_0(u_1,u_2))$ is an IFC of the type $K_{22}$ considered in subsection IV-B and the region

$$\bigcup_{P_1,P_2} \mathcal{R}_1(h_0,P_1,P_2)$$

which occurs in (144) becomes the region $\mathcal{S}_1(K_{22})$ defined in (44a), so that Theorem 41 can also be obtained via Theorem 15. Indeed, communication situation $(K_{12},I)$ bears much resemblance to $K_{22}$. E. Blackwell's Example

Blackwell [48] gave the following example of a d.m. BC. Let $\mathcal{A} = \{0,1,2\}$, $\mathcal{B}_1 = \mathcal{B}_2 = \{0,1\}$, and let the transition probabilities be defined by

$$P(y = 0,z = 1|x = 0) = Pr(y = 1,z = 0|x = 1)$$

$$= Pr(y = 1,z = 1|x = 2) = 1.$$  

(145)

It is easily verified that this example does not fall into the class of DBC's considered in Section VI. Clearly, the capacities of the component channels are both equal to one. By time-sharing, all pairs $(R_1,R_2)$, such that $R_1 > 0$, $R_2 > 0$ and $R_1 + R_2 \leq 1$, are attainable in situation $(K_{12},I)$. Blackwell [48] noticed that the point $(0.793,0.793)$ is outside the capacity region $\mathcal{C}(K_{12},I)$.

Applying Theorem 41, the author [38] has found an achievable rate region for Blackwell's example in situation $(K_{12},I)$, which considerably exceeds the set of rates found by time-sharing. Let $h(p)$ denote the entropy function defined in (5) and let

$$C(p) = \begin{cases} \log_2 \left[ 1 + \exp \left( - \frac{h(p)}{1-p} \right) \right], & \text{if } 0 \leq p < 1 \\ 0, & \text{if } p = 1. \end{cases}$$  

(146)

Using test channels of length one, the author found that all pairs $(R_1,R_2)$ such that

$$0 \leq R_1 \leq h(p) \text{ and } 0 \leq R_2 \leq C(p)$$  

are achievable for $0 \leq p \leq 1$. Similarly, all pairs satisfying

$$0 \leq R_1 \leq C(p) \text{ and } 0 \leq R_2 \leq h(p)$$  

are achievable for this example in situation $(K_{12},I)$. In particular, it was found that the points $(R_1,R_2) = (1,0.3215)$ and $(R_1,R_2) = (0.694,0.694)$ are in the capacity region $\mathcal{C}(K_{12},I)$.

The corresponding rate region is sketched in Fig. 18.

S. Gelfand [65], in a lecture at the 1975 IEEE-USSR Moscow Workshop on Information Theory, has stated that by using test channels of length two it is possible to achieve a point $(R_1,R_2)$ with $R_1 = 1$ and $R_2 > 0.3215$. Sato [66] has shown that no rate pair dominating the point $(R_1,R_2) = (1,0.5)$ can be achieved for the present example. The latter result was conjectured in [63].

F. Outer Bounds

In his original article [42], Cover gave an outer bound for the capacity region $\mathcal{C}(K_{12},I)$. This bound has recently been improved by Sato [66].

Let a BC $K_{12}$ with marginals $P_1(y|x), P_2(z|x)$, and joint transition probabilities $P(y,z|x)$ be given. Denote by $K_{12}$ the class of all BC's with the same marginals as $K_{12}$. Let $Q(x)$ be a PD on inputs. Denote by $\mathcal{S}(K_{12},Q)$ the set of all pairs $(R_1,R_2)$ such that

$$0 \leq R_1 \leq I(X;Y)$$  

(149a)

$$0 \leq R_2 \leq I(X;Z).$$  

(149b)

Let

$$\mathcal{S}_0(K_{12},I) = \bigcup_Q \mathcal{S}(K_{12},Q).$$  

(150)

Cover [42] proved the following.

Theorem 42: An outer bound on the capacity region of the discrete memoryless broadcast channel with two receivers, where separate information is sent to both, is given
by the region $\mathcal{C}_0(K_{12},I)$ defined in (150). That is,
$$
\mathcal{C}(K_{12},I) \subseteq \mathcal{C}_0(K_{12},I).
$$

Sato [66] tightened this outer bound as follows. Let $\mathcal{J}^*(K_{12},Q)$ be the set of pairs $(R_1,R_2)$ satisfying (149) and, in addition,
$$
R_1 + R_2 \leq I(X;Y,Z). \quad (151)
$$
Define the region
$$
\mathcal{J}(K_{12},I) = \co \left( \bigcup_Q \mathcal{J}^*(K_{12},Q) \right). \quad (152)
$$
Clearly, $\mathcal{J}(K_{12},I) \subseteq \mathcal{C}_0(K_{12},I)$. Because of (151), $\mathcal{J}(K_{12},I)$ depends on the joint transition probabilities $P(y,z|x)$. However, the capacity region of the BC depends only on the marginals. Therefore, define
$$
\mathcal{J}(K_{12},I) = \bigcap_{K'_{12} \subseteq K_{12}} \mathcal{J}(K'_{12},I) \quad (153)
$$
where the intersection is taken over all BC’s with the same marginals as $K_{12}$. Sato [66] proved the following.

