Multiple User Information Theory

ABBAS EL GAMAL, MEMBER, IEEE, AND THOMAS M. COVER, FELLOW, IEEE

Invited Paper

Abstract—A unified framework is given for multiple user information networks. These networks consist of several users communicating to one another in the presence of arbitrary interference and noise. The presence of many senders necessitates a tradeoff in the achievable information transmission rates. The goal is the characterization of the capacity region consisting of all achievable rates. The focus is on broadcast, multiple access, relay, and other channels for which the recent theory is relatively well developed. A discussion of the Gaussian version of these channels demonstrates the concreteness of the encoding and decoding necessary to achieve optimal information flow. We also offer speculations about the form of a general theory of information flow in networks.

I. INTRODUCTION

THE SHANNON theory of channel capacity has been extended successfully to many interesting communication networks in the past 10 years. We shall attempt to achieve three goals in our exposition of this theory: (i) make the theory accessible to researchers in communication theory, (ii) provide conditionally novel proofs of the theory for researchers in information theory, and (iii) present an overview of the basic problems in constructing a theory of information flow in networks.

The primary ideas can be obtained by reading the introduction and the sections on the Gaussian examples, the Shannon's theorem, and the summary. The heretofore unpublished information theoretic proofs are those for the sections on the multiple access channel, Slepian–Wolf data compression, and the degraded broadcast channel. All proofs, both new and old, are based on the idea of jointly typical sequences.

No claim for comprehensive coverage is given. For that the reader is referred to van der Meulen [1]. Rather, we are concerned with providing a unified approach to the theory. This leads naturally to a discussion of some of the major results. We begin by discussing some of the building blocks for networks.

Suppose m ground stations are simultaneously communicating to a common satellite as in Fig. 1. This is known as the multiple access channel. What are the achievable rates of communication? Does the total amount of information flow tend to infinity with the number of stations—or does the interference put an upper limit on the total communication? Does
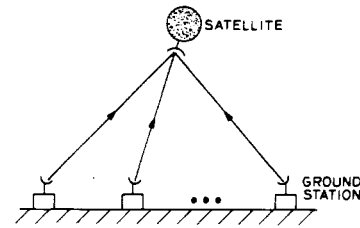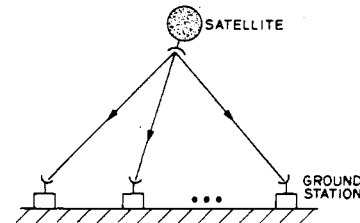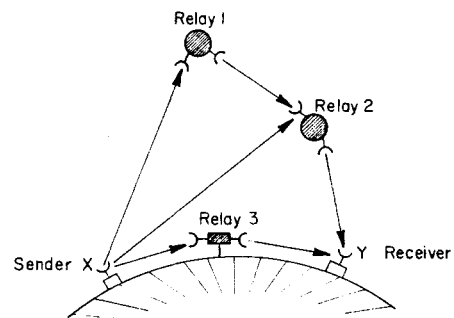
Fig. 1. Multiple access network.

Fig. 2. Broadcast network.

A COMMUNICATION NETWORK WITH RELAYS

Fig. 3. Relay network.

the signaling strategy change with m? Here the theory is completely known (Ahlswede [2] and Liao [3]), and all of these questions have quite satisfying answers (Section V).

In contrast, we can reverse the network, and consider one satellite broadcasting simultaneously to m stations as shown in Fig. 2. This is the broadcast channel. Here the achievable set of rates is not known except in special cases (Section VII). Yet another example consists of only one sender and one receiver, but includes extra channels serving as relays. This is the relay channel shown in Fig. 3.

In general, the underlying goal of work on these special channels is a theory for networks of the general form given in Fig. 4.

The interpretation of Fig. 4 is that at each instant of time the ith node sends a symbol x_i that depends on the messages
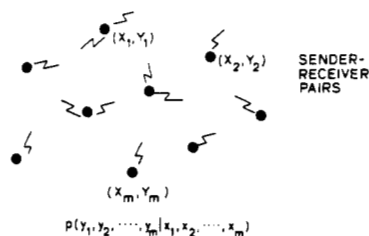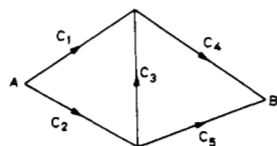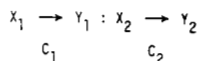
Fig. 4. General multiple user network.

Fig. 5. Capacity is min $\{C_1 + C_2, C_1 + C_3 + C_5, C_2 + C_4, C_4 + C_5\}$. The maximum flow minimum cut theorem.
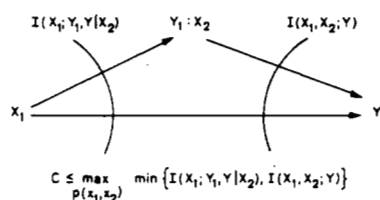

Fig. 6.

Fig. 7. Degraded and deterministic relay channel capacity.

he wishes to send and (perhaps) his past received y_i symbols. We assume that the result of the simultaneous transmission of $(x_1, x_2, \cdots, x_m)$ is a random collection of received symbols $(Y_1, Y_2, \cdots, Y_m)$ drawn according to a conditional probability $p(y_1, \cdots, y_m | x_1, \cdots, x_m)$, where $p(\cdot|\cdot)$ describes all of the effects of interference and noise in the network.

In a more restricted domain, such as the flow of water in networks of pipes, the existing theory is very satisfying. For example, in the single source, single sink network in Fig. 5, the maximum flow from A to B is easily computed from the maximum flow minimum cut theorem of Ford and Fulkerson [4].

Assume that the edges have capacities C_i as shown. Clearly, the maximum flow across any cutset is no greater than the sum of the capacities of the cut edges. Thus minimizing the maximum flow over cutsets yields an upper bound to the total flow from A to B. This flow can actually be achieved, as the Ford and Fulkerson theorem demonstrates.

However, the information flow problem involves "soft" quantities rather than "hard" commodities. A choice of node symbols x results in a random response Y, and it is difficult to see how to choose as many distinguishable x's as possible in this random environment. Consequently, it is gratifying to find that information problems like the relay channel and cascade channel admit min flow max cut interpretations. For example, the informally defined cascade network in Fig. 6, has capacity $C = \min \{C_1, C_2\}$, where C_i denotes the Shannon capacity of the ith channel. Also, for the degraded or deterministic relay channel (Section VII) we have a similar max flow min cut interpretation as shown in Fig. 7.

The structure of this paper is presented in miniature in Section II. In that section, we use Gaussian channels to run through the major results that will be given in greater generality and detail in the subsequent sections. The physically motivated Gaussian channel lends itself to concrete and easily interpreted answers. Some preliminary technical details on the properties of joint typicality are given in Section III, followed by a simple proof in Section IV of Shannon's original capacity theorem. Then treated are the multiple access channel (Section V), Slepian–Wolf data compression theorem (Section VI), the combination of both (Section VII), the broadcast channel (Section VIII), and the relay channel (Section IX).

The final summary (Section X) is a recapitulation of the paper paralleling Section II, this time in greater generality.

II. GAUSSIAN MULTIPLE USER CHANNELS

We shall begin our treatment of multiple user information theory by investigating Gaussian multiple user channels. This allows us to give concrete descriptions of the coding schemes and associated capacity regions. The proofs of capacity for the discrete memoryless counterparts of these channels will be given in later sections.

The basic discrete time additive white Gaussian noise channel with input power P and noise variance N is modeled by

$$Y_i = x_i + Z_i, \quad i = 1, 2, \cdots$$

where Z_i are independent identically distributed Gaussian random variables with mean zero and variance N. The signal $x = (x_1, x_2, \cdots, x_n)$ has a power constraint

$$1/n \sum_{i=1}^{n} x_i^2 \leq P.$$

The Shannon capacity C, obtained by maximizing $I(X; Y)$ over all random variables X such that $EX^2 \leq P$ is given by

$$C = (1/2) \log (1 + P/N) \text{ bits/transmission.} \quad (2.1)$$

The continuous time Gaussian channel capacity is simply related to the discrete time capacity. If the signal $x(t), 0 \leq t \leq T$, has power constraint P and bandwidth constraint W, and the white noise $Z(t), 0 \leq t \leq T$, has power spectral density N, then the capacity of the channel $Y(t) = x(t) + Z(t)$ is given by

$$C = W \log (1 + P/NW) \text{ bit/s.} \quad (2.2)$$

The relationship between (2.1) and (2.2) can be seen informally by replacing the continuous time processes by $n = 2TW$ independent samples from the process and calculating the noise variance per sample. The full theory establishing (2.2) can be found in Wyner [5], Gallager [6], and Slepian and Pollack [7].

Having said this, we restrict our treatment to time discrete Gaussian channels.

Random Codebook: Shannon observed in 1948 that a randomly selected codebook is good (with high probability) when the rate R of the codebook is less than the channel capacity $C = \max I(X; Y)$. As mentioned above, for the Gaussian channel the capacity is given by $C = (1/2) \log (1 + P/N)$ bits per transmission.

We now set up a codebook that will be used in all of the multiple user channel models below. The codewords comprising the codebook are vectors of length n and power P. To generate such a random codebook, simply choose 2^{nR} independent identically distributed random n-vectors $\{x(1), x(2),$

$\cdots, x(2^{nR})\}$, each consisting of n independent Gaussian random variables with mean zero and variance P. The rate R will be specified later. Sometimes we will need two or more independently generated codebooks.

In the continuous channel case, one simply lets the white noise generator of power P and bandwidth W run for T seconds. Every T seconds, a new codeword is generated and we list them until we fill up the codebook.

Now we analyze the Gaussian channels shown in Fig. 8.

A. The Gaussian Channel

Here $Y = x + Z$. Choose an $R < C = 1/2 \log (1 + P/N)$. Choose any index i in the set 2^{nR}. Send the ith vector $x(i)$ from the codebook generated above. The receiver observes $Y = x(i) + Z$, then finds the index $\hat{i}$ of the closest codeword to Y. If n is sufficiently large, the probability of error $P(i \neq \hat{i})$ will be arbitrarily small. As will be seen in the definitions on joint typicality, this minimum distance decoding scheme for the Gaussian channel is essentially equivalent to finding the codeword in the codebook that is jointly typical with the received vector Y.

B. The Multiple Access Channel with m Users

We consider m transmitters, each of power P. Let

$$Y = \sum_{1}^{m} x_i + Z.$$

Specializing the results of Section IV to the Gaussian channel shows that the achievable rate region for the Gaussian channel takes on the simple form given in the following equations:

$$R_i < C(P/N)$$

$$R_i + R_j < C(2P/N)$$

$$R_i + R_j + R_k < C(3P/N)$$

$$\vdots \qquad (2.3)$$

$$\sum_{1}^{m} R_i < C(mP/N)$$

where

$$C(x) = (1/2) \log (1 + x) \qquad (2.4)$$

denotes the capacity of the Gaussian channel with signal to noise ratio x. When all the rates are the same, the last inequality dominates the others.

Here we need m codebooks, the ith codebook having 2^{nR_i} codewords of power P. Transmission is simple. Each of the independent transmitters chooses whatever codeword he wishes from his own codebook. The users simultaneously send these vectors. The receiver sees the codewords added together with the Gaussian noise Z.

Optimal decoding consists of looking for the m codewords, one from each codebook, such that the vector sum is closest to Y in Euclidean distance. The set of m codewords achieving the minimum distance to Y corresponds to the hypothesized collection of messages sent.

If $(R_1, R_2, \cdots, R_m)$ is in the capacity region given above, then the probability of error goes to zero as n tends to infinity.

Remarks: It is exciting to see in this problem that the sum
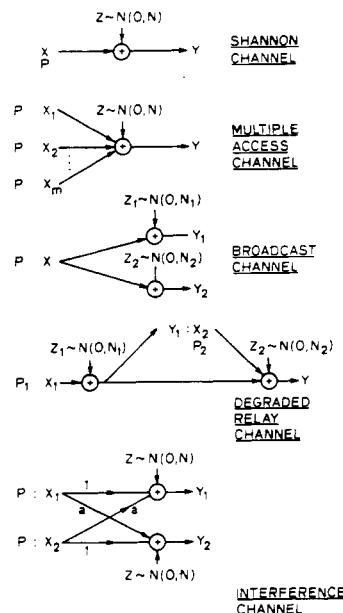

Fig. 8. Gaussian multiple user channels.

of the rates $C(mP/N)$ of the users goes to infinity with m. Thus in a cocktail party with m celebrants of power P in the presence of ambient noise N, although the interference grows as the number of speakers increases, the intended listener receives an unbounded amount of information as the number of people goes to infinity. A similar conclusion holds of course for ground communications to a satellite.

It is also interesting to note that the optimal transmission scheme here does not involve time division multiplexing. In fact, each of the transmitters utilizes the entire time to send his message.

A practical consideration for ground transmission to a satellite involves the possible inability of the ground communicators to synchronize their transmissions. Nonetheless, it can be shown that the capacity is unchanged when there is a lack of synchronization [8].

C. The Broadcast Channel

Here we assume that we have a sender of power P and two distant receivers, one with noise spectral density N_1 and the other with noise spectral density N_2. Without loss of generality, assume $N_1 < N_2$. Thus, in some sense receiver, Y_1 is better than receiver Y_2. The model for the channel is $Y_1 = x + Z_1$ and $Y_2 = x + Z_2$, where Z_1 and Z_2 are arbitrarily correlated Gaussian random variables with variances N_1 and N_2, respectively. The sender wishes to send independent messages at rates R_1 and R_2 to receivers Y_1 and Y_2, respectively.

Fortunately, all Gaussian broadcast channels belong to the class of degraded broadcast channels which will be discussed in Section VIII. Specializing that work, we find the following capacity region for the Gaussian broadcast channel:

$$R_1 < C(\alpha P/N_1)$$

$$R_2 < C(\bar{\alpha} P/(\alpha P + N_2)) \qquad (2.5)$$

where $0 \leqslant \alpha \leqslant 1$, $\bar{\alpha} = 1 - \alpha$, may be arbitrarily chosen to trade off rate R_1 for rate R_2 as the transmitter wishes.

To encode the messages, the transmitter generates two codebooks, one with power αP at rate R_1, and another codebook with power $\bar{\alpha} P$ and rate R_2. He has chosen R_1 and R_2 to sat-

isfy the equation above. Then, to send an index $i \in \{1, 2, \cdots, 2^{nR_1}\}$ and $j \in \{1, 2, \cdots, 2^{nR_2}\}$ to Y_1 and Y_2, respectively, he takes codeword $x(i)$ from the first codebook and codeword $x(j)$ from the second codebook and computes the sum. He then sends the sum over the channel.

Two receivers must now do the decoding. First consider the bad receiver Y_2. He merely looks through the second codebook for the closest codeword to his received vector Y_2. His effective signal-to-noise ratio is $\overline{\alpha}P/(\alpha P + N_2)$ since Y_1's message acts as noise to Y_2. The good receiver Y_1 first decodes Y_2's codeword, which he can do because of his lower noise N_1. He subtracts this codeword $\hat{x}_2$ from Y_1. This leaves him with a channel of power αP and noise N_1. He then looks for the closest codeword in the first codebook to $Y_1 - \hat{x}_2$. The resulting probability of error can be made as low as wished.

A nice dividend of optimal encoding for degraded broadcast channels is that the better receiver Y_1 always knows the message intended for receiver Y_2 in addition to the extra information intended for himself.

D. The Relay Channel

For the relay channel, we have a sender X_1 and an ultimate intended receiver Y. Also present, however, is the relay channel intended solely to help the sender. The channel is given by

$$Y_1 = x_1 + Z_1$$
$$Y_2 = x_1 + Z_1 + x_2 + Z_2 \qquad (2.6)$$

where Z_1, Z_2 are independent zero mean Gaussian random variables with variance N_1, N_2, respectively. The allowed encoding by the relay is the causal sequence

$$x_{2i} = f_i(y_{11}, y_{12}, \cdots, y_{1i-1}). \qquad (2.7)$$

The sender X_1 has power P_1 and the relay X_2 has power P_2. The results of Section IX yield the capacity

$$C = \max_{0 < \alpha < 1} \min \left\{ C\left(\frac{P_1 + P_2 + 2\sqrt{\overline{\alpha}P_1P_2}}{N_1 + N_2}\right), C\left(\frac{\alpha P_1}{N_1}\right)\right\}$$

where

$$\overline{\alpha} = 1 - \alpha.$$

Note that if

$$P_2/N_2 \geqslant P_1/N_1$$

it can be seen that $C^* = C(P_1/N_1)$. (This is achieved by $\alpha = 1$.) The channel appears to be noise-free after the relay, and the capacity $C(P_1/N_1)$ from x_1 to the relay can be achieved. Thus the rate without the relay $C(P_1/(N_1 + N_2))$ is increased by the relay to $C(P_1/N_1)$. For large N_2, and for $P_2/N_2 \geqslant P_1/N_1$, we see that the increment in rate is from $C(P_1/(N_1 + N_2)) \approx 0$ to $C(P_1/N_1)$.

Encoding of Information: Two codebooks are needed. The first codebook has 2^{nR_1} words of power αP_1. The second has 2^{nR_0} codewords of power $\overline{\alpha}P_1$. We shall use words from these codebooks successively in order to create the opportunity for cooperation by the relay. We start by sending a codeword from the first codebook. The relay now knows the index of this codeword since $R_1 < C(\alpha P_1/N_1)$, but the intended receiver does not. However, the intended receiver has a list of possible codewords of size $2^{n(R_1 - C(\alpha P_1/N_1 + N_2))}$. The list calculation involves a result on list codes.

In the next block the relay and the transmitter would like to cooperate to resolve the receiver's uncertainty about the previously sent codeword on the receiver's list. Unfortunately, they cannot quite be sure what this list is. They do not know the received signal Y. Thus they randomly partition the first codebook into 2^{nR_0} cells with an equal number of codewords in each cell. The relay, the receiver, and the transmitter agree on this partition. The relay and the transmitter find the cell of the partition in which the codeword from the first codebook lies and cooperatively send the codeword from the second codebook with that index. That is, both X_1 and X_2 send the same designated codeword. The relay, of course, must scale this codeword so that it meets his power constraint P_2. They now simultaneously transmit their codewords. An important point to note is that the cooperative information sent by the relay and the transmitter is sent coherently. So the power of the sum as seen by the receiver Y is $(\sqrt{\overline{\alpha}P_1} + \sqrt{P_2})^2$.

However, this does not exhaust what the transmitter does in the second block. He also chooses a fresh codeword from his first codebook, adds it "on paper" to the cooperative codeword from his second codebook, and sends this sum over the channel.

The reception by the ultimate receiver Y in the second block involves first finding the cooperative index from the second codebook by looking for the closest codeword in the second codebook. He subtracts it off, and then calculates a list of indices of size 2^{nR_0} corresponding to all transmitted words from the first book which might have been sent in that second block.

Now it is time for the intended receiver Y to finish computing the codeword from the first codebook sent in the first block. He takes his list of possible codewords that might have been sent in the first block and intersects it with the cell of the partition that he has learned from the cooperative relay transmission in the second block. The rates and powers have been chosen so that it is highly probable that there will be only one codeword in this intersection. This is Y's guess about the information sent in the first block.

We are now in steady state. In each new block, the transmitter and the relay cooperate to resolve the list uncertainty from the previous block. In addition, the transmitter superimposes some fresh information from his first codebook to his transmission from the second codebook and transmits the sum.

The receiver is always one block behind, but for sufficiently many blocks, this does not affect his overall rate of reception.

E. The Interference Channel

The interference channel has two senders and two receivers. Sender 1 wishes to send information to receiver 1. He does not care what receiver 2 receives or understands. Similarly, with sender 2 and receiver 2. Each channel interferes with the other. It is not quite a broadcast channel because there is only one intended receiver for each sender, nor is it a multiple access channel because each receiver is only interested in what is being sent by the corresponding transmitter.

This channel has not been solved in general, even in the Gaussian case. But remarkably, in the case of high interference, Carleial [9] has shown that the solution to this channel is the same as if there were no interference whatsoever. To achieve this, generate two codebooks, each with power P and rate $C(P/N)$. Each sender independently chooses a word from his book and sends it. Now, if the interference a satisfies $C(a^2P/(P + N)) > C(P/N)$, the first receiver perfectly understands the index of the second transmitter. He finds it by the

usual technique of looking for the closest codeword to his received signal. Once he finds this signal, he subtracts it from his received waveform. Now there is a clean channel between him and his sender. He then searches his sender's codebook to find the closest codeword and declares that codeword to be the one sent.

III. THE ASYMPTOTIC EQUIPARTITION THEOREM AND SHANNON CHANNEL CAPACITY

Every sequence of n fair coin flips has equal probability $(1/2)^n$. If the coin has bias p, all of the sequences having roughly np heads are nearly equally probable and exhaust almost all of the probability. A formalization of this idea for arbitrary random variables is known as the asymptotic equipartition property (AEP) (Shannon [10], MacMillan [11], and Breiman [12]).

Consider a sequence of independent identically distributed random variables $X = (X_1, X_2, \cdots, X_n)$, where X_i is drawn according to probability mass function $p(x)$. We are interested in defining a set A of possible outcomes, each roughly equally probable, such that $P(X \in A) \approx 1$. Toward this end, we shall say that a given sequence $x = (x_1, x_2, \cdots, x_n)$ is ϵ-typical if

$$|-(1/n) \log p(X) - H(X)| \leqslant \epsilon \qquad (3.1)$$

where

$$H(X) = - \sum p(x) \log p(x) \qquad (3.2)$$

is the Shannon *entropy* of $p(\cdot)$. Let us define the typical set A_ϵ to be the set of all ϵ-typical n-sequences x.

By the law of large numbers, for a random i.i.d. sequence X

$$-(1/n) \log p(X) = (1/n) \sum_{i=1}^{n} -\log p(X_i)$$

$$\rightarrow H(X), \text{ with probability one.} \qquad (3.3)$$

Consequently, the following results are true.
1) $P(A_\epsilon) \rightarrow 1$, as $n \rightarrow \infty$.
2) $x \in A_\epsilon$ implies $2^{-n(H+\epsilon)} \leqslant p(x) \leqslant 2^{-n(H-\epsilon)}$.
3) The number of elements $\|A_\epsilon\|$ in A_ϵ is $\leqslant 2^{n(H+\epsilon)}$.
The third assertion follows from

$$1 \geqslant \sum_{A_\epsilon} p(x) \geqslant \sum_{A_\epsilon} 2^{-n(H+\epsilon)} = \|A_\epsilon\| \, 2^{-n(H+\epsilon)}. \qquad (3.4)$$

Roughly speaking, a sequence is typical if the proportion of occurrences of each of its symbols is close to the true probability of occurrence.

An immediate application of the AEP yields Shannon's source coding theorem:

Theorem 1 (Shannon 1948): Given an independent identically distributed source $X_1, X_2, \cdots$ with entropy $H(X)$, there exists, for every $\epsilon > 0$, an integer n and an encoding i: $\mathfrak{X}^n \rightarrow \{1, 2, \cdots, 2^{n(H+\epsilon)}\}$ and decoding rule g: $\{1, 2, \cdots, 2^{n(H+\epsilon)}\} \rightarrow \mathfrak{X}^n$, such that

$$P\{g(i(X)) \neq X\} \leqslant \epsilon.$$

Proof: Simply index the elements x in the typical set A_ϵ.
Remark: Since the index set has $\leqslant 2^{n(H+\epsilon)}$ elements, we see that only $(H + \epsilon)$ bits/symbol are necessary to describe x.

Now, to make progress with multiple user information theory, the idea of *joint typicality* is needed. A pair of sequences x and y are said to be jointly ϵ-typical if x is individu-
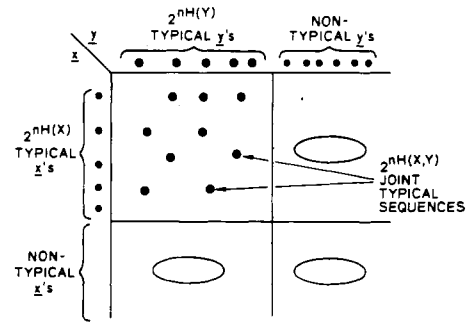

Fig. 9. Jointly typical sequences.

ally ϵ-typical, i.e.,

$$|-(1/n) \log p(x) - H(X)| \leqslant \epsilon$$

y is individually ϵ-typical, and (x, y) is ϵ-typical, i.e.,

$$|-(1/n) \log p(x, y) - H(X, Y)| \leqslant \epsilon.$$

The picture of jointly typical pairs is given in Fig. 9.

The dots in the matrix denote jointly typical pairs. It can be shown by the method of (3.4) that there are $\leqslant 2^{nH(Y|X)}$ dots in each row and $\leqslant 2^{nH(X|Y)}$ dots in each column.

The definitions and proofs of the needed results now follow. Let $\{X^{(1)}, X^{(2)}, \cdots, X^{(k)}\}$ denote a finite collection of discrete random variables with some fixed joint distribution $p(x^{(1)}, x^{(2)}, \cdots, x^{(k)})$, $(x^{(1)}, x^{(2)}, \cdots, x^{(k)}) \in \mathfrak{X}^{(1)} \times \mathfrak{X}^{(2)} \times \cdots \times \mathfrak{X}^{(k)}$. Let S denote an ordered subset of these r.v.'s, and consider n independent copies of S. Thus

$$\Pr\{S = s\} = \prod_{i=1}^{n} \Pr\{S_i = s_i\}, \quad s \in S^n. \qquad (3.5)$$

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

$$\Pr\{S = s\} = \Pr\{(X^{(j)}, X^{(l)}) = (x^{(j)}, x^{(l)})\}$$

$$= \prod_{i=1}^{n} p(x_i^{(j)}, x_i^{(l)}). \qquad (3.6)$$

By the law of large numbers, for any subset S of random variables,

$$-(1/n) \log p(S_1, S_2, \cdots, S_n) = -(1/n) \sum_{i=1}^{n} \log p(S_i) \rightarrow H(S).$$

$$(3.7)$$

Convergence in (3.7) take place simultaneously with probability one for all 2^k subsets

$$S \subseteq \{X^{(1)}, X^{(2)}, \cdots, X^{(k)}\}.$$

Definition: The set A_ϵ of ϵ-typical n-sequences $(x^{(1)}, x^{(2)}, \cdots, x^{(k)})$ is defined by

$$A_\epsilon(X^{(1)}, X^{(2)}, \cdots, X^{(k)}) = A_\epsilon$$

$$= \{(x^{(1)}, x^{(2)}, \cdots, x^{(k)}):$$

$$|-(1/n) \log p(s) - H(S)| < \epsilon,$$

$$\forall S \subseteq \{X^{(1)}, X^{(2)}, \cdots, X^{(k)}\}\}.$$

$$(3.8)$$

Let $A_\epsilon(S)$ denote the restriction of A_ϵ to the coordinates

corresponding to S. Thus if $S = (X^{(1)}, X^{(2)})$ we have

$$A_\epsilon(X^{(1)}, X^{(2)}) = \{(x^{(1)}, x^{(2)}):$$

$$|-1/n \log p(x^{(1)}, x^{(2)}) - H(X^{(1)}, X^{(2)})| < \epsilon$$

$$|-1/n \log p(x^{(1)}) - H(X^{(1)})| < \epsilon$$

$$|-1/n \log p(x^{(2)}) - H(X^{(2)})| < \epsilon\}. \qquad (3.9)$$

Notation: We shall use the notation

$$a_n \doteq 2^{nb} \qquad (3.10)$$

to mean

$$(1/n) \log a_n \to b. \qquad (3.11)$$

Lemma 1: For any $\epsilon > 0$, for sufficiently large n,

(i) $P\{A_\epsilon(S)\} \geqslant 1 - \epsilon, \quad \forall S \subseteq \{X^{(1)}, \cdots, X^{(k)}\}$

(ii) $\|A_\epsilon(S)\| \doteq 2^{nH(S)}$

(iii) $s \in A_\epsilon(S) \Rightarrow p(s) \doteq 2^{-n(H(S)-\epsilon)}$

(iv) Let $S_1, S_2 \subseteq \{X^{(1)}, X^{(2)}, \cdots, X^{(k)}\}$.