**Theorem 43:** A new outer bound on the capacity region of the discrete memoryless broadcast channel with two receivers in communication situation ($K_{12}$ is given by the region $\mathcal{J}(K_{12},I)$ defined in (153). This bound is tighter than the bound $\mathcal{C}_0(K_{12},I)$ defined in (150). That is,
$$
\mathcal{C}(K_{12},I) \subseteq \mathcal{J}(K_{12},I) \subseteq \mathcal{C}_0(K_{12},I).
$$

This outer bound can be extended to the case ($K_{12},III$). Outer bounds like this one are useful as long as simple expressions for the capacity regions $\mathcal{C}(K_{12},I)$ and $\mathcal{C}(K_{12},III)$ remain unknown.

**VIII. GAUSSIAN BROADCAST CHANNELS**

The time-discrete Gaussian BC was first discussed by Cover [42]. An additive white Gaussian noise (AWGN) BC with two receivers has sequences of real numbers $x = (x_1, \ldots, x_n)$ as inputs, and sequences $y_1 = (y_{11}, \ldots, y_{1n})$ and $y_2 = (y_{21}, \ldots, y_{2n})$ as corresponding outputs. The transitions are characterized by $Y_1 = X + Z_1$, $Y_2 = X + Z_2$, where $Z_1 = (Z_{11}, \ldots, Z_{1n})$, $Z_2 = (Z_{21}, \ldots, Z_{2n})$ are sequences of independent, identically distributed (i.i.d) normal RV’s with mean zero and noise variances $N_1$ and $N_2$. Moreover, $X$ is independent of $(Z_1, Z_2)$. In the broadcast formulation, it is irrelevant whether $Z_1$ and $Z_2$ are correlated or not.

Assume that there is an average power constraint on inputs, given for every $n$ by
$$
\frac{1}{n} \sum x_i^2 \leq E. \quad (154)
$$
Let $N_1 < N_2$. We denote this AWGN-BC by $K_{12}(N_1,N_2,E)$. We are interested in the capacity region $\mathcal{C}(K_{12},AWGN)$ which is defined as the set of all pairs $(R_1,R_2)$ of simultaneously achievable rates. It is well known [67] that the individual capacities are
$$
C_i(E) \triangleq \frac{1}{2} \log \left( 1 + \frac{E}{N_i} \right), \quad i = 1,2. \quad (155)
$$
From these one-way capacities, one can derive the time-sharing rates $(\tau C_1(E) + (1-\tau) C_2(E))$, for $0 \leq \tau \leq 1$.

Naive superposition (cf. Theorems 15 and 41) yields the simultaneous rates
$$
C_i(E_i) \triangleq \frac{1}{2} \log \left( 1 + \frac{E_i}{E_j + N_i} \right), \quad \text{for } i = 1,2; \quad j = 3 - i; \quad E_1 + E_2 = E. \quad (156)
$$
for $i = 1,2$; $j = 3 - i$; $E_1 + E_2 = E$.

Cover’s innovative superposition scheme [42] allows the simultaneous transmission at rates which dominate the two previous rate pairs. Let the signal $S_2 = (S_{21}, \ldots, S_{2n})$, intended for the high noise receiver $Y_2$, be a sequence of $n$ i.i.d. $N(0,E_2)$ RV’s, and let the signal $S_1 = (S_{11}, \ldots, S_{1n})$, intended for the low noise receiver $Y_1$, be a sequence of $n$ i.i.d. $N(0,E_1)$ RV’s with $E_1 + E_2 = E$. The two input sequences are superimposed on each other yielding the actual
input sequence $X = S_1 + S_2$. Signal $S_1$ and noise $Z_2$ are considered to be noise by receiver 2. Therefore, messages can be sent to receiver 2 at rates $R_2 \leq \bar{C}_2(E_2)$. But any message which can be decoded without error by receiver 2 can also be decoded without error by receiver 1. Therefore, receiver 1 can subtract $S_2$ from $Y_1$ yielding $Y_1 \triangleq Y_1 - S_2 = S_1 + Z_1$. Thus the interference effect of the receiver 2 codeword can be suppressed in the transmission to receiver 1. This allows transmission to receiver 1 at rates $R_1 \leq \bar{C}_1(E_1)$. For a full and clear exposition of these ideas, see [42], [19], and [20]. Cover [42] proved the following.

**Theorem 44:** For the additive white Gaussian noise broadcast channel $K_0(N_1, N_2, E)$ defined above, every rate pair $(R_1, R_2)$ such that

$$0 \leq R_1 \leq \bar{C}_1(\alpha) \leq 1 - \frac{1}{\alpha} \log \left(1 + \frac{\alpha E}{N_1}\right)$$

$$0 \leq R_2 \leq \bar{C}_2(\alpha) \leq 1 - \frac{1}{\alpha} \log \left(1 + \frac{1 - \alpha E}{E + N_2}\right)$$

is achievable for any $0 \leq \alpha \leq 1$.

We denote the set of all $(R_1, R_2)$ satisfying (157) by $\mathcal{S}(K_{12}, \text{AWGN})$. Bergmans [44], by proving a weak converse, showed that this rate region is the true capacity region. Following [44], we use the equivalent notation

$$\bar{C}_1(\alpha) = g(N_1 + \alpha E) - g(N_1)$$

$$\bar{C}_2(\alpha) = g(N_2 + E) - g(N_2 + \alpha E)$$

where

$$g(E) \triangleq \frac{1}{2} \log (2\pi eE).$$

Bergmans [44] derived the following theorem which is the Gaussian counterpart of Theorem 24.