If

$$(s_1, s_2) \in A_\epsilon(S_1, S_2)$$

then

$$p(s_1|s_2) \doteq 2^{-n(H(S_1|S_2)-2\epsilon)}.$$

Proof: (i) follows from the convergence of the random variables in the definition of $A_\epsilon(S)$. (ii) follows immediately from the definition of $A_\epsilon(S)$. (iii) follows from

$$1 \geqslant \sum_{s \in A_\epsilon(S)} p(s) \overset{\text{(ii)}}{\geqslant} \sum_{A_\epsilon(S)} 2^{-n(H(S)+\epsilon)}$$

$$= \|A_\epsilon(S)\| 2^{-n(H(S)+\epsilon)} \qquad (3.12)$$

and

$$(1 - \epsilon) \leqslant \sum_{s \in A_\epsilon(S)} p(s) \overset{\text{(ii)}}{\leqslant} \|A_\epsilon(S)\| 2^{-n(H(S)-\epsilon)} \qquad (3.13)$$

Finally, $(s_1, s_2) \in A_\epsilon(S_1, S_2)$ implies

$$p(s_1, s_2) \doteq 2^{-n(H(S_1,S_2)-\epsilon)}$$

and

$$p(s_2) \doteq 2^{-(H(S_2)-\epsilon)}.$$

(iv) follows from

$$p(s_1|s_2) = p(s_1, s_2)/p(s_2).$$

Lemma 2: Let S_1, S_2 be two subsets of $X^{(1)}, \cdots, X^{(k)}$. For any $\epsilon > 0$, define $\|A_\epsilon(S_1|s_2)\|$ to be the number of s_1 sequences, jointly ϵ-typical with a particular s_2 sequence. If $s_2 \in A_\epsilon(S_2)$, then for arbitrarily large n

$$\|A_\epsilon(S_1|s_2)\| \leqslant 2^{n(H(S_1|S_2)+2\epsilon)}.$$

Moreover,

$$(1 - \epsilon)2^{n(H(S_1|S_2)-2\epsilon)} \leqslant \sum_{s_2} p(s_2)\|A_\epsilon(S_1|s_2)\|.$$

Proof: Similar to part (ii) of Lemma 1.

In achievability proofs, we shall need to know the probability that conditionally independent sequences are jointly typical. Let S_1, S_2, and S_3 be three subsets of $\{X^{(1)}, X^{(2)}, \cdots, X^{(k)}\}$.

Lemma 3:

(a) Let

$$P(S_1' = s_1, S_2' = s_2, S_3' = s_3) = \prod_{i=1}^n p(s_{1i}|s_{3i})p(s_{2i}|s_{3i})p(s_{3i}).$$

Then,

$$P\{S_1', S_2', S_3'\} \in A_\epsilon\} \doteq 2^{-nI(S_1;S_2|S_3)}.$$

Proof:

$$P\{(S_1', S_2', S_3') \in A_\epsilon\}$$

$$= \sum_{(s_1,s_2,s_3) \in A_\epsilon} p(s_3)p(s_1|s_3)p(s_2|s_3)$$

$$\doteq \|A_\epsilon(S_1, S_2, S_3)\| 2^{-nH(S_3)} 2^{-nH(S_1|S_3)} 2^{-nH(S_2|S_3)}$$

$$\doteq 2^{nH(S_1,S_2,S_3)} 2^{-n(H(S_3)+H(S_1|S_3)+H(S_2|S_3))}$$

$$= 2^{-nI(S_1;S_2|S_3)}.$$

IV. SHANNON'S CAPACITY THEOREM

We begin the theoretical discussion by reproving Shannon's original channel capacity theorem for the discrete memoryless channel (Fig. 10). When Shannon's work was published in 1948, the proofs were considered to be only heuristic—mere plausibility arguments. The first rigorous proof by Feinstein [13] 8 years later was completely different from Shannon's random coding idea. Further proofs by Wolfowitz, Fano, Gallager, and many others, were also different. The rigorous proof we give here supports the idea of Shannon's outline. It can be found as a problem in Gallager [6] and in Forney's unpublished class notes [14]. This proof technique will come up many times in subsequent sections. The fact that Shannon's original proof is the natural proof for multi-user networks is really quite remarkable.

The basic Shannon model for a communication channel is the discrete memoryless channel $(\mathfrak{X}, p(y|x), \mathfrak{Y})$ consisting of two alphabets—input alphabet $\mathfrak{X}$ and output alphabet $\mathfrak{Y}$—and a channel probability matrix $p(y|x)$. The interpretation is that if x is sent, then the received symbol Y is drawn according to probability mass function $p(y|x)$. *Discrete memoryless* means if a sequence $x_1, x_2, \cdots, x_n$ is sent, then the received sequence $y = (y_1, y_2, \cdots, y_n)$ is drawn according to $\Pi_{i=1}^n p(y_i|x_i)$.

A $(2^{nR}, n)$ code for a channel consists of a set of integers $[1, 2^{nR}]$ called the message set, an *encoding function*

$$x : [1, 2^{nR}] \to \mathfrak{X}^n$$

and a *decoding function*

$$g : \mathfrak{Y}^n \to [1, 2^{nR}].$$

If the message $w \in [1, 2^{nR}]$ is sent, let

$$\lambda(w) = P\{g(Y) \neq w | w \text{ sent}\} \qquad (4.1)$$

denote the conditional probability of error. Define the average probability of error of the code assuming a uniform distribution over the set of messages $[1, 2^{nR}]$, as

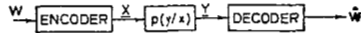$$\bar{P}_e^n = \frac{1}{2^{nR}} \sum_w \lambda(w). \qquad (4.2)$$

Fig. 10. The discrete memoryless channel.

The rate R of an $(2^{nR}, n)$ code is said to be *achievable* by the discrete memoryless channel if, for any $\epsilon > 0$, there exists a $(2^{nR}, n)$ code for all n sufficiently large such that $\bar{P}_e^n < \epsilon$.

The *capacity* C of the discrete memoryless channel is the supremum of the set of achievable rates. In 1948, Shannon showed that the capacity C of the discrete memoryless finite alphabet channel is given by the following.

Theorem 2 (Shannon): The capacity of the discrete memoryless channel $(\mathfrak{X}, p(y|x), \mathcal{Y})$ is given by

$$C = \sup_{p(x)} I(X; Y). \tag{4.3}$$

Proof: The proof that C is the capacity of a channel involves proving:

(i) *achievability* of any $R < C$, i.e., that there exists a sequence of $(2^{nR}, n)$ codes such that $\bar{P}_e^n \to 0$, and

(ii) a *converse* showing that given any sequence of $(2^{nR}, n)$ codes with $\bar{P}_e^n \to 0$ then $R < C$.

We shall only prove the achievability part of the theorem. The converses are discussed in the following sections of this paper.

Achievability: Assume that the maximization in (4.3) is achieved for a distribution $p^*(x)$ on $\mathfrak{X}$.

Random code: Choose 2^{nR} i.i.d. x sequences each with probability

$$p(x) = \prod_{i=1}^{n} p^*(x_i).$$

Decoding: Given y, choose the message w such that

$$(y, x(w)) \in A_\epsilon(X, Y)$$

is such a $w \in [1, 2^{nR}]$ exists and is unique, otherwise declare an error.

By symmetry of the random code construction, the probability of error (averaged over the random code) is independent of the index w sent. Thus without loss of generality assume that 1 is sent. Consider the events

$$E_w = \{(X(w), Y) \in A_\epsilon\}, \ w \in [1, 2^{nR}].$$

Then by the union of events bound

$$P_n = P\left(\bigcup_{w \neq 1} E_w \bigcup E_1^c\right)$$

$$\leq P(E_1^c) + \sum_{w \neq 1} P(E_w).$$

From Lemma 1, it follows that

$$P(E_1^c) \to 0.$$

From Lemma 3,

$$P(E_w) \doteq 2^{-nC}, \qquad \forall w \neq 1.$$

Therefore, if $R < C$, then

$$\sum_{w \neq 1} P(E_w) \doteq 2^{n(R-C)} \to 0.$$

If the error averaged over the random code tends to zero, then there must exist a sequence of codes with $\bar{P}_e^n \to 0$, and achievability is proved.

V. THE MULTIPLE ACCESS CHANNEL

The most completely understood multiple user channel is the multiple access channel shown in Fig. 11.

Many senders each simultaneously attempt to communicate a message to a common receiver. The most common example of this is a satellite receiver with many independent ground stations. After reception, the satellite will broadcast the information back to the ground, but that is the subject of the section on broadcast channels.

We see that the senders must contend not only with the receiver noise but with interference from each other as well.

The discrete memoryless multiple access channel $(\mathfrak{X}_1 \times \mathfrak{X}_2, p(y|x_1, x_2), \mathcal{Y})$ consists of three alphabets $\mathfrak{X}_1, \mathfrak{X}_2,$ and $\mathcal{Y}$ and a probability transition matrix $p(y|x_1, x_2)$.

A $((2^{nR_1}, 2^{nR_2}), n)$ code for the multiple access channel consists of two sets of integers $\mathcal{W}_1 = [1, 2^{nR_1}]$, $\mathcal{W}_2 = [1, 2^{nR_2}]$ called the message sets, two *encoding functions*

$$X_1 : \mathcal{W}_1 \to \mathfrak{X}_1^n$$

$$X_2 : \mathcal{W}_2 \to \mathfrak{X}_2^n$$

and a decoding function

$$g : \mathcal{Y}^n \to \mathcal{W}_1 \times \mathcal{W}_2.$$

Assuming uniform distribution over the product message sets $\mathcal{W}_1 \times \mathcal{W}_2$ i.e., that the messages are independent and equally likely, we define the *average probability of error* for a $((2^{nR_1}, 2^{nR_2}), n)$ code to be

$$\bar{P}_e^n = \frac{1}{2^{n(R_1 + R_2)}}$$

$$\sum_{(w_1, w_2) \in \mathcal{W}_1 \times \mathcal{W}_2} P\{g(Y) \neq (w_1, w_2)|(w_1, w_2) \text{ sent}\}.$$

$$(5.1)$$

A rate pair (R_1, R_2) is said to be *achievable* for the multiple access channel if there exists a sequence of $((2^{nR_1}, 2^{nR_2}), n)$ codes with $\bar{P}_e^n \to 0$.

The *capacity region* of the multiple access channel is the closure of the set of all achievable (R_1, R_2) rate pairs.

The capacity region of the multiple access channel was established by Ahlswede [2] and Liao [3]. The following proof is different from theirs.

Theorem 3: Multiple access capacity.

The capacity of the multiple access channel $(\mathfrak{X}_1 \times \mathfrak{X}_2, p(y|x_1, x_2), \mathcal{Y})$ is given by the convex hull of the union of the sets

$$C(p_1 p_2) = \{(R_1, R_2) : R_1 \leq I(X_1; Y|X_2)$$

$$R_2 \leq I(X_2; Y|X_1)$$

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

where

$$p(x_1, x_2) = p_1(x_1) p_2(x_2)\} \tag{5.2}$$

over all input product probability distributions on $\mathfrak{X}_1 \times \mathfrak{X}_2$.

Proof:

(i) *Achievability.*

Random coding: Fix $p_1(x_1)$, $p_2(x_2)$. Let $p(x_1, x_2) = p_1(x_1) p_2(x_2)$. Choose a random code of 2^{nR_1} x_1's $\in \mathfrak{X}_1^n$
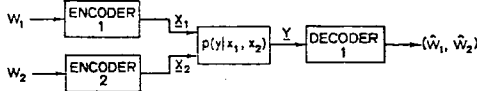
Fig. 11. The multiple access channel.

i.i.d $\sim \Pi_{i=1}^n p_1(x_{1i})$, and independently choose 2^{nR_2} x_2's $\in \mathfrak{X}_2^n$ i.i.d $\sim \Pi_{i=1}^n p_2(x_{2i})$. Index these sequences as $x_1(i)$, $x_2(j)$, $i \in [1, 2^{nR_1}], j \in [1, 2^{nR_2}]$.

Decoding: For decoding, given y, simply choose the pair (i, j) such that

$$(x_1(i), x_2(j), y) \in A_\epsilon$$

if such an $(i, j) \in [1, 2^{nR_1}] \times [1, 2^{nR_2}]$ exists and is unique— otherwise declare an error.

By symmetry of the random code construction, the probability of error (averaged over the random code) is independent of the index (i, j) sent. Thus without loss of generality assume that $(i, j) = (1, 1)$ is sent. Consider the events

$$E_{ij} = \{(X_1(i), X_2(j), Y) \in A_\epsilon^n\}.$$

Then, by the union of events bound

$$P_n = P\left(E_{11}^c \bigcup \bigcup_{(i,j) \neq (1,1)} E_{ij}\right)$$

$$\leqslant P(E_{11}^c + \sum_{i \neq 1, j=1} P(E_{i1}) + \sum_{i=1, j \neq 1} P(E_{1j}) + \sum_{i \neq 1, j \neq 1} P(E_{ij}).$$

(5.3)

From Lemma 1, $P(E_{11}^c) \to 0$.

Next, for $i \neq 1$,

$$P(E_{i1}) = \sum_{(x_1, x_2, y) \in A_\epsilon} p(x_1) p(x_2, y)$$

$$\doteq \|A_\epsilon\| 2^{-nH(X_1)} 2^{-nH(X_2, Y)}$$

$$\doteq 2^{-nI(X_1; Y|X_2)}$$

where the first equality follows from $i \neq 1$, which implies the independence of X_1 from (X_2, Y). The second and third inequalities follow from the definition of A_ϵ, Lemma 1, and from $I(X_1; X_2, Y) = I(X_1; Y|X_2)$ when X_1 and X_2 are independent.

Similarly, for $j \neq 1$,

$$P(E_{1j}) \doteq 2^{-nI(X_2; Y|X_1)}$$

and for $i \neq 1, j \neq 1$,

$$P(E_{ij}) \doteq 2^{-nI(X_1; X_2; Y)}.$$

If follows that

$$P_n \leqslant P(E_{11}^c) + 2^{nR_1} 2^{-nI(X_1; Y|X_2)} + 2^{nR_2} 2^{-nI(X_2; Y|X_1)}$$
$$+ 2^{n(R_1 + R_2)} 2^{-nI(X_1, X_2; Y)}.$$

(5.4)

Thus the conditions of the theorem cause each term to tend to zero as $n \to \infty$.

Time-sharing allows any (R_1, R_2) in the convex hull to be achieved, and the theorem is proved.

We now prove the converse to Theorem 3. This should provide the reader with some of the basic converse proof techniques.

(ii) Converse.