**Theorem 45:** Consider a time-discrete additive white Gaussian noise one-way channel $Y = X + Z$ with arbitrary input sequence $X$ of length $n$, and independent noise vector $Z$ consisting of $n$ components which are independent identically distributed $\mathcal{N}(0, N)$ random variables. If, for any $n \geq 1$, $H(X) \geq n\mu$, then $H(Y) \geq n g(N + g^{-1}(\mu))$.

This theorem is a consequence of an inequality by Shannon [68] on the entropy power of the sum of two ensembles, a proof of which was given by Stam [69] and Blachman [70]. Using the conditional version of Theorem 45, Bergmans [44] proved the following converse to Theorem 44.

**Theorem 46:** If $(R_1, R_2)$ is achievable for the additive white Gaussian noise broadcast channel $K_0(N_1, N_2, E)$, and $R_1 \geq \bar{C}_1(\alpha)$ for some $\alpha$, $0 \leq \alpha \leq 1$, then $R_2 \leq \bar{C}_2(\alpha)$, where $\bar{C}_1(\alpha)$ and $\bar{C}_2(\alpha)$ are as defined in (157).

XXV. Since Theorem 46 is a weak converse, and a strong converse was established for the d.m. DBC in Theorem 30, it would be desirable to prove also a strong converse for the AWGN-BC.

In [71] and [44], Bergmans generalized Theorem 44 and Theorem 46 to the case of AWGN BC's with any finite number of outputs. We combine Theorems 44 and 46 into Theorem 47.

**Theorem 47:** The capacity region of the additive white Gaussian noise broadcast channel with two receivers, with noise powers $N_1 < N_2$ and average input power $E$, is equal to the set of rates $(R_1, R_2)$ defined in (157) as $\alpha$ varies between 0 and 1. That is,

$$\mathcal{C}(K_{12}, \text{AWGN}) = g(K_{12}, \text{AWGN}).$$

Bergmans and Cover [72] also studied the continuous-time band-limited Gaussian BC. Suppose a transmitter is transmitting to two receivers 1 and 2 in the presence of additive white Gaussian noise of one-sided spectral density $N_1$ and $N_2$, respectively, with $N_1 \leq N_2$. Let the input be power-constrained by $E$ and suppose that the total available bandwidth is $W$. The situation is completely analogous to the case of the time-discrete AWGN-BC $K_0(N_1 W, N_2 W, E)$. We call this channel the band-limited Gaussian noise (BLGN) BC and denote it by $K_0(N_1 W, N_2 W, E)$. We are interested in its capacity region, $\mathcal{C}(K_{12}, \text{BLGN})$.

The individual capacities are now given by [67]

$$C_i^* = W \log \left(1 + \frac{E}{N_i W}\right), \quad i = 1, 2. \quad (160)$$

Naive superposition yields the simultaneously achievable rates

$$\bar{C}_i \triangleq W \log \left(1 + \frac{E_i}{E_j + N_i W}\right) \quad (161)$$

for $i = 1, 2; j = 3 - i; E_1 + E_2 = E$.

A third approach which is available in this case is standard frequency division or band allocation. Here, the simultaneous transmitter operates with power $E_i$ in a band of width $W_i$ when sending to receiver $i$ $(i = 1, 2)$, whereby $W_1 + W_2 = W$ and $E_1 + E_2 = E$. For each fixed frequency division $W_i, W_2$, the transmitter can send at simultaneous rates

$$\bar{C}_i(W_i, E_i) \triangleq W_i \log \left(1 + \frac{E_i}{N_i W_i}\right). \quad (162)$$

$i = 1, 2$. As $E_1$ varies, $0 \leq E_1 \leq E$, one obtains a curve $(\bar{C}_1(W_1, E_1), \bar{C}_2(W_2, E_2))$. Bergmans and Cover [72] have shown that the envelope of all these curves as $W_1$ varies, $0 \leq W_1 \leq W$, strictly dominates the naive time-sharing bound derived from the capacities in (160).

As a fourth approach, Bergmans and Cover [72] suggested a superposition scheme similar to the one developed for the AWGN-BC in which one receiver subtracts out of his received signal the signal intended for the other receiver. This procedure is called "cooperative broadcasting." The cooperation needed for this superposition scheme need not be active; the signals can be independently generated and "mixed in the air." Both in the naive superposition scheme and in the cooperative broadcasting scheme, the transmitter makes use of the entire band while
Theorem 48. For a band-limited Gaussian noise two-receiver broadcast channel with spectral densities \( N_1 \leq N_2 \), available bandwidth \( W \), and average input power \( E \), every rate pair \((R_1, R_2)\) given by

\[
0 \leq R_1 \leq C_1^*(E_1) = W \log \left( 1 + \frac{E_1}{WN_1} \right) \tag{163a}
\]

\[
0 \leq R_2 \leq C_2^*(E_1) = W \log \left( 1 + \frac{E - E_1}{WN_2 + E_1} \right) \tag{163b}
\]

is achievable for any \( E_1 \), \( 0 \leq E_1 \leq E \).

We denote the set of rates given by (163) by \( \mathcal{S}(K_{12},BLGN) \). The achievability of the pairs (163) follows from Theorem 44 due to Cover [42], whereas the optimality of the region \( \mathcal{S}(K_{12},BLGN) \) is a consequence of Theorem 46 which is due to Bergmans [44]. Thus we have [72] the following.