Given a $((2^{nR_1}, 2^{nR_2}), n)$ code for the MAC, the empirical probability distribution on $\mathfrak{W}_1 \times \mathfrak{W}_2 \times \mathfrak{X}_1^n \times \mathfrak{X}_2^n \times \mathfrak{Y}^n$ is of

the form

$$p(w_1, w_2, x_1, x_2, y) = \frac{1}{2^{n(R_1 + R_2)}}$$

$$\cdot p(x_1|w_1) p(x_2|w_2) p(y|x_1, x_2).$$

(5.5)

Fano's inequality [6] requires that

$$H(W_1, W_2|Y) \leqslant n(R_1 + R_2) \bar{P}_e^n + h(\bar{P}_e^n) \triangleq n\epsilon_n. \quad (5.6)$$

The assumption that $P_e^n \to 0$ implies that $\epsilon_n \to 0$. Consequently,

$$H(W_1 | Y) \leqslant n\epsilon_n$$
$$H(W_2 | Y) \leqslant n\epsilon_n. \quad (5.7)$$

Now consider,

(i) $$nR_1 = H(W_1) = I(W_1; Y) + H(W_1|Y)$$
$$\leqslant I(W_1; Y) + n\epsilon_n. \quad (5.8)$$

By the data processing inequality it follows that

$$nR_1 \leqslant I(W_1; Y) + n\epsilon_n \leqslant I(X_1; Y) + n\epsilon_n.$$

But since X_1 and X_2 are independent, we have

$$nR_1 \leqslant I(X_1; Y|X_2) + n\epsilon_n. \quad (5.9)$$

By symmetry it can be shown that

(ii) $$nR_2 \leqslant I(X_2; Y|X_1) + n\epsilon_n. \quad (5.10)$$

Finally,

(iii) $$n(R_1 + R_2) \leqslant I(W_1, W_2; Y) + n\epsilon_n. \quad (5.11)$$

The data processing theorem yields

$$I(W_1, W_2; Y) \leqslant I(X_1, X_2; Y). \quad (5.12)$$

Now the discrete memorylessness of the channel is easily applied to yield

$$R_1 \leqslant (1/n) I(X_1; Y|X_2) + \epsilon_n \leqslant (1/n) \sum_{i=1}^n I(X_{1i}; Y_i | X_{2i}) + \epsilon_n$$

$$R_2 \leqslant (1/n) I(X_2; Y|X_1) + \epsilon_n \leqslant (1/n) \sum_{i=1}^n I(X_{2i}; Y_i | X_{1i}) + \epsilon_n$$

and

$$R_1 + R_2 \leqslant (1/n) I(X_1, X_2; Y)$$

$$+ \epsilon_n \leqslant (1/n) \sum_{i=1}^n I(X_{1i}, X_{2i}; Y_i) + \epsilon_n \quad (5.13)$$

and the converse is proved.

VI. THE SLEPIAN-WOLF SOURCE CODING THEOREM

We know how to encode a source X. A rate $R > H(X)$ is sufficient. Now suppose that there are two sources $(X, Y) \sim p(x, y)$. A rate $R > H(X, Y)$ is sufficient. But what if the X-source and the Y-source must be separately described for some user who wishes to reconstruct X and Y? Clearly, by separately encoding X and Y, it is seen that a rate $R = R_x + R_y > H(X) + H(Y)$ is sufficient. However, in the surprising
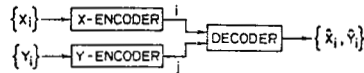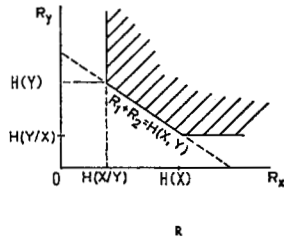
Fig. 12. Correlated source coding.

Fig. 13. Rate region for Slepian-Wolf data compression.

Fig. 14. Bins for Slepian and Wolf.

and fundamental paper of Slepian and Wolf [15], it is shown that a total rate $R = H(X, Y)$ suffices.

Example: Let X be a sequence of 1's and 0's generated by flipping a fair coin. Let Z be a sequence of 0's and 1's generated by flipping a coin with $\Pr(Z = 1) = p = 0.11$. Let $Y = X \oplus Z$. Note that Y is also a fair coin sequence.

Suppose that $R_y = 1$, so that Y is perfectly described. Then the information rate needed to describe X is not $R_x = 1$ but is $R_x = H(X|Y) = h(p) = \frac{1}{2}$. This is true despite the fact that the describer of X does not know the Y on which his description will be conditioned.

We now give a formal description of the Slepian-Wolf problem depicted in Fig. 12. The proof is different from that in [15] and [16] in that no time sharing is required.

A sequence $\{(X_i, Y_i)\}_{i=1}^n$ of independent copies of the pair (X, Y) of discrete random variables is to be encoded by two separate encoders; an X-encoder that observes the $\{X_i\}_{i=1}^n$ sequence and maps it into an integer $i \in [1, 2^{nR_x}]$, and a Y-encoder that observes the $\{Y_i\}_{i=1}^n$ sequence and maps it into the integer $j \in [1, 2^{nR_y}]$. The integers (i, j) are then communicated to a common decoder that tries to reproduce the $\{(X_i, Y_i)\}_{i=1}^n$ sequence. A $((2^{nR_x}, 2^{nR_y}), n)$ compression scheme for this problem consists of an integer n, two encoding functions

$$i : \mathfrak{X}^n \to [1, 2^{nR_x}]$$

$$j : \mathfrak{Y}^n \to [1, 2^{nR_y}]$$

and a decoding function

$$g : [1, 2^{nR_x}] \times [1, 2^{nR_y}] \to \mathfrak{X}^n \times \mathfrak{Y}^n.$$

The average probability of incorrect reproduction is given by

$$\bar{P}_e^n \triangleq P\{g(i(X), j(Y)) \neq (X, Y)\}.$$

A rate pair (R_x, R_y) is said to be *achievable* if there exist a sequence of compression schemes with $\bar{P}_e^n \to 0$. The problem is to find the set R of *achievable* compression rates (R_x, R_y). The rate region R (shown in Fig. 13) was established by Slepian and Wolf [15].

Theorem 4 (Slepian-Wolf): The set R of all achievable rates is given by

$$R = \{(R_x, R_y) : R_x > H(X|Y)$$

$$R_y > H(Y|X)$$

$$R_x + R_y > H(X, Y)\}. \tag{6.1}$$

The idea is to divide the $\mathfrak{X}^n$-space into 2^{nR_x} bins and the $\mathfrak{Y}^n$-space into 2^{nR_y} bins as shown in Fig. 14.
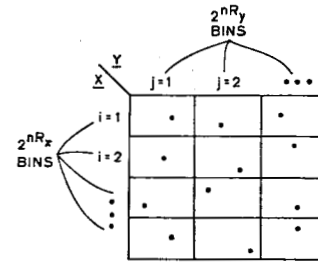
The dots in this figure correspond to jointly typical (x, y) pairs, and they ϵ-exhaust the probability. If R_x and R_y are large enough, there will be no more than one dot per product bin. Thus the name i of the x bin and the name j of the y bin will uniquely define the sequence (x, y) that falls in the product bin (i, j). We now prove this result.

Proof:

Encoding: Randomly assign every $x \in \mathfrak{X}^n$ to one of 2^{nR_x} bins, according to a uniform distribution over the integers $[1, 2^{nR_x}]$. More precisely, for every $x \in \mathfrak{X}^n$, let $p(i(x) = i) = 2^{-nR_x}$ for $i \in [1, 2^{nR_x}]$. Similarly, randomly assign every $y \in \mathfrak{Y}^n$ to one of 2^{nR_y} bins, such that $p(j(Y) = j) = 2^{-nR_y}, j \in [1, 2^{nR_y}]$.

Decoding: Given (i_0, j_0) declare $(\hat{x}, \hat{y}) = (x, y)$ was sent if there is one and only one pair of sequences (x, y) such that

$$i(x) = i_0 \qquad j(y) = j_0$$

and

$$(x, y) \in A_\epsilon(X, Y).$$

Otherwise declare an error. To bound $\bar{P}_e^n$, define the events

$$E_0 = \{(X, Y) \notin A_\epsilon(X, Y)\}$$

$$E_1 = \{\exists \text{ some } (x', y') \neq (X, Y), (x', y') \in A_\epsilon,$$

$$\text{such that } i(x') = i_0, j(y') = j_0\}.$$

Using the union of events bound, we obtain

$$\bar{P}_n = P(E_0 \cup E_1) \leq P(E_0) + P(E_1). \tag{6.2}$$

The first term in (6.1) $\to 0$ as $n \to \infty$ by typicality (Lemma 1). Notice that the event E_1 is equal to the union of the events

$$E_{11} = \{\exists \text{ some } x' \text{ such that } x' \neq X$$

$$\text{and } i(x') = i_0, (x', y) \in A_\epsilon, y = Y\}$$

$$E_{12} = \{\exists \text{ some } y' \text{ such that } y' \neq Y \text{ and } j(y') = j_0,$$

$$(x, y') \in A_\epsilon, \text{ and } x = X\}$$

and

$$E_{13} = \{\exists \text{ some } (x', y') \in A_\epsilon \text{ such that } x' \neq X \text{ and } y' \neq Y$$

$$\text{and } i(x') = i_0 \text{ and } j(y') = j_0\}.$$

First consider the probability of E_{11}. By the union of events bound

$$P(E_{11}) \leq \sum_{\substack{x' \neq x \\ (x', y) \in A_\epsilon(X, Y)}} P\{i(x') = i_0\}$$

$$= 2^{-nR_x} \|A_\epsilon(X|y)\|.$$

But from Lemma 2

$$\|A_\epsilon(X|y)\| \leqslant 2^{n(H(X|Y)+2\epsilon)}.$$

Therefore,

$$P(E_{11}) \to 0 \quad \text{as} \quad n \to \infty$$

if

$$R_x > H(X|Y).$$

Similarly, the probability of the events E_{12} and $E_{13} \to 0$ as $n \to \infty$ if $R_y > H(Y|X)$ and $R_x + R_y > H(Y, X)$, respectively.

Finally, the probability of error is averaged over random partitions. Thus there exists maps $i^*(\cdot)$, $j^*(\cdot)$ such that $\overline{P}_e \to 0$.

The Slepian–Wolf theorem presented in this section is historically the first fundamental result in multiple user source coding theory. For a comprehensive overview of that theory, the reader is referred to Berger [17].

VII. THE MULTIPLE ACCESS CHANNELS WITH CORRELATED SOURCES

From Slepian–Wolf, we know how to give efficient separate descriptions of correlated sources. From the multiple access channel, we know how to send two separate independent descriptions over a noisy channel. Apparently we can weld these two problems together to obtain a theory of sending correlated sources over a multiple access channel (Fig. 15).

We do so in this section and find that the combined theory involves a new ingredient—correlation of the inputs for the multiple access channel. As a byproduct, we shall find a common proof of the Slepian–Wolf theorem and the multiple access channel capacity region [18].

Assume we have two information sources $U_1, U_2, \cdots$ and $V_1, V_2, \cdots$ generated by repeated independent drawings of a pair of discrete random variables U and V from a given bivariate distribution $p(u, v)$.

A block code for the channel consists of an integer n, and two encoding functions

$$x_1^n : \mathcal{U}^n \to \mathcal{X}_1^n \qquad x_2^n : \mathcal{V}^n \to \mathcal{X}_2^n$$

assigning codewords to the source outputs, and a decoding function

$$d^n : \mathcal{Y}^n \to \mathcal{U}^n \times \mathcal{V}^n.$$

The probability of error is given by

$$P_n = P\{(U^n, V^n) \neq d^n(Y^n)\}$$

$$= \sum_{(u,v) \in \mathcal{U}^n \times \mathcal{V}^n} p(u^n, v^n) P\{d^n(y^n) \neq (u^n, v^n)|(U^n, V^n)$$

$$= (u^n, v^n)\} \tag{7.1}$$

where the joint probability mass function is given, for a code assignment $\{x_1(u^n), x_2(v^n)\}$, by

$$p(u, v, y) = \prod_{i=1}^{n} p(u_i, v_i) p(y_i|x_{1i}(u^n), x_{2i}(v^n)). \tag{7.2}$$

Definition: The source (U, V) can be reliably transmitted over the multiple access channel $(\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{Y}, p(y|x_1, x_2))$ if there exists a sequence of block codes $(x_1^n(u^n), x_2^n(v^n), d^n(y^n))$ such that
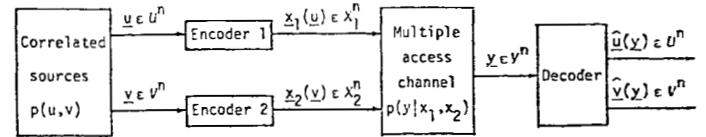
$$P_n = P\{d^n(Y^n) \neq (U^n, V^n)\} \to 0.$$

Fig. 15. The multiple access channel with arbitrarily correlated sources.

Example: Consider the transmission of the correlated sources (U, V) with the joint distribution $p(u, v)$ given by

V \ U	0	1
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

over the multiple access channel defined by

$$\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$$

$$\mathcal{Y} = \{0, 1, 2\}$$

$$Y = X_1 + X_2.$$

Here $H(U, V) = \log 3 = 1.58$ bits. On the other hand, if X_1 and X_2 are independent,

$$\max_{p(x_1)p(x_2)} I(Y; X_1, X_2) = 1.5 \text{ bits.}$$

Thus $H(U, V) > I(Y; X_1, X_2)$ for all $p(x_1) p(x_2)$. Consequently there is no way, even with the use of Slepian–Wolf data compression on U and V, to use the standard multiple access channel capacity region to send U and V reliably to Y. However, it is easy to see that with the choice $X_1 \equiv U$, and $X_2 \equiv V$, error-free transmission of the source over the channel is possible. This example shows that the separate source and channel coding described above is not optimal—the partial information that each of the random variables U and V contains about the other is destroyed in this separation.

To allow partial cooperation between the two transmitters, we allow our codes to depend statistically on the source outputs. This induces dependence between codewords.

We shall outline here a proof of a special case of Theorem 1 [18], in which U and V have no common part. In this case, we must show that U and V can be reliably sent to Y if, for $p(u, v) p(x_1|u) p(x_2|v) p(y|x_1, x_2)$,

$$H(U|V) < I(X_1; Y|X_2, V)$$

$$H(V|U) < I(X_2; Y|X_1, U) \tag{7.3}$$

$$H(U, V) < I(X_1, X_2; Y). \tag{7.4}$$

The proof will employ random coding. We first describe the random code generation and encoding–decoding schemes, and then analyze the probability of error.

Generating Random Codes: Fix $p(x_1|u)$ and $p(x_2|v)$; for each $u \in \mathcal{U}^n$ generate *one* x_1 sequence drawn according to $\prod_{i=1}^{n} p(x_{1i}|u_i)$ and for each $v \in \mathcal{V}^n$ generate *one* x_2 sequence drawn according to $\prod_{i=1}^{n} p(x_{2i}|v_i)$. Call these sequences $x_1(u)$ and $x_2(v)$, respectively.