Theorem 49: The capacity region of the band-limited Gaussian noise broadcast channel \( K_{12}^p(N_1,N_2,W,E) \) is precisely equal to the set of rates given by (163). That is,

\[
\mathcal{O}(K_{12},BLGN) = \mathcal{O}(K_{12,3},BLGN). \]

Bergmans and Cover [72] also proved that the curve described by (163) dominates the envelope of frequency curves \((C_1,C_2)\) described by (162). The general result of [72] can be formulated as "cooperative broadcasting dominates frequency multiplexing which in turn dominates time multiplexing." The arguments contained in [72] can easily be extended to the situation of one transmitter and \( r \) receivers.

Hughes–Hartogs [73] has investigated the situation of \( m \) parallel Gaussian BC’s with \( r \) receivers. This structure combines aspects of both the BC of Cover [42] and the parallel channel of Shannon [74]. We shall describe the results obtained by Hughes–Hartogs only for the case \( m = 3, r = 2 \).

Let three AWGN-BC’s, denoted by \( K_1, K_2, K_3 \), respectively, with transition probability densities \( P(y_{ij},y_{2i} | x_1) \), \( P(y_{ij},y_{2j} | x_2) \) and \( P(y_{ij},y_{2k} | x_3) \) be given. The noise variances in the six component channels are given by the pairs \((N_{11,21},N_{21,32})\), \((N_{12,22},N_{22,33})\), and \((N_{13,23},N_{23,31})\). The total power available at the input is \( E \). When placed in parallel, these three BC's form a parallel AWGN-BC with two receivers. This parallel structure can be regarded as a single BC with input \( x = (x_1,x_2,x_3) \), and outputs \( y_1 = (y_{11},y_{12},y_{13}) \) and \( y_2 = (y_{21},y_{22},y_{23}) \). Considered this way, the parallel AWGN-BC is a BC with two receivers and noise spectra \( N_1 = (N_{11,21},N_{12,32}) \) and \( N_2 = (N_{21,31},N_{22,33}) \), respectively. In other words, the parallel AWGN-BC can also be viewed as a spectral Gaussian BC with two receivers and spectrum dimension equal to three. We denote this spectral Gaussian BC by \( K_{12}^{p,3}(N_1,N_2,E) \). The problem is to determine the region \( \mathcal{O}(K_{12}^{p,3},AWGN) \) of pairs \((R_1,R_2)\) which are simultaneously achievable for the spectral AWGN-BC \( K_{12}^{p} \). This capacity region is the two-dimensional analog of the single capacity associated with the parallel structure of a certain number of OWC's, in this case, three.

Hughes–Hartogs [73] has considered two situations, viz., a general unrestricted situation and a degraded or restricted one. In the general system, not all three BC’s are degraded in the same direction. Without loss of generality, we assume that \( N_{11} < N_{21} \), \( N_{12} < N_{22} \) but that \( N_{13} > N_{23} \). Thus, \( K_1 \) and \( K_2 \) are degraded in the same direction, whereas \( K_3 \) is degraded in the opposite one. Hughes–Hartogs [73] has demonstrated that the following rate region is achievable for this parallel structure. Let \( \mathcal{S}(K_{12}^{p},AWGN) \) be the set of all rates \((R_1,R_2)\) such that

\[
R_1 = \frac{1}{2} \log \left( 1 + \frac{E_{11}}{N_{11}} \right) \left( 1 + \frac{E_{12}}{N_{12}} \right) \left( 1 + \frac{E_{13}}{N_{13}} \right) \tag{164a}
\]

\[
R_2 = \frac{1}{2} \log \left( 1 + \frac{E_{21}}{E_{11} + N_{21}} \right) \left( 1 + \frac{E_{22}}{E_{12} + N_{22}} \right) \left( 1 + \frac{E_{23}}{N_{23}} \right) \tag{164b}
\]

\[
\sum_i (E_{1i} + E_{2i}) \leq E \text{ and } E_{1i}, E_{2i} \geq 0; \quad i = 1,2,3. \tag{164c}
\]

Hughes–Hartogs [73] has found the following.

Theorem 50: An achievable rate region for the parallel structure of three Gaussian broadcast channels with two receivers is given by the set of rates \((R_1,R_2)\) which satisfy (164). That is,

\[
\mathcal{S}(K_{12}^{p},AWGN) \subseteq \mathcal{E}(K_{12}^{p},AWGN). \]

This follows directly from Theorem 44 due to Cover [42], and from [2, theorem 7.5.1] due to Ebert [75]. Hughes–Hartogs [73] has shown that the boundary of the rate region \( \mathcal{S}(K_{12}^{p},AWGN) \) is obtained by a water-filling process similar to Shannon’s maximization process for the one-way parallel structure [74]. But, unlike Shannon’s solution for the OWC, the water-filling solution for the BC involves scalings of the powers and noises.