Encoding: Transmitter 1, upon observing u at the output of source 1, transmits $x_1(u)$ and transmitter 2, after observing v at the output of source 2, transmits $x_2(v)$. Assume the maps $x_1(\cdot)$, $x_2(\cdot)$ are known to the receiver.
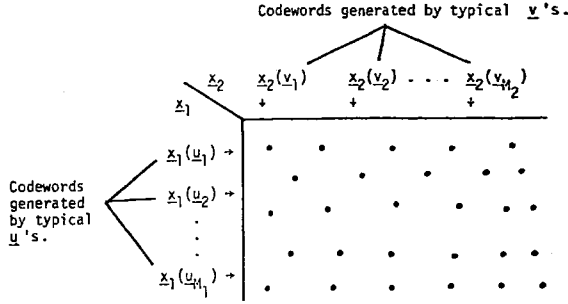
Codewords generated by typical $\underline{v}$'s.

Fig. 16. *Picture of joint typicality for multiple access channel.* The dots correspond to jointly typical (X_1, X_2) pairs. Note that only $2^{nH(U,V)}$ $(x_1(u), x_2(v))$ pairs are likely to occur.

Decoding: Upon receiving y, the decoder finds the only (u, v) pair such that $(u, v, x_1(u), x_2(v), y) \in A_\epsilon$, where A_ϵ is the set of jointly ϵ-typical sequences. If there is no such (u, v) pair, or there exists more than one such pair, the decoder declares an error. A helpful picture is given in Fig. 16.

Error: Suppose (U_0, V_0) is the source output, and define the events

$$E_{(u, v)} = \{(u, v, x_1(u), x_2(v), y) \in A_\epsilon\}.$$

The probability of error averaged over (U_0, V_0) and the random codes is given by

$$P_n = P \left\{ E^c_{(U_0, V_0)} \bigcup \bigcup_{(u,v) \neq (U_0, V_0)} E_{(u,v)} \right\}$$

$$\leqslant P \left\{ E^c_{(U_0, V_0)} \right\} + P \left\{ \bigcup_{(u,v) \neq (U_0, V_0)} E_{(u,v)} \right\}.$$

The first term $\to 0$ by Lemma 1. The second term is bounded above by

$$\sum_{(u_0, v_0) \in A_\epsilon} p(u_0, v_0)$$

$$\cdot \sum_{\substack{u' \neq u_0 \\ v' = v_0}} P \{(u', v', X_1(u'), X_2(v'), Y) \in A_\epsilon | (u_0, v_0)\}$$

$$+ \sum_{(u_0, v_0) \in A_\epsilon} p(u_0, v_0) \sum_{\substack{u' = u_0 \\ v' \neq v_0}} P \{\cdot\}$$

$$+ \sum_{(u_0, v_0) \in A_\epsilon} p(u_0, v_0) \sum_{\substack{u' \neq u_0 \\ v' \neq v_0}} P \{\cdot\}$$

$$\leqslant 2^{n(H(U|V)+\epsilon)} 2^{-n(I(X_1; Y|X_2, V)-\epsilon)}$$

$$+ 2^{n(H(V|U)+\epsilon)} 2^{-n(I(X_2; Y|X_1, U)-\epsilon)}$$

$$+ 2^{nH(U,V)} 2^{-n(I(X_1, X_2; Y)-\epsilon)}. \tag{7.5}$$

Consequently, $P_n \to 0$ if the conditions in (7.3) are satisfied.

VIII. THE BROADCAST CHANNEL

The broadcast channel [19] is a communication network with one transmitter and many receivers.

The basic problem is to find the set of simultaneously achievable rates (R_0, R_1, R_2). To date this problem has not been solved. The special case of sequentially degraded channels has been solved by Bergmans [20] and Gallager [21]. An achievable rate region for the general broadcast channel has
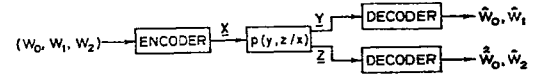

Fig. 17. The discrete memoryless broadcast channel.

been put forth by Marton [22], but is not known to be the capacity region.

We now give the formal definition of the two-receiver broadcast channel problem. A discrete memoryless broadcast channel $(\mathfrak{X}, p(y, z \mid x), \mathfrak{Y} \times \mathfrak{Z})$ as depicted in Fig. 17 consists of three finite alphabets $\mathfrak{X}, \mathfrak{Y}, \mathfrak{Z}$ and a probability transition matrix $p(y, z \mid x)$.

An $((2^{nR_0}, 2^{nR_1}, 2^{nR_2}), n)$ code for a broadcast channel consists of three sets of integers

$$\mathbb{U}_0 = [1, 2^{nR_0}], \quad \mathbb{U}_1 = [1, 2^{nR_1}], \quad \text{and} \quad \mathbb{U}_2 = [1, 2^{nR_2}]$$

an *encoding function*

$$X: \mathbb{U}_0 \times \mathbb{U}_1 \times \mathbb{U}_2 \to \mathfrak{X}^n$$

and two *decoding functions*

$$g_1: \mathfrak{Y}^n \to \mathbb{U}_0 \times \mathbb{U}_1$$
$$g_2: \mathfrak{Z}^n \to \mathbb{U}_0 \times \mathbb{U}_2.$$

The integer w_0 has the interpretation of the common part of the message, while the integers w_1, w_2 are called the independent parts of the message. Assuming uniform distribution on the set of messages $\mathbb{U}_0 \times \mathbb{U}_1 \times \mathbb{U}_2$, define

$$\bar{P}_e^n = \frac{1}{2^{n(R_0+R_1+R_2)}}$$

$$\cdot \sum_{(w_0, w_1, w_2) \in \mathbb{U}_0 \times \mathbb{U}_1 \times \mathbb{U}_2} P \{g_1(Y) \neq (w_0, w_1) \text{ or }$$

$$g_2(Z) \neq (w_0, w_2) | (w_0, w_1, w_2) \text{ sent}\} \tag{8.1}$$

to be the *average probability of error* of the code.

The rate triple (R_0, R_1, R_2) is said to be *achievable* by a broadcast channel if there exists a sequence of $((2^{nR_0}, 2^{nR_1}, 2^{nR_2}), n)$ codes with $\bar{P}_e^n \to 0$. The *capacity region* C for the broadcast channel is the closure of the set of all achievable rate triples (R_0, R_1, R_2). We now look at a family of broadcast channels in which one receiver can be considered worse than the other. The mathematical definition corresponding to this physical idea is as follows.

Definition: The broadcast channel $(\mathfrak{X}, p(y, z \mid x), \mathfrak{Y} \times \mathfrak{Z})$ is called *degraded* if there exists a probability transition matrix $p(z \mid y)$ such that for all $z \in \mathfrak{Z}$ and $x \in \mathfrak{X}$

$$p(z \mid x) = \sum_{y \in \mathfrak{Y}} p(y, z \mid x) = \sum_{y \in \mathfrak{Y}} p(z \mid y) p(y \mid x). \tag{8.2}$$

Theorem 5: The capacity region of the degraded broadcast channel $(\mathfrak{X}, p(y, z \mid x), \mathfrak{Y} \times \mathfrak{Z})$ is given by

$$C = \{(R_0, R_1, R_2): R_0 + R_2 < I(U; Z), R_1 < I(X; Y | U)$$

for some $p(u) p(x \mid u)$,

$$\|\mathfrak{U}\| \leqslant \min \{\|\mathfrak{X}\|, \|\mathfrak{Y}\|, \|\mathfrak{Z}\|\}\}. \tag{8.3}$$

Proof: Fix $p(u), p(x \mid u)$.

a) Achievability: The natural coding idea necessary to solve this channel is that of superposition coding [19], [20]. We first find a low rate code for the worst receiver Z. Since

this rate is below the capacity from X to Z, it is possible to further randomize the codewords chosen for Z into a small cloud of satellite codewords intended for the good receiver Y. Receiver Z then decodes the cloud center. Receiver Y first decodes the cloud center, which he can do because he can "see" at least as much as Z_1. Receiver Y then decodes the satellite codeword.

(i) Random Code: First generate $2^{n(R_2+R_0)}$ i.i.d. sequences u, each with probability

$$p(u) = \prod_{i=1}^{n} p(u_k).$$

Label these $u(i), i \in [1, 2^{n(R_0+R_2)}]$.

For every $u(i)$, generate 2^{nR_1} conditionally independent x sequences, each with probability

$$p(x \,|\, u(i)) = \prod_{k=1}^{n} p(x_k | u_k(i)).$$

Label these $x(i, j), j \in [1, 2^{nR_1}]$.

Here $u(i)$ plays the role of the cloud center understandable to both Y and Z, while $x(i, j)$ is the jth satellite codeword in the ith cloud.

(ii) Decoding: If y is received, declare $(\hat{i}, \hat{j}) = (i, j)$ was sent if there is one and only one pair $(i, j) \in [1, 2^{n(R_0+R_2)}] \times [1, 2^{nR_1}]$, such that $(u(i), x(i, j), y) \in A_\epsilon(U, X, Y)$. If z is received, declare $\hat{\hat{i}} = i$ was sent if there is one and only one i such that $(u(i), z) \in A_\epsilon(U, Z)$. Let I, J be independent r.v.'s drawn according to uniform distributions on $[1, 2^{n(R_0+R_2)}] \times [1, 2^{nR_1}]$. Let the code be chosen randomly according to the encoding description. Then the probability of error averaged over I, J and the random code is given by

$$P_n = P\{(\hat{I}, \hat{J}) \neq (I, J) \text{ or } \hat{\hat{I}} \neq I\}. \tag{8.4}$$

By the symmetry induced by the random coding, we see that each transmitted message (i, j) yields the same probability of error. Thus, setting $(i, j) = (1, 1)$, we have

$$P_n = P\{(\hat{I}, \hat{J}) \neq (1, 1) \text{ or } \hat{\hat{I}} \neq 1 | (I, J) = (1, 1)\}.$$

Define the events

$$E_{Zi} = \{(U(i), Z) \in A_\epsilon(U, Z)\}$$
$$E_{Yi} = \{(U(i), Y) \in A_\epsilon(U, Y)\}$$
$$E_{Yij} = \{(U(i), X(i, j), Y) \in A_\epsilon(U, X, Y)\}.$$

Then

$$P_n = P\{E_{Z1}^c \cup E_{Y11}^c \bigcup_{i \neq 1} E_{Zi} \bigcup_{i \neq 1} E_{Yi} \bigcup_{j \neq 1} E_{Y1j}\} \leq P\{E_{Z1}^c\}$$

$$+ P\{E_{Y11}^c\} + \sum_{i \neq 1} P\{E_{Zi}\} + \sum_{i \neq 1} P\{E_{Yi}\} + \sum_{j \neq 1} P\{E_{Y1j}\}.$$

The first two terms correspond to the event that the correct codeword does not belong to the decoding set. The last terms correspond to the event that some incorrect codeword belongs to the decoding set. From Lemma 1, it follows that

$$P\{E_{Z1}^c\} \to 0, \quad P\{E_{Y11}^c\} \to 0.$$

Consider the event E_{Zi}. We observe, for $i \neq 1$, that $U(i)$ and Z are independent. Thus, by Lemma 3, for $i \neq 1$

$$P\{E_{Zi}\} \doteq 2^{-nI(U;Z)}.$$

Consequently,

$$\sum_{i \neq 1} P\{E_{Zi}\} \doteq 2^{-n(I(U;Z) - (R_0 + R_2))}.$$

This term goes to zero if

$$R_0 + R_2 < I(U; Z).$$

Similarly, for the event $E_{Yi}, i \neq 1$

$$P\{E_{Yi}\} \doteq 2^{-nI(U;Y)}.$$

and

$$\sum_{i \neq 1} P\{E_{Yi}\} \to 0, \quad \text{if } R_0 + R_2 < I(U; Y) \tag{8.5}$$

But since by the degradedness of the broadcast channel $I(U; Y) \leq I(U; Z)$, the condition (8.5) is redundant.

Next, consider the event E_{Y1j}. From Lemma 3, it follows that

$$P\{E_{Y1j}\} \doteq 2^{-nI(X;Y|U)}.$$

Thus the term

$$\sum_{j \neq 1} P\{E_{Y1j}\} \to 0, \quad \text{if } R_1 < I(X; Y|U). \tag{8.6}$$

b) Converse: (See Gallager [21] and Bergmans [23].)

The capacity region for the general broadcast channel is still an open problem. Several special cases have been recently solved (e.g., degraded message sets [24], more capable class [25], deterministic class [26], [27], and parallel degraded channels [28]). The most general inner bound to the capacity region of the broadcast channel is that given by Marton [22]. In the following theorem, we outline a proof of a special case of Marton's general result where it is assumed that no common part is decoded. However, this special case isolates a new coding idea. This, together with superposition, yields Marton's theorem. The proof of the following theorem is due to El Gamal and van der Meulen [29].

Theorem: Let

$$\mathcal{R}_0 = \{(R_1, R_2): R_1, R_2 \geq 0,$$

$$R_1 \leq I(U; Y),$$

$$R_2 \leq I(V; Z),$$

$$R_1 + R_2 \leq I(U; Y) + I(V; Z) - I(U; V),$$

$$\text{for some } p(u, v, x) \text{ on } \mathcal{U} \times \mathcal{V} \times \mathcal{X}\} \tag{8.7}$$

Then any rate pair $(R_1, R_2) \in \mathcal{R}_0$ is achievable for

$$(\mathcal{X}, p(y, z | x), \mathcal{Y} \times \mathcal{Z}).$$

Proof: (Outline)

Fix $p(u, v), p(x | u, v)$. The channel $p(y, z | x)$ is given. The idea is to send u to y and v to z.

Random Coding: Generate $2^{nI(U;Y)}$ typical u's in U^n. Generate $2^{nI(V;Z)}$ typical V's in V^n. Now randomly throw the u's into 2^{nR_1} bins and the v's into 2^{nR_2} bins. For each product bin, find a *jointly* typical (u, v) pair. This can be done, as shown by the circled dots in Fig. 18, if

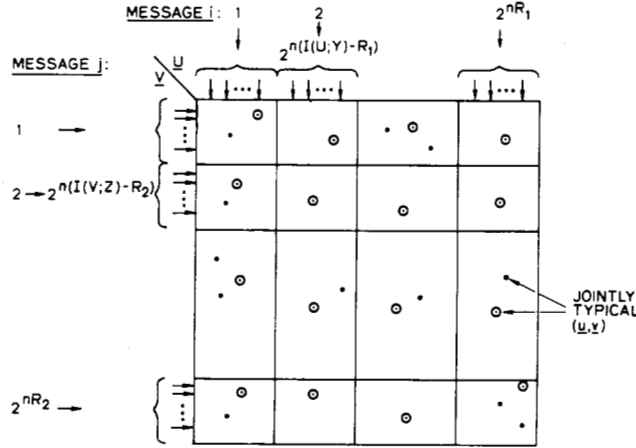$$R_1 + R_2 < I(U; Y) + I(V; Y) - I(U; V). \tag{8.8}$$

Fig. 18. Coding for Marton's Theorem.

To see this, we recall from Lemma 3 that independent choices of u and v result in a jointly typical (u, v) with probability $2^{-nI(U;V)}$. Now there are $2^{n(I(U;Y)-R_1)}$ U's in any U-bin, and $2^{n(I(V;Z)-R_2)}$ V's in any V-bin. Thus the expected number of jointly typical (U, V) pairs in a given product $U \times V$ bin is

$$2^{n(I(U;Y)-R_1)} 2^{n(I(V;Z)-R_2)} 2^{-nI(U;V)}. \qquad (8.9)$$

The desired jointly typical (u, v) pair can be found if this expected number is much greater than 1. This is guaranteed if (8.8) holds.

Continuing with the coding, for each $U \times V$ bin and its designated jointly typical (u, v) pair, generate $x(u, v)$ according to the conditional distribution $\prod_{k=1}^{n} p(x_k | u_k, v_k)$.

Encoding: To send i to Y and j to Z, send $x(u, v)$, where (u, v) is the designated pair in the product bin (i, j).

Decoding: Receiver Y, upon receiving y, finds the u such that (u, y) is jointly typical. Thus it is necessary that $R_1 < I(U; Y)$. He then finds the index i of the bin in which u lies. Receiver Z finds the v such that (v, z) is jointly typical. Thus we need $R_2 < I(V; Z)$. He then finds the index j of the bin in which v lies.

A Note on Feedback: In [30], it was shown that feedback cannot increase the capacity of the physically degraded broadcast channel, i.e., broadcast channels for which $p(y, z | x) = p(y | x) p(z | y)$. Surprisingly, it was later shown by Ozarow [31] and Dueck [32] that feedback can in fact increase the capacity of general broadcast channels.

IX. The Relay Channel

The discrete memoryless relay channel denoted by $(\mathfrak{X}_1 \times \mathfrak{X}_2, p(y, y_1 | x_1, x_2), \mathfrak{Y} \times \mathfrak{Y}_1)$ consists of four finite sets $\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{Y}, \mathfrak{Y}_1$ and a collection of probability distributions $p(\cdot, \cdot | x_1, x_2)$ on $\mathfrak{Y} \times \mathfrak{Y}_1$ one for each $(x_1, x_2) \in \mathfrak{X}_1 \times \mathfrak{X}_2$. The interpretation is that x_1 is the input to the channel and y is the output, y_1 is the relay's output and x_2 is the input symbol chosen by the relay as shown in Fig. 19. The problem is to find the capacity of the channel between the sender x_1 and the receiver y.

The relay channel was introduced by van der Meulen [33]. The following discussion is based on Cover and El Gamal [34].

A $(2^{nR}, n)$ code for the relay channel consists of a set of integers $\mathfrak{V}$ $[1, 2^{nR}]$, an encoding function
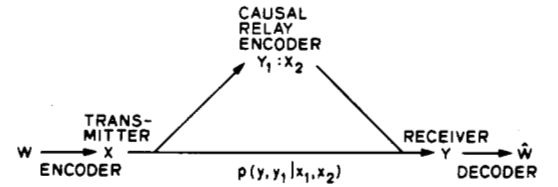
$$x_1 : [1, 2^{nR}] \to \mathfrak{X}_1^n$$

Fig. 19. The relay channel.

a set of relay functions $\{f_i\}_{i=1}^{n}$ such that

$$x_{2i} = f_i(Y_{11}, Y_{12}, \cdots, Y_{1i-1}), \qquad 1 \leqslant i \leqslant n$$

and a decoding function

$$g : \mathfrak{Y}^n \to [1, 2^{nR}].$$

Note that the allowed encoding functions actually form part of the definition of the relay channel because of the nonanticipatory relay condition. The relay channel input x_{2i} is allowed to depend only on the past $y_{11}, y_{12}, \cdots, y_{1i-1}$. This is the definition used by van der Meulen [33]. The channel is memoryless in the sense that (y_i, y_{1i}) depends on the past only through the current transmitted symbols (x_{1i}, x_{2i}). Thus for any choice $p(w)$, $w \in \mathfrak{V}$, and code choice $x_1 : [1, 2^{nR}] \to \mathfrak{X}_1^n$ and relay functions $\{f_i\}_{i=1}^{\infty}$, the joint probability mass function on $\mathfrak{V} \times \mathfrak{X}_1^n \times \mathfrak{X}_2^n \times \mathfrak{Y}^n \times \mathfrak{Y}_1^n$ is given by

$$p(w, x_1, x_2, y, y_1) = p(w) \prod_{i=1}^{n} p(x_{1i} | w) p(x_{2i} | y_{11}, y_{12}, \cdots,$$

$$y_{1i-1}) p(y_i, y_{1i}, x_{1i}, x_{2i}). \qquad (9.1)$$

If the message $w \in [1, 2^{nR}]$ is sent, let

$$\lambda(w) = \Pr \{g(Y) \neq w | w \text{ sent}\}$$

denote the conditional probability of error. We define the average probability of error of the code as

$$\bar{P}_e^n = 2^{-nR} \sum_w \lambda(w).$$

The probability of error is calculated under uniform distribution over the codewords $w \in [1, 2^{nR}]$. The rate R is said to be *achievable* by a relay channel if there exists a sequence of $(2^{nR}, n)$ codes with $\bar{P}_e^n \to 0$. The *capacity* C of the relay channel is the supremum of the set of achievable rates.

We first give an upper bound to the capacity of any relay channel [34]:

Theorem 9.1: For any relay channel $(\mathfrak{X}_1 \times \mathfrak{X}_2, p(y, y_1 | x_1, x_2), \mathfrak{Y} \times \mathfrak{Y}_1)$ the capacity C is bounded above by

$$C \leqslant \sup_{p(x_1, x_2)} \min \{I(X_1, X_2; Y), I(X_1; Y, Y_1 | X_2)\} \quad (9.2)$$

This upper bound has a nice max flow min cut interpretation. The first term in (9.2) upper bounds the rate of information flow from senders X_1 and X_2 to receiver Y. The second term bounds the rate of transmission from X_1 to Y and Y_1.

We now consider a family of relay channels in which the relay receiver Y_1 is better than the ultimate receiver Y in the sense defined below. Here the max flow min cut upper bound in (9.2) is achieved.

Definition: The relay channel $(\mathfrak{X}_1 \times \mathfrak{X}_2, p(y, y_1 | x_1, x_2), \mathfrak{Y} \times \mathfrak{Y}_1)$ is said to be *degraded* if $p(y, y_1 | x_1, x_2)$ can be written in the form

$$p(y, y_1 | x_1, x_2) = p(y_1 | x_1, x_2) p(y | y_1, x_2). \qquad (9.3)$$

Thus Y is a random degradation of the relay signal Y_1.

For the degraded relay channel, the capacity is given by the following.

Theorem 9.2: The capacity C of the degraded relay chanel is given by

$$C = \sup_{p(x_1, x_2)} \min \{I(X_1, X_2; Y), I(X_1; Y_1|X_2)\} \quad (9.4)$$

where the supremum is over all joint distributions $p(x_1, x_2)$ on $\mathfrak{X}_1 \times \mathfrak{X}_2$.

Proof: Converse.

The converse follows from Theorem 9.1 and by degradedness. Thus

$$I(X_1; Y, Y_1|X_2) = I(X_1; Y_1|X_2). \quad (9.5)$$

This capacity region proof involves 1) random coding, 2) list codes, 3) Slepian–Wolf partitioning, 4) coding for the cooperative multiple access channel, 5) superposition coding, and 6) block Markov encoding at the relay and transmitter.

Achievability (Outline): We consider B blocks of transmission, each of n symbols. A sequence of $B - 1$ indices $W_i \in [1, 2^{nR}]$, $i = 1, 2, \cdots, B - 1$ will be sent over the channel in nB transmissions. (Note that as $B \to \infty$, for fixed n, that the rate $R(B - 1)/B$ is arbitrarily close to R.)

In each n-block $b = 1, 2, \cdots, B$, we shall use the same doubly indexed set of codewords

$$C = \{x_1(w|s), x_2(s)\}, w \in [1, 2^{nR}], s \in [1, 2^{nR_0}], x_1(\cdot|\cdot)$$
$$\in \mathfrak{X}_1^n, x_2(\cdot) \in \mathfrak{X}_2^n. \quad (9.6)$$

We shall also need a partition

$$\mathcal{S} = \{S_1, S_2, \cdots, S_{2^{nR_0}}\} \quad \text{of} \quad \mathfrak{W} = \{1, 2, \cdots, 2^{nR}\}$$

into

$$2^{nR_0} \text{ cells}, S_i \cap S_j = \phi, i \neq j, \cup S_i = \mathfrak{W}.$$

The partition $\mathcal{S}$ will allow us to send side information to the receiver in the manner of Slepian and Wolf [15].

Random Coding: First generate at random 2^{nR_0} independent identically distributed n-sequences in $\mathfrak{X}_2^n$, each drawn according to $p(x_2) = \prod_{i=1}^n p(x_{2i})$. Index them as $x_2(s)$, $s \in [1, 2^{nR_0}]$. For each $x_2(s)$, generate 2^{nR} conditionally independent n-sequences $x_1(w|s)$, $w \in [1, 2^{nR}]$ drawn according to $p(x_1|x_2(s)) = \prod_{i=1}^n p(x_{1i}|x_{2i}(s))$. This defines a random codebook $C = \{x_1(w|s), x_2(s)\}$.

The random partition $\mathcal{S} = \{S_1, S_2, \cdots, S_{2^{nR_0}}\}$ of $\{1, 2, \cdots, 2^{nR}\}$ is defined as follows. Let each integer $w \in [1, 2^{nR}]$ be assigned independently, according to a uniform distribution over the indices $s = 1, 2, \cdots, 2^{nR_0}$, to cells S_s.

Encoding: Let $w_i \in [1, 2^{nR}]$ be the new index to be sent in block i, and assume that $w_{i-1} \in S_{s_i}$. Then the encoder sends $x_1(w_i|s_i)$. The relay has an estimate $\hat{w}_{i-1}$ of the previous index w_{i-1}. (This will be made precise in the decoding section). Assume that $\hat{w}_{i-1} \in S_{\hat{s}_i}$. Then the relay encoder sends the codeword $X_2(\hat{s}_i)$ in block i.

Decoding: We assume that at the end of block $(i - 1)$ the receiver knows $(w_1, w_2, \cdots, w_{i-2})$ and $(s_1, s_2, \cdots, s_{i-1})$ and the relay knows $(w_1, w_2, \cdots, w_{i-1})$ and consequently $(s_1, \cdots, s_i)$.

The decoding procedures at the end of block i are as follows.

1) Knowing s_i, and upon receiving $y_1(i)$, the *relay* receiver estimates the message of the transmitter $\hat{w}_i = w$ iff there exists a *unique* w such that $(x_1(w|s_i), x_2(s_i), y_1(i))$ are jointly ϵ-

typical. Using Lemma 3 it can be shown that $\hat{w}_i = w_i$ with arbitrarily small probability of error if

$$R < I(X_1; Y_1|X_2). \quad (9.7)$$

and n is sufficiently large.

2) The *receiver* declares that $\hat{s}_i = s$ was sent iff there exists one and only one s such that $(x_2(s), Y(i))$ are jointly ϵ-typical. From Lemma 3 we know that s_i can be decoded with arbitrarily small probability of error if

$$R_0 < I(X_2; Y) \quad (9.8)$$

and n is sufficiently large.

3) Assuming that s_i is decoded successfully at the receiver, then $\hat{w}_{i-1} = w$ is declared as the index sent in block $(i - 1)$ iff there is a unique $w \in S_{s_i} \cap \mathfrak{L}(Y(i - 1))$ where $\mathfrak{L}(Y(i - 1))$ is the list of indices w that the receiver y considered to be "typical" with $Y(i - 1)$ in the $(i - 1)$th block. If n is sufficiently large and if

$$R < I(X_1; Y|X_2) + R_0 \quad (9.9)$$

then $\hat{w}_{i-1} = w_{i-1}$ with arbitrarily small probability of error. Combining the two constraints (9.8) and (9.9) R_0 drops out, leaving

$$R < I(X_1; Y|X_2) + I(X_2; Y) = I(X_1, X_2; Y). \quad (9.10)$$

For a detailed analysis of the probability of error, the reader is referred to [34].

Theorem 9.1 can also be shown to be the capacity for the following classes of relay channels.

(i) Reversely degrades, i.e.,

$$p(y, y_1|x_1, x_2) = p(y_1|x_1, x_2)p(y|y_1, x_2).$$

(ii) Relay channel with feedback.

(iii) Deterministic relay channel [35]

$$y_1 = f(x_1, x_2) \quad y = g(x_1, x_2).$$

A general lower bound to the capacity of any relay channel can be found in [34].

X. SUMMARY AND OPEN PROBLEMS

Now that we have presented the Gaussian multiple user channels in Section II and proved some of the results in detail in subsequent sections, it is time to abstract the salient points of the theory. We do so by paralleling the discussion of Section II for the channels shown in Fig. 20.

1) The Shannon Channel: The codewords are n-vectors $x(1), x(2), \cdots, x(2^{nR})$. First suppose $R < I(X; Y)$. The intuitive idea is that 2^{nR} x sequences, each independent identically distributed according to $\prod_{i=1}^n p(x_i)$ will be mutually far apart in the sense that if one sends an $x(i)$ and receives a Y, then, looking back from the Y, one finds a unique x in the codebook that is jointly typical with Y. This is the correct $x(i)$.

2) List Codes: Suppose that 2^{nR} codewords are generated as above for the Shannon channel, but $R > I(X; Y)$. Then there will be exponentially many codewords jointly typical with Y. In fact, the number of codewords on the list associated with Y will be $2^{n(R-I)}$. Thus the cutoff at capacity is very sharp. One goes from one codeword in the inverse fan for $R < I$ to an exponential number of codewords for $R > I$.