In the restricted situation the various component BC’s are all assumed to be degraded in the same direction. In our present example, this means that \( N_{1i} < N_{2i} \) for \( i = 1,2,3 \). It follows that the second receiver spectral channel \( P_2(y_2 | x) \) is a strictly degraded version of the first receiver spectral channel \( P_1(y_1 | x) \). Therefore, one may speak of a degraded parallel AWGN-BC or a degraded spectral Gaussian BC, denoted by \( (K_{12}^{p},D) \). Let \( \mathcal{E}(K_{12}^{p},D) \) be the set of rates \((R_1,R_2)\) such that

\[
R_1 = \frac{1}{2} \log \left( 1 + \frac{E_{11}}{N_{11}} \right) \left( 1 + \frac{E_{12}}{N_{12}} \right) \left( 1 + \frac{E_{13}}{N_{13}} \right) \tag{165a}
\]

\[
R_2 = \frac{1}{2} \log \left( 1 + \frac{E_{21}}{E_{11} + N_{21}} \right) \left( 1 + \frac{E_{22}}{E_{12} + N_{22}} \right) \left( 1 + \frac{E_{23}}{E_{13} + N_{23}} \right) \tag{165b}
\]
where again
\[
\sum_i (E_{1i} + E_{2i}) \leq E \text{ and } E_{1i}, E_{2i} \geq 0; \quad i = 1, 2, 3.
\] (165c)

As in the previous theorem, it is easily seen that every pair \((R_1, R_2)\) satisfying (165) is achievable for the degraded parallel AWGN-BC \((K_{12}^{D})\). Hughes-Hartogs [73] has established a converse, thereby showing that the rate region given by (165) is the true capacity region of the degraded parallel AWGN-BC \((K_{12}^{D})\). This capacity region is denoted by \(C(K_{12}^{D})\). The converse obtains only when one receiver is a strictly degraded version of the other. Thus, Hughes-Hartogs [73] has proved the following.

**Theorem 51**: The capacity region of the parallel degraded additive white Gaussian broadcast channel described above is precisely equal to the set of rates \((R_1, R_2)\) which satisfy (165). That is,
\[
C(K_{12}^{P}) = \mathcal{S}(K_{12}^{P}).
\]

Hughes-Hartogs [73] has extended the above results to the case of \(m \geq 3\) parallel AWGN-BC’s with \(r \geq 2\) receivers, both for the general situation and the degraded one. He has also given the extension of the parallel AWGN-BC to a band-limited time-continuous model, and indicated a solution for it.

XXVI. Hughes-Hartogs [73] has conjectured that the rate region \(\mathcal{S}(K_{12}^{P}, \text{AWGN})\) given by (164) is the true capacity region of the unrestricted parallel AWGN-BC \(K_{12}^{P}\). For an additional discussion of the same problem, see [76].

XXVII. Another problem suggested in [76], is to determine the capacity region of the parallel MAC.

We conclude this section with a remark on the use and application of BC’s. The degraded BC, the Gaussian BC, and also BC’s for which one component is less noisy than the other one (cf. Definition 2, Section VI), all possess the characteristic property that \((R_1, R_2, R_0) \in \mathcal{C}(K_{12}^{III})\) if and only if \((R_1, R_0 + R_2) \in \mathcal{C}(K_{12}^{II})\) and in turn if and only if \((R_1, R_0 + R_2) \in \mathcal{C}(K_{12}^{II})\). This implies that every message which is received by the high noise receiver 2 is also received by the low noise receiver 1; but, in addition, receiver 1 receives a separate message. Thus, the transmitter can send information at rate \(R_2 \leq R_0 + R_2\) to receiver 2, and at rate \(R_1 \leq R_1 + R_0 = R_1 + R_2\) to receiver 1.

Cover [42] has envisioned such use of the DBC and the Gaussian BC where information is sent at different levels to the two receivers. We call this procedure “compound broadcasting,” as it combines the two procedures of simultaneous communication [37] and superposition coding [42] into one single procedure. The component channels must of course be compatible in some sense so that compound broadcasting has an effect. In practice, compound broadcasting may be the most appropriate procedure to follow. Several examples are given in [42] and [72]. We provide here yet another example of the use of the same idea.

Imagine a newspaper editor who wants to convey essential information to all his readers (assume for this category that the maximum noise power is \(N_2\)) and more detailed information to a smaller group of readers with more time and interest (assume for this latter group that the noise power is \(N_1 < N_2\)). The editor can now adopt the following strategy. He may condense the most important news into the headlines at rate \(R_2\). This news will be caught by all readers. On the other hand, he may supply the more detailed news in smaller print in the articles at rate \(R_1\). This kind of information will be obtained only by the low noise receivers, i.e., those readers with more time than is required to skim through the headlines. The latter category of readers will receive the total news at rate \(R_1 + R_2\). This strategy of distributing the information between the two groups may very well be optimal. What Cover [42] wrote in his original article applies to this example: “for a slight degradation in the rate for the worst channel an incrementally larger increase in the rate of transmission can be made for the better channels.”

IX. THE RELAY CHANNEL

The relay channel was studied by the author for the case of three terminals in [77], and for more than three terminals in [78]. Consider a discrete communication channel with three different terminals and transition probabilities given by \(P(y_1, y_2, y_3|x_1, x_2, x_3)\). Here \(x_i\) is the input at terminal \(t\) and \(y_i\) the corresponding output, \(t = 1, 2, 3\). As in the case of the TWC, all terminals are mixed in the sense that each one contains both an input and an output. For the other channels studied in this survey paper (the IFC, MAC, and BC), all terminals are pure, i.e., each one is either a strict input or a strict output terminal.

In [77], the author considered the following problem. Suppose the sender at terminal 1 wants to transmit information from a single source to the receiver at terminal 3, whereas terminal 2 acts as a relay (Fig. 19). The transmitter at terminal 2 observes the output letter \(y_{2(k)}\) received at time \(h\) at his terminal before he selects the next input letter \(x_{2(k+1)}\) for transmission over the channel. Thus, terminal 2 can influence the transmission procedure from terminal 1 to terminal 3. The problem is how to transmit information as effectively as possible in the specific direction 1 \(\rightarrow\) 3 over the relay channel, assuming that all terminals, in particular terminal 2, cooperate in the transmission procedure. The input \(x_{2}\) and output \(y_{1}\) are not so crucial in this context, and are therefore omitted in the representation of Fig. 19. However, they may play a role when one considers a relay channel with feedback, or simultaneous communication in various directions over the channel.

We denote the present three-terminal relay channel, with one-way transmission from terminal 1 to terminal 3, by \(K^{R}(1 \rightarrow 3; 2)\). A four-terminal relay channel where two terminals communicate with each other in opposite directions with the aid of two relay terminals could be denoted by \(K^{R}(1 \leftrightarrow 3; 2, 4)\). As a practical example, consider a radio station that wants to transmit to a receiver located...
at the other side of a mountain top, or a transmitter on earth that wants to send information to the dark side of the moon. In both cases, a relay station or satellite could be of great help.

The author, in [77], has investigated various ways of transmitting information over $K^R(1 \rightarrow 3; 2)$. It turns out that there are several unexpected methods which can be used and which are optimal in some specific cases. The author [77] has found four different transmission methods which can be roughly described as follows. Not only can one send directly from terminal 1 to terminal 3, but one may also time-share between the two procedures of sending first from terminal 1 to terminal 2 and then from terminal 2 to terminal 3. However, sometimes it is better to send first from terminal 1 to the pair of terminals (2,3) and then, within the pair, from terminal 2 to terminal 3. Or, finally, one may prefer to send first from terminal 1 to terminal 2 and then jointly from the pair (1,2) of terminals to terminal 3. Like the second procedure, the last two involve time-sharing between the two underlying procedures. At the time of the writing of [77], the author was not aware yet of the results on the BC [42]; it is quite conceivable that the results in [77] can be greatly improved by allowing a broadcast formulation and superposition coding.

A code $(n,M,F_2,F_3,X)$ for $K^R(1 \rightarrow 3; 2)$ consists of $M$ codewords $u_1, \ldots, u_M$ at terminal 1; $M$ decoding sets $D_1, \ldots, D_M$ at terminal 3; a strategy $F_2$ of length $n$ at terminal 2 which specifies at each transmission period $k$ the input letter $x_{sk}$ on the basis of the previously transmitted and received sequences at that terminal, and a similar strategy at terminal 3, such that

$$P^n(D_1|u_i,F_2,F_3) \geq 1 - \lambda$$ (166)

for $i = 1, \ldots, M$. A number $R' \geq 0$ is an attainable rate for $K^R(1 \rightarrow 3; 2)$ if, for any $\epsilon > 0$ and any $0 < \lambda < 1$, for $n$ sufficiently large, there exists a code $(n,M,F_2,F_3,X)$ such that $\log M \geq n(R' - \epsilon)$. The transmission capacity of the system $K^R(1 \rightarrow 3; 2)$ is denoted by $C^R(1,3)$.

In [77], the author found a limiting expression for $C^R(1,3)$, and obtained a necessary and sufficient condition for $C^R(1,3)$ to be positive.

XXVIII. It is still an open problem to find a simple computable characterization for $C^R(1,3)$.

Given the transition probabilities $P(y_1,y_2,y_3|x_1,x_2,x_3)$, one can derive various subchannels which could be used to send information away from terminal 1 and into terminal 3. We distinguish between five such subchannels. For each $(x_2,x_3)$, $P(y_3|x_1,x_2,x_3)$ is a OWC from terminal 1 to terminal 3, and we denote its capacity, maximized over $(x_2,x_3)$, by $C_1(1,3)$. Similarly, for each $(x_2,x_3)$, $P(y_2|x_2,x_3)$ is a OWC from terminal 1 to terminal 2 with maximum capacity $C_1(1,2)$. Likewise, for each $(x_1,x_3)$, $P(y_3|x_2,x_1)$ is a OWC from terminal 2 to terminal 3 with highest capacity $C_1(2,3)$. Next, for each pair $(x_2,x_3)$ we define a subchannel from terminal 1 to the pair $(2,3)$ of terminals by $P(y_2,y_3|x_1,x_3)$. The highest capacity in that direction is denoted by $C_1(1,[2,3])$. Finally, for each $x_3$, $P(y_3|x_1,x_2)$ defines a channel from the pair $(1,2)$ of terminals to terminal 3; the largest capacity, over $x_3$, is denoted by $C_1([1,2),3]$. In [77], the author derived the following relationships among these five capacities.