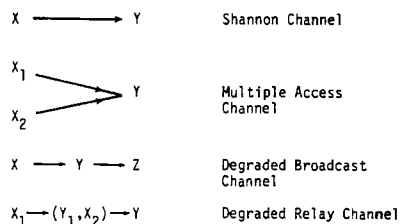*3) The Multiple Access Channel:* Again, random coding works. Fix p_1, p_2. Choose 2^{nR_1} x_1 sequences according to

Fig. 20. Multiple user channels.

Fig. 21. Communication system.

$\Pi p_1(x_{1i})$ and choose 2^{nR_2} x_2 sequences according to $\Pi p_2(x_{2i})$. Now send one of the codewords from the first codebook, and one of the codewords from the second codebook. These two together generate a random response Y drawn according to the conditional probability distribution of Y given the two sequences. With high probability, Y will be jointly typical with these two sequences. If R_1 and R_2 satisfy $R_1 < I(X_1; Y|X_2), R_2 < I(X_2; Y|X_1), R_1 + R_2 < I(X_1, X_2; Y)$ for some $p_1(x_1)p_2(x_2)$, then the receiver Y will see that, although there are exponentially many x_1's jointly typical with Y and exponentially many x_2's jointly typical with Y, there will be only one (x_1, x_2) pair in the codebook that is jointly typical with Y. Thus no error will be made in the decoding.

4) The Degraded Broadcast Channel: We assume that X sends to two receivers Y and Z, where Z is stochastically noisier than Y. Here, the idea of superposition is needed. We first introduce an auxiliary sender U and an associated channel from U to X. Thus the overall channel becomes $U \to X \to Y \to Z$. We now use a random code for the channel $U \to Z$ with 2^{nR_2} codewords u independently drawn according to $\Pi_{i=1}^n p(u_i)$. For each such codeword u, we generate 2^{nR_1} codewords x according to $\Pi_{i=1}^n p(x_i|u_i)$. If $R_2 < I(U; Z)$, then Z perceives u correctly. And if $R_1 < I(X; Y|U)$, then receiver Y perceives x correctly after decoding u. In summary, we send crude codewords to the poor receiver Z, but each of these crude codewords is really made up of small satellite codewords distinguishable by Y.

5) Degraded Relay Channel: Random coding works, but here we also need the Slepian-Wolf theorem. The idea is that we randomly choose codewords which are distinguishable by the relay, but not by the ultimate receiver. Nonetheless, by the argument for list codes, the number of codewords on the ultimate receiver's list is given by $2^{n(R - I(X_1; Y|X_2))}$. Then the relay, in cooperation with the sender, attempts to send the name of the codeword on the (unknown) list of known size $2^{n(R - I(X_1; Y|X_2))}$. He does this by randomly partitioning all the codewords into 2^{nR_0} sets each of equal size. This is the Slepian-Wolf step. He then sends the index of that set in cooperation with the transmitter. They do so by superimposing this information in the manner suggested by the degraded broadcast channel. Yet one more ingredient is necessary. The resolution of the index of the sent codeword on Y's list does not take place until the next block transmission. Thus we have a block Markov encoding of the resolution information, while superimposing the fresh information for the relay.

Summing up, we have the following points.

1) We use a random code to send information to the relay.

2) We use the list coding idea to see how many codewords are left to be resolved by the ultimate receiver Y.

3) We have a block Markov encoding scheme in which the resolution of the information is sent in the next block.

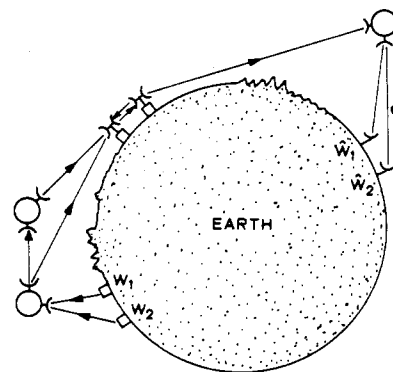4) We use superposition to put cooperative information on top of the new information to the relay.
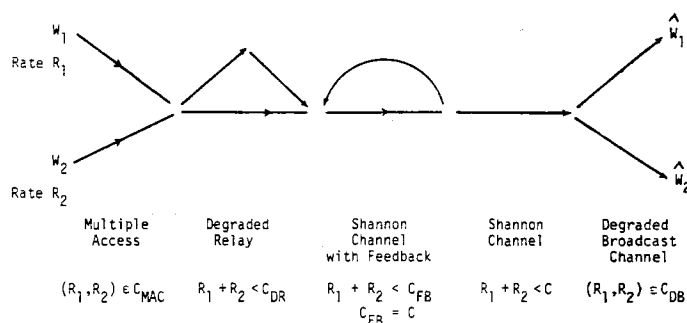

Fig. 22. Communication network with known capacity.

When we put all of these ideas together, we get the achievable rate region for the degraded relay channel. Moreover, a converse, which we do not prove here, shows that this is in fact the capacity for the degraded relay channel.

We can obviously combine these building blocks to solve networks like that in Fig. 21 and schematized in Fig. 22.

We have suppressed the input and output labels at the nodes of this graph. In this network, sender 1 wishes to send an index W_1 at rate R_1 and sender 2 wishes to send an index W_2 at rate R_2 to the receivers respectively designated as $\hat{W}_1$ and $\hat{W}_2$ with overall probability of error $P\{(\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)\}$ tending to zero. This can be accomplished if and only if R_1 and R_2 simultaneously satisfy all of the constraints in Fig. 22. This overall capacity region is the convex subset of the (R_1, R_2) plane given by the intersection of the individual capacity regions. Thus max flow min cut holds for the capacity regions in this network.

Optimism is created each time a new capacity theorem is obtained, but a general theory does not yet exist for information flow in networks. Even for the building blocks of networks, some problems remain open. The multiple access capacity is known, but the capacity of the multiple access channel with feedback is still open. The capacity of the degraded broadcast channel is known, but the capacity of the general broadcast channel remains unknown. The capacity of the general relay channel is still unknown. Progress on Shannon's two-way channel [36] and on the interference channel [37] has just begun. A general theory for networks is years away. Indeed there may not be a mathematically nice answer. Even if there is a nice answer, it may not yield a practical implementation. But in communication theory, as in physics, one must have some faith. If a theory is found, it will have enormous qualitative theoretical implications. Indeed, even if optimal implementations prove to be too complex, the theory will tell the communication designer when

he is close to optimality. Finally, new coding schemes will be suggested from inspection of the proofs. Fortunately, all of the results achieved to date are consistent with an information theoretic description solely in terms of quantities like $I(X_1, X_2; Y_3 | X_3)$, where $I(\cdot; \cdot | \cdot)$ denotes conditional mutual information. Whether this is a result of limited technique or is true of the entire theory is yet to be revealed.

ACKNOWLEDGMENT

The authors are thankful to Chris Heegard for many valuable remarks during the preparation of this paper.

REFERENCES

[1] E. C. van der Meulen, "A survey of multi-way channels in information theory: 1961–1976," *IEEE Trans. Inform. Theory,* vol. IT-23, no. 2, Jan. 1977.

[2] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Inform. Theory* (Tsahkadsor, Armenian S.S.R.), pp. 23–52, 1971. (Publishing House of the Hungarian Academy of Sciences, 1973.)

[3] H. Liao, "A coding theorem for multiple access communications," presented at the Int. Symp. Information Theory, Asilomar, 1972. Also Ph.D. dissertation, "Multiple Access Channels," Dep. Elec. Eng., Univ. Hawaii, Honuolulu, 1972.

[4] L. R. Ford and D. R. Fulkerson, *Flows in Networks.* Princeton, NJ: Princeton Univ. Press, 1962.

[5] A. D. Wyner, "The capacity of the band-limited Gaussian channel," *Bell Syst. Tech. J.,* vol. 45, pp. 359–371, Mar. 1965.

[6] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[7] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty–I," *Bell Syst. Tech. J.,* vol. 40, pp. 43–64. (Also see Landau and Pollak for Parts II and III.)

[8] T. Cover, R. J. McEliece, and E. Posner, "Asymchronous multiple access channel capacity," Tech. Rep. 35, Dep. Statistics, Stanford Univ., Stanford, CA, Dec. 1978. (To appear in *IEEE Trans. Inform. Theory.*)

[9] A. B. Carleial, "A case where interference does not reduce capacity," *IEEE Trans. Inform. Theory,* vol. IT-21, pp. 569–570, Sept. 1975.

[10] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.,* vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.

[11] B. McMillan, "The basic theorems of information theory," *Ann. Math. Stat.,* vol. 24, pp. 196–219, June 1953.

[12] L. Breiman, "The individual ergodic theorem of information theory," *Ann. Math. Stat.,* vol. 28, pp. 809–811, 1960.

[13] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. Inform. Theory,* vol. IT-4, pp. 2–22, Sept. 1954.

[14] G. D. Forney, "Information theory," unpublished course notes.

[15] D. Slepian and J. W. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory,* vol. IT-19, pp. 471–480, July 1973.

[16] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inform. Theory,* vol. IT-22, pp. 226–278, Mar. 1975.

[17] T. Berger, "Multi-terminal source coding," Lecture notes presented at the 1977 CISM Summer School, Udine, Italy, July 18–20, 1977.

[18] T. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," to appear in the *IEEE Trans. Inform. Theory.*

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

[20] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory,* vol. IT-19, pp. 197–207, Mar. 1973.

[21] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredaci Informaccii,* vol. 10, no. 3, pp. 3–14, 1974.

[22] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory,* vol. IT-25, pp. 306–311, May 1979.

[23] P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inform. Theory,* p. 279, Mar. 1974.

[24] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inform. Theory,* vol. IT-23, no. 1, Jan. 1977.

[25] A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Inform. Theory,* vol. IT-25, pp. 166–169, Mar. 1979.

[26] S. Gelfand, "Capacity of one broadcast channel," *Problems of Information Transmission,* vol. 13, no. 3, 1977, pp. 240–242 (English translation).

[27] M. Pinsker, "The capacity region of noiseless broadcast channels," *Problems of Information Transmission,* vol. 14, no. 2, pp. 97–102, 1978 (English translation).

[28] A. El Gamal, "The capacity of the product and sum of two inconsistently degraded broadcast channels," *Problems of Information Transmission,* vol. 1, 1980.

[29] A. El Gamal and E. van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," to appear in *IEEE Trans. on Inform. Theory.*

[30] A. El Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inform. Theory,* vol. IT-24, no. 3, May 1978.

[31] L. Ozarow, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, June 1979.

[32] G. Dueck, "Semi-feedback for multiple-user channels," submitted for publication.

[33] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.,* vol. 3, pp. 120–154, 1971.

[34] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory,* vol. IT-25, no. 5, Sept. 1979.

[35] A. El Gamal, "The capacity of the deterministic relay channel," presented at the 1979 IEEE Int. Symp. Information Theory.

[36] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Stat. Prob.,* 1, Univ. California Press, pp. 611–644, 1961.

[37] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Ann. Probability,* vol. 2, pp. 805–814, Oct. 1974.

BIBLIOGRAPHY

This bibliography of multiple user information theory papers since 1975 is not purported to be complete. Material in this area through 1975 is thoroughly characterized in van der Meulen's 1977 survey paper.

R. Ahlswede and G. Dueck, "The best known codes are highly probable and can be produced by a few permutations," submitted to *IEEE Trans. Inform. Theory.*

R. Ahlswede and T. S. Han, "On source coding with side information via a multiple-access channel and related problems," submitted to *IEEE Trans. Inform. Theory.*

R. Ahlswede and I. Csiszar, "To get a bit of information may be as hard as to get full information," submitted to *IEEE Trans. Inform. Theory.*

R. Ahlswede, "A method of coding and its application to arbitrarily varying channels," to appear in *J. Comb., Inform. Syst. Sci.*

——, "An elementary proof of the strong converse theorem for the multiple-access channel," submitted to *J. Comb., Inform. Syst. Sci.,* 1980.

——, "Coloring hypergraphs: A new approach to multi-user source coding, Part I," *J. Comb., Inform. Syst. Sci.,* vol. 4, no. 1, pp. 76–115, 1979. Part II, to appear.

——, "Elimination of correlation in random codes for arbitrarily varying channel," *Z. Wahrscheinlichkeitstheorie u. Verw. Geb.,* vol. 44, pp. 159–175, 1978.

R. Ahlswede and G. Dueck, "Every bad code has a good subcode: A local converse to the coding theorem," *Z. Wahrscheinlichkeitstheorie u. Verw. Geb.,* vol. 34, pp. 179–182, 1976.

R. Ahlswede and P. Gacs, "Spreading of sets in product spaces and hypercontraction of the Markov operator," *Ann. Probability,* vol. 4, no. 6, pp. 925–939, 1976.

R. Ahlswede, P. Gacs, and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Z. Wahrscheinlichkeitstheorie u, Verw. Geb.,* vol. 34, pp. 157–177, 1976.

R. Ahlswede and P. Gacs, "Two contributions to information theory," *Colloquia Mathematica Societatis Janos Bolyai,* vol. 16. (Topics in Information Theory, Keszthely (Hungary), 1975.)

R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channel," *IEEE Trans. Inform. Theory,* vol. IT-21, pp. 629–637, Nov. 1975.

——, "On common information and related characteristics of correlated information sources," presented at the 7th Prague Conf. Information Theory, Stat. Dec. Fct's, and Rand. Proc., Sept. 1974, to be published in the *IEEE Trans. Inform. Theory.*

M. Aref, "Information flow in relay networks," Ph.D. dissertation, Stanford Univ., Stanford, CA, Oct. 1980.

T. Basar and T. Basar, "Optimum linear coding in continuous-time communication systems with noisy side information at the decoder," *IEEE Trans. Inform. Theory,* vol. IT-25, pp. 612–616, Sept. 1979.

R. Benzel, "The capacity region of a class of discrete additive degraded

interference channels," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 228-231, Mar. 1979.

T. Berger, K. Housewright, J. Omura, S. Tung, and J. Wolfowitz, "An upper bound on the rate distortion function for source coding with partial side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 25, pp. 664-666, Nov. 1979.

T. Berger, "Multiterminal source coding," *The Information Theory Approach to Communications*, G. Longo, Ed. (CISM Courses and Lectures, No. 229), New York: Springer Verlag, 1978, pp. 171-231.

T. Berger and M. Chang, "Rate-distortion with a fully informed decoder and partially informed encoder," *Abstract, 1977 IEEE Int. Symp. Information Theory* (Cornell Univ., Ithaca, NY, Oct. 1977).