Theorem 52: In a discrete memoryless three-terminal relay channel with transmission of information from terminal 1 to terminal 3 where terminal 2 acts as a relay, the capacities of the one-way subchannels that can be derived from the relay channel in the manner described above satisfy the following inequalities and relationships.

\[
\begin{align*}
&\text{a)} \quad C_1(1,3) \leq C_1(1,[2,3]). \\
&\text{b)} \quad C_1(1,2) \leq C_1([1,2),3]. \\
&\text{c)} \quad \text{If } C_1(1,[2,3]) = 0, \text{ then } C_1(1,3) = C_1(1,2) = 0, \text{ but the converse is not true.} \\
&\text{d)} \quad C_1(1,3) \leq C_1([1,2),3]. \\
&\text{e)} \quad C_1(2,3) < C_1([1,2),3]. \\
&\text{f)} \quad C_1([1,2),3] = 0 \text{ if and only if } C_1(1,3) = C_1(2,3) = 0.
\end{align*}
\]

It appears from part c) that $C_1(1,[2,3])$ can be positive while $C_1(1,3)$ and $C_1(2,3)$ are both equal to zero. A numerical example illustrating this fact is given in [77]. The following example describes a real-life situation where this phenomenon could occur.

Suppose an official in Washington wants to send a joint message to London and Paris, but the separate component channels to these cities are so noisy that the receivers cannot decode the message individually. In this situation the person in Washington may still be able to get a joint message across because if the two receivers can combine the two received sequences, they may be able to decode the joint message correctly. A spy who intercepts the signal in one of the two communication links will not get any information at all. Only a person who sees the outputs of both channels can decode the message originally sent.

Based on the four methods of transmitting information described above, the author [77] has found several achievable rates for $K^R(1 \rightarrow 3; 2)$. He proved the following.

Theorem 53: For a discrete memoryless three-terminal relay channel, described as above, the following rates $R'$
are achievable:

a) \(0 \leq R' \leq C_{1(1,3)},\)

b) \(0 \leq R' \leq \frac{C_{1(1,2)}C_{2(3)} + C_{1(1,2)}C_{2(3)}}{C_{1(1,2)} + C_{1(1,2)}}\)

provided \(C_{1(1,2)} > 0\) and \(C_{1(2,3)} > 0,\)

c) \(0 \leq R' \leq \frac{C_{1(1,2)}C_{1(1,2)} + \log b}{C_{1(2,3)} + \log b}\)

d) \(0 \leq R' \leq \frac{C_{1(1,2)}C_{1(1,2)} + \log a}{C_{1(1,2)} + \log a}\)

Here, \(a \geq 2\) and \(b \geq 2\) denote the size of the input and output alphabets, respectively, at terminal 2, whereas \(C_{1(1,3)}, C_{1(1,2)}, C_{1(2,3)}, C_{1(1,2,3)},\) and \(C_{1(1,2,3)}\) denote the capacities of the various one-way subchannels that can be derived from the relay channel.

The author [77] has proved the following.

**Theorem 54:** The transmission capacity \(C^R(1,3)\) of the relay channel described above is positive if and only if the one-way capacities \(C_{1(1,2,3)}\) and \(C_{1(1,2,3)}\) are both positive. That is,

\(C^R(1,3) > 0,\) if and only if \(C_{1(1,2,3)} > 0\) and \(C_{1(1,2,3)} > 0.\)

This theorem states that the transmission capacity of the three-terminal relay channel \(K^R(1 \rightarrow 3;2)\) is positive if and only if the one-way capacity \(C_{1(I,J)}\) is positive for every partition \((I,J)\) of the set of terminals \([1,2,3]\) for which the transmitting terminal 1 is in \(I\) and the receiving terminal 3 is in \(J.\) In the present case there are two such partitions, viz. \((1,2,3))\) and \((1,2,3).\) If \(C_{1(I,J)} > 0,\) the group \(I\) of sending terminals can as a whole transmit information to the group \(J\) of receiving terminals. If \(C_{1(I,J)} > 0\) for all partitions just described, the final receiver 3 can decode correctly the message originally sent by sender 1 by subtracting out of his received signal the sequences which are sent and received by relay terminal 2 and passed on to him. Theorem 54 also holds in the case of a \(T\)-terminal \((T > 3)\) relay channel, as was shown in [78].

Many problems still remain in the study of the relay channel, one of which we stated above.

**XXIX.** As was mentioned earlier, one could conceivably obtain better rates than the four exhibited in Theorem 53 by allowing a broadcast formulation and employing random coding arguments. These possibilities have not yet been explored.

**XXX.** Also, relay channels with feedback and the problem of simultaneous communication in various directions over the relay channel have yet to be investigated.

**XXXI.** Very little work has been done on relay channels with more than three terminals, apart from the investigations in [78].

In connection with the example given above, in which the relay channel is used to send secret information to the two output users, we should like to mention some other papers in information theory which deal with secrecy considerations arising by the communication over a noisy channel. Recently, Wyner [79] introduced the so-called wiretap channel which in a sense is related to the DBC. Here the objective is to maximize the rate of reliable communication from the legitimate receiver, subject to the constraint that the wiretapper learns as little as possible about the source output. Wyner's results for the d.m. wiretap channel were extended to the case of the Gaussian wiretap channel by Leung-Yan-Cheong and Hellman [80]. In [81], Leung-Yan-Cheong has investigated the wiretap channel with feedback.

The wiretap problem is related to problems in cryptography [82], [83]. However, while the objectives are the same from both points of view, the techniques used to achieve privacy are very different. For an application of cryptography to multi-user problems, see the recent article by Diffie and Hellman [84]. Although cryptography is also a very active area at present, a detailed discussion of this fascinating field would lead beyond the scope of this article.

**X. GENERAL MULTI-USER NETWORKS**

Consider a three-terminal communication channel \(P(x_1,y_2,y_3|x_1,x_2,x_3),\) as discussed in Section IX, with inputs and outputs at each terminal. Suppose that at each terminal the transmitter wants to broadcast separate information to the receivers at the other two terminals, whereas for each transmission direction the remaining terminal acts as a relay terminal.

**XXXII.** Consider the problem of sending simultaneously in the six directions over the three-terminal channel, \(1 \rightarrow 2, 1 \rightarrow 3, 2 \rightarrow 1, 2 \rightarrow 3, 3 \rightarrow 1,\) and \(3 \rightarrow 2.\) This problem involves aspects of the TWC, MAC, IFC, BC, and the relay channel \(K^R(1 \rightarrow 3;2)\) discussed in the previous section. Unfortunately, the general solution of this problem must wait until the capacity regions of the TWC and BC are settled and the capacity of the relay channel is determined.

**XXXIII.** However, a preliminary study of this problem could shed much light on additional problems involved.

The extension of the previous setup to a \(T\)-terminal communication channel, \(T > 3,\) is of course formidable.

**XXXIV.** Here the problem is to determine the \(T(T - 1)\)-dimensional capacity region, assuming that every terminal sends only separate information to every other terminal.

In the most general setup, discussed in [7], each terminal has \(2^{T-1} - 1\) different information sources, with each source sending information to precisely one of the \(2^{T-1} - 1\) nonempty subsets of the set of \(T - 1\) remaining terminals.

**XXXV.** In this case the problem would be to determine the \(T(2^{T-1} - 1)\)-dimensional capacity region.
XXXVI. A special case would be to determine the capacity region of a T-terminal Gaussian communication channel under either of the two setups just described.

For a discussion of the general multi-user framework and the Gaussian case in particular, the reader is referred to [20].

XXXVII. Another aspect of large communication networks which needs to be studied is their possible separation into smaller networks. At first sight, one would expect here to obtain results similar to Theorem 54.

XXXVIII. Yet another fundamental problem in the theory of multi-way channels is the source-channel in terconnection problem, mentioned in [19]. Whereas this problem was solved by Shannon [68] in the case of the OWC, it remains open for even the simplest multi-way channels. For a formulation in the case of the BC, see [20] and [42].

XXXIX. Carleial [7] has discussed the problems which arise in communication over a network with dependent information sources.

Of immediate relevance here is the basic paper by Slepian and Wolf [86] on the coding of correlated information sources, the results of which were extended to the case of ergodic sources by Cover [87]. The concept of common information, (see [92], [93], [21], and [94]), may also play an important role. According to [7], if common information generated at several correlated sources can be encoded and transmitted from any one of them, one may have more flexibility than with independent sources. In particular, if one could avoid duplication in encoding, an advantageous use of capacity would result.

Finally, in connection with the general problem of sending information over a network, we should mention the work done by Ovseevich and Pinsker on multi-path channels which is discussed in the survey article by Dobrushin [88], and the article by Elias [89] on networks of Gaussian channels. In a related paper, Wozencraft and Horstein [99] have investigated how two-way channels can be used to provide noiseless feedback. More generally, for the study of coding problems for multi-way channels with feedback, the article by Schalkwijk and Kailath [100] may prove to be useful.

XI. Concluding Remarks

It is interesting to observe that all multi-way channels considered in this survey article involve the notion of conditional decoding. Here the receiver subtracts out of his received signal another signal in order to determine the message sent. This aspect is entirely absent in the theory of the one-way channel.

The concept of superposition coding, developed in [42] and [45], can be regarded as an entirely novel idea that originated during the past decade and has become an important standard technique in the theory of multi-way channels.

It can be fairly said that among the multi-way channels considered in this article, the multiple-access channel and the degraded broadcast channel are best understood. The theory for these two channels is rather complete now. Since Shannon’s article [1], the characterization of the capacity region of both the multiple-access channel and the degraded broadcast channel has stood centrally in the development of the theory of multi-way channels. The variety of refined results on the degraded broadcast channel shows that much more theory could be developed for more complicated channels.

The interference channel and the general broadcast channel, which have much in common, seem to be next in difficulty. One of the open problems which most urgently awaits solution is the characterization of the capacity region of these channels.

The two-way channel remains the most difficult and perhaps the least understood among the five multi-way channels considered in this paper. This difficulty can be partly ascribed to the extra complication caused by feedback. It is to be hoped that as progress is made on the interference channel and the general broadcast channel, further insight will be gained in the structure of the two-way channel, leading eventually to the solution of its capacity region.

As a final remark, a survey paper on multi-way channels would not be complete without mentioning that besides the theory of information transmission over multi-terminal channels, as discussed in this article, there is another related area, namely that of multiple-user source coding. The latter area began with the fundamental paper by Slepian and Wolf [86]. Since then, many important contributions have been made.

In this paper we have addressed ourselves only to the channel coding aspects of multiple-user information theory and not at all to data compression. However, the interrelationship between the two areas seems to be quite strong, even more so than in the case of the one-way channel.

ACKNOWLEDGMENT

The author wishes to thank several friends and colleagues who let him share their current research and provided him with helpful comments on drafts of this paper.

REFERENCES


R. Ahlswede, private communication.


S. I. Gelfand, "Broadcast Channels." Lecture given at the 1st IEEE-USSR Workshop on Information Theory, December 1976, Moscow, USSR.