T. Berger and S. Tung, "Encoding of correlated analog sources," in *Proc. 1975 IEEE-USSR Joint Workshop on Information Theory* (December 1975, Moscow, USSR), Piscataway, NJ: IEEE Press, pp. 7-10.

P. Bergmans, "The Gaussian interference channel," presented at the 1976 IEEE Int. Symp. Information Theory, Ronneby, Sweden, June 1976.

M. Bierbaum and H. Wallmeier, "A note on the capacity region of the multiple-access channel," *IEEE Trans. Inform. Theory*, vol. IT-25, p. 484, July 1979.

A. Carleial, "Interference channels," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 60-70, Jan. 1978.

A. Carleial and M. Hellman, "A note on Wyner's wiretap channel," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 387-390, May 1977.

A. Carleial, "A case where interference does not reduce capacity," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 569-570, Sept. 1975.

S. Chang and E. Weldon, "Coding for T-user multiple-access channels," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 684-691, Nov. 1979.

T. Cover, R. McEliece, and E. Posner, "Asynchronous multiple access channel capacity," to be published in *IEEE Trans. Inform. Theory*, July 1981.

T. Cover and S. K. Leung, "A rate region for the multiple access channel with feedback," to be published in *IEEE Trans. Inform. Theory*, Mar. 1981.

T. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inform. Theory*, vol. IT-26, Nov. 1980.

T. Cover and A. El Gamal, "Capacity Theorems for the Relay Channel," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 572-584, Sept. 1979.

T. Cover, "Some advances in broadcast channels," ch. 4 in *Advances in Communication Systems, Vol. 4, Theory and Applications*, A. Viterbi, Ed. New York: Academic Press, 1975.

——, "An achievable rate region for the broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 399-404, July 1975.

——, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 2, pp. 226-228, Mar. 1975. Reprinted in *Ergodic and Information Theory*, L. Davisson and R. Gray, Eds., (Benchmark Papers in Electrical Engineering and Computer Science). Stroudsburg, PA: Dowden, Hutchinson, and Ross, Part V, pp. 305-307.

I. Csiszar and J. Körner, "Towards a general theory of source networks," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 155-166, Mar. 1980.

——, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, to be published in 1981.

——, "Graph decomposition: A new key to coding theorems," in *Preprint of Math. Inst. Hungarian Acad. Sci.* (Budapest), 1979.

——, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 339-348, May 1978.

——, "Source networks with unauthorized users," *J. Combinatorics, Inform., and Syst. Sci.*, vol. 1, no. 1, pp. 25-40, 1976.

M. Deaett and J. Wolf, "Some very simple codes for the nonsynchronized two-user multiple access adder channel with binary inputs," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 635-636, Sept. 1978.

C. Downey and J. Karlof, "Group codes for the Gaussian broadcast channel with two receivers," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 406-412, July 1980.

G. Dueck, "Semi-feedback for multiple-user channels," *Preprint*. Submitted for publication.

——, "The strong converse of the coding theorem for the multiple access channel," to be published in *J. Comb., Inform. Syst. Sci.*

——, "Partial feedback for two-way and broadcast channels," to be published in *Inform. Contr.*

——, "A note on the multiple access channel with correlated sources," to be published in *IEEE Trans. Inform. Theory*, vol. IT-27, Mar. 1981.

——, "The capacity region of the two-way channel can exceed the inner bound," *Inform. Contr.*, vol. 40, pp. 258-266, 1979.

——, "Maximal error capacity regions are smaller than average error capacity regions for multi-user channels," *Prob. Cont. Inform. Theory*, vol. 7, pp. 11-19, 1978.

A. El Gamal and T. Cover, "Information theory of joint descriptions," submitted to *IEEE Trans. Inform. Theory*.

A. El Gamal, "The capacity of the deterministic relay channel," to be published in *IEEE Trans. Inform. Theory*.

——, "The capacity of the physically degraded Gaussian broadcast channel with feedback," to be published in *IEEE Trans. Inform. Theory*.

A. El Gamal and E. van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," to be published in *IEEE Trans. Inform. Theory*, vol. IT-27, Jan. 1981.

A. El Gamal, "The capacity of the product and sum of two inconsistently degraded broadcast channels," *Probl. Peredachi Inform.*, vol. 1, 1980.

——, "The capacity of a class of broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 166-169, Mar. 1979.

——, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 379-381, May 1978.

A. El Gamal and C. Heegard, "On the capacity of a computer memory with defects and noise," submitted for publication in *IEEE Trans. Inform. Theory*.

N. Gaarder and J. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 100-102, Jan. 1975.

S. Gelfand and M. Pinsker, "Coding for channel with random parameters," *Probl. Contr. Inform. Theory*, vol. 9, no. 1, pp. 19-31, 1980.

——, "Capacity of a broadcast channel with one deterministic component," *Probl. Inform. Transmission*, vol. 16, no. 1, 1980.

S. Gelfand, "Capacity of one broadcast channel," *Probl. Inform. Transmission*, vol. 13, no. 3, pp. 240-242 (English transl.), 1977.

R. Gray and P. Shields, "The maximum mutual information between two random processes," *Inform. Contr.*, vol. 33, no. 4, pp. 273-280, Apr. 1977.

B. Hajek and M. Pursley, "Evaluation of an achievable rate region for the broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 36-46, Jan. 1979.

T. S. Han, "Slepian-Wolf-Cover theorem for networks of channels," Res. Memo. RMI 79-1, Univ. Tokyo. (Submitted for publication to *Inform. Contr.*)

T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel and its evaluation," Res. Memo. RMI 79-9, Univ. Tokyo. (To be published in *IEEE Trans. Inform. Theory*, vol. IT-27, Jan. 1981.)

T. S. Han, "Deterministic broadcast channels with a common message," *IEEE Trans. Inform. Theory*, vol. IT-26, Nov. 1980.

T. S. Han and R. Ahlswede, "On source coding with side information via a multiple-access channel and related problems," submitted to *IEEE Trans. Inform. Theory*.

T. S. Han and K. Kobayashi, "An achievable rate region for noiseless source coding in a general multiterminal situation," *IEEE Trans. Inform. Theory*, vol. IT-26, May 1980.

T. S. Han, "Source coding with cross observations at the encoders," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 360-361, May 1979.

——, "The capacity region of general multiple-access channel with certain correlated sources," *Inform. Contr.*, vol. 40, no. 1, pp. 37-60, Jan. 1979.

C. Heegard, H. DePedro, and J. Wolf, "Permutation codes for the Gaussian broadcast channel with two receivers," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 5, pp. 569-578, Sept. 1978.

G. D. Hu and S. Y. Shen, "Some new results of Shannon theory," Nankai University, Tianjin, China, Mar. 1980.

D. Hughes-Hartogs, "The capacity of the degraded spectral Gaussian broadcast channel," Tech. Rep. 7002-2, Information Systems Lab., Stanford Univ., Stanford, CA. Also Ph.D. dissertation, July 1975.

K. Jahn, "Coding of arbitrarily varying multiuser channels," Stanford Univ., Stanford, CA, Tech. Rep., 1979.

——, "Coding with side information and a rate-distortion theorem for arbitrarily varying correlated sources," Stanford Univ., Stanford, CA, Tech. Rep., 1979.

K. Jarett and T. Cover, "Asymmetries in relativistic information flow," to be published in *IEEE Trans. Inform. Theory*, vol. IT-27, 1981.

R. Kahn and M. Hellman, "On the wiretap channel with feedback," to be published in *IEEE Trans. Inform. Theory*.

T. Kasami and S. Lin, "Decoding of linear δ-decodable codes for a multiple access channel," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 633-635, Sept. 1978.

——, "Bounds on the achievable rates of block coding for a memoryless multiple access channel," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 187-197, Mar. 1978.

——, "Coding for a multiple access channel," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 129-137, Mar. 1976.

A. Kaspi and T. Berger, "Rate-distortion for correlated sources with partially separated encoders," submitted to *IEEE Trans. Inform. Theory*.

C. Keilers, "The capacity of the spectral Gaussian multiple-access channel," Ph.D. dissertation Dep. Elec. Eng., Stanford Univ.,

Stanford, CA, Apr. 1976.

J. Kieffer, M. Pursley, M. Wallace, and J. Wolfowitz, "Coding for correlated sources with unknown parameters," in *Proc. 1980 Conf. Information Sciences and Systems* (Princeton, NJ). pp. 168–171, Mar. 1980.

R. King, "Multiple access channels with generalized feedback," Ph.D. dissertation, Stanford Univ., Stanford, CA, ISL Rep. 6152-1, Mar. 1978.

J. Körner, "Source networks and new information quantities," in *Coll. Int. CNRS*, no. 276 (Theorie de l'Information), pp. 235–242, Paris, France, 1977.

——, "Some methods in multi-user communication," in *Information Theory, New Trends and Open Problems*, G. Longo, Ed. Vienna, Austria: Springer Verlag, 1976, pp. 172–176.

J. Körner and K. Marton, "How to encode the mod 2 sum of two binary sources," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 219–221, Mar. 1979.

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

——, "Comparison of two noisy channels," *Coll. Math. Soc. F. Boliya*, vol. 16, in *Topics in Information Theory*, Amsterdam, The Netherlands: North Holland, 1977, pp. 411–423.

——, "General broadcast channels with degraded message sets," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 60–64, Jan. 1977.

J. Körner and A. Sgarro, "Universally attainable error exponents for broadcast channels with degraded message sets," *IEEE Trans. Inform. Theory*, vol. IT-25, Nov. 1980.

V. Koshelev, "On a problem of separate coding for two dependent sources," *Probl. Trans. Inform.*, vol. 13, pp. 26–32, Jan.-Mar. 1977.

S. K. L-Y. Cheong and M. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 451–456, July 1978.

S. Lu, "On secrecy systems with side information about the message available to a cryptanalyst," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 472–475, July 1979.

K. Mackenthun, "An evaluation of the Wyner–Ziv rate region," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 347–354, May 1980.

K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 306–311, May 1979.

——, "The capacity region of deterministic broadcast channels," in *Trans. Int. Symp. Inform. Theory* (Paris-Cachan, France) 1977.

L. Ozarow, "Coding and capacity for additive white Gaussian noise multi-user channels with feedback," Ph.D. dissertation, Dep. Elec. Eng. and Comput. Sci., M.I.T., Cambridge, MA.

R. Peterson and D. Costello, "Comments on 'Convolutional tree codes for multiple access channels,' " *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 627–628, Sept. 1980.

M. Pinsker, "The capacity region of noiseless broadcast channels," *Probl. Inform. Transmission*, vol. IT-14, pp. 97–102, 1978 (English transl.).

G. Sh. Poltyrev, "Capacity of parallel broadcast channels with degraded subchannels," *Probl. Trans. Inform.*, vol. 13, no. 2, pp. 23–36, Apr.-June 1977 (English transl.).

——, "Capacity for a sum of broadcast channels," *Probl. Inform. Trans*, vol. 15, no. 2, pp. 111–114, Apr.-June 1979 (English transl.).

H. Sato, "A study on Gaussian two-user channel," submitted to *IEEE Trans. Inform. Theory*.

——, "The capacity region of two-user channels—Method of extension channels," in *Abstract Proc. 1979 IEEE Int. Symp. Inform. Theory*, July 1979.

H. Sato and M. Tanabe, "A discrete two-user channel with strong interference," *Trans. IECE Japan*, vol. E61, no. 11, p. 880, Nov. 1978.

H. Sato, "On degraded Gaussian two-user channels," *IEEE Trans.*

Inform. Theory, vol. IT-24, p. 637, Sept. 1978.

——, "An outer bound to the capacity region of broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 3, p. 374, May 1978.

——, "On the capacity region of a discrete two-user channel for strong interference," *IEEE Trans. Inform. Theory*, vol. IT-24, p. 377, May 1978.

——, "Two-user communication channels," *IEEE Trans. Inform. Theory*, vol. IT-23, p. 295, May 1977.

——, "Information transmission through a channel with relay," Tech. Report ALOHA SYSTEM B-76-7, Mar. 1976.

M. Ulrey, "The capacity region of a channel with s senders and r receivers," *Inform. Contr.*, vol. 29, no. 3, pp. 185–203, Nov. 1975.

E. C. van der Meulen, "A survey of multi-way channels in information theory: 1961-1976," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 1–37, Jan. 1977.

——, "Random coding theorems for the general discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 180–190, Mar. 1975.

E. Verriest and M. Hellman, "Convolutional encoding for Wyner's wiretap channel," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 234–236, Mar. 1979.

J. Willems, "The feedback capacity region of a class of discrete memoryless multiple access channels," *Preprint*, Univ. Leuven, Belgium, 1980.

H. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. Inform. Theory*, vol. IT-26, Sept. 1980.

——, "On source networks with minimal breakdown degradation," *Bell Syst. Tech. J.*, vol. 59, pp. 1083–1087, July-Aug. 1980.

——, "Some aspects of convexity useful in information theory," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 265–271, May 1980.

——, "The zero-error side-information problem and chromatic numbers," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 592–593, Sept. 1976.

——, "Values and bounds for the common information of two discrete random variables," *SIAM J. Appl. Math.*, vol. 31, pp. 313–333, Sept. 1976.

H. Witsenhausen and A. Wyner, "A conditional entropy bound for a pair of discrete random variables," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 493–501, Sept. 1975.

H. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM J. Appl. Math.*, vol. 28, pp. 100–113, Jan. 1975.

J. Wolf, A. Wyner, and J. Ziv, "Source coding for multiple descriptions," *Bell Syst. Tech. J.*, vol. 59, Oct. 1980.

J. Wolf, "Multi-user communication networks," in *Commun. Syst. and Random Process Theory*, J. K. Skwirzynski, Ed. Alphen aaan den Rijn-The Netherlands: Sijthoff & Noordhoff, 1978, pp. 37–53.

——, "The AEP property of random sequences and applications to information theory: Part I basic principles," *Inform. Theory, New Trends and Open Problems*, G. Longo, Ed., (CISM Courses and Lectures No. 219). New York: Springer-Verlag, 1975.

J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. New York: Springer-Verlag, 1978.

A. Wyner, "The rate-distortion function for source coding with side-information at the Decoder II: General sources," *Inform. Contr.*, vol. 38, pp. 60–80, July 1980.

A. Wyner and J. Ziv, "The rate-distortion function for source coding with side-information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 1–10, Jan. 1976.

A. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, pp. 1355–1387, Oct. 1975.

——, "On source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 294–300, May 1975.

——, "The common information of two dependent random variables," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 163–170, Mar. 1975.