ACKNOWLEDGMENT

In conclusion, the author wishes to express his grateful indebtedness to Professors Solomon Golomb and Albert Whiteman for their guidance and encouragement.

REFERENCES

[1] J. J. Sylvester, *Mathematical Questions from the Educational Times (London)*, vol. X, p. 7406, 1868; vol. LVI, pp. 97–99, 1892; cf. W. W. Rouse Ball (revised by H. S. M. Coxeter), *Mathematical Recreations and Essays*, American ed., New York: MacMillan, 1947, pp. 108–111.

[2] J. Hadamard, "Résolution d'une question relative aux déterminants," *Bull. Sci. Math. (2)*, part 1, vol. 17, pp. 240–246.

[3] R. E. A. C. Paley, "On orthogonal matrices," *J. Math. and Phys.*, vol. 12, pp. 311–320, 1933.

[4] J. Williamson, "Hadamard's determinant theorem and the sum of four squares," *Duke Math. J.*, vol. 11, pp. 65–81, 1944.

[5] S. W. Golomb and L. D. Baumert, "The search for Hadamard matrices," *Amer. Math. Mon.*, vol. 70, pp. 12–17, 1963.

[6] A. L. Whiteman, "An infinite family of skew Hadamard matrices," *Pac. J. Math.*, vol. 38, pp. 817–822, 1971.

[7] A. L. Whiteman, "An infinite family of Hadamard matrices of the Williamson type," *J. Combinat. Theor., Ser. A*, vol. 14, pp. 334–340, 1973.

[8] R. T. Chien, "A class of optimal noiseless load-sharing matrix switches," *IBM J. Res. Develop.*, vol. 4, pp. 414–417, 1960.

[9] W. K. Pratt, J. Kane, and H. C. Andrews, "Hadamard transform image coding," *Proc. IEEE*, vol. 57, pp. 58–68, 1969.

[10] J. E. Whelchel and D. F. Guinn, "The fast Fourier-Hadamard transform and its use in signal representation and classification," *1968 EASCOM Rec.*, pp. 561–573.

[11] J. A. Decker, Jr., "Hadamard transform spectrometry: a new analytical tool," *Anal. Chem.*, vol. 44, pp. 127A–134A, 1973.

[12] R. O. Rowlands, "Might Walsh functions solve your problem?" *European Scientific Notes*, vol. ESN-25-8, London, Office of Naval Research, pp. 261–262, Aug. 1971.

[13] P. J. Shlichta, "Three- and four dimensional Hadamard matrices," *Bull. Amer Phys. Soc.*, ser. 11, vol. 16, pp. 825–826, 1971.

[14] G. A. Korn and T. A. Korn, in *Mathematical Handbook for Scientists and Engineers*, New York, McGraw-Hill, 1961, sect. 13.2-11, p. 361.

[15] H. C. Andrews, *private communication*.

[16] H. S. M. Coxeter, in *Regular Polytopes*, second ed. New York, MacMillan, 1963, p. 24 and pp. 223–225.

[17] A. Heydayat and W. D. Wallis, "Hadamard matrices and their applications," *Annals of Statistics*, vol. 6, pp. 1184–1238, 1978.

Capacity Theorems for the Relay Channel

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

***Abstract*—A relay channel consists of an input x_1, a relay output y_1, a channel output y, and a relay sender x_2 (whose trasmission is allowed to depend on the past symbols y_1). The dependence of the received symbols upon the inputs is given by $p(y,y_1|x_1,x_2)$. The channel is assumed to be memoryless. In this paper the following capacity theorems are proved.**

1) If y is a degraded form of y_1, then
$C = \max_{p(x_1,x_2)} \min\{I(X_1,X_2;Y), I(X_1;Y_1|X_2)\}$.

2) If y_1 is a degraded form of y, then
$C = \max_{p(x_1)} \max_{x_2} I(X_1;Y|x_2)$.

3) If $p(y,y_1|x_1,x_2)$ is an arbitrary relay channel with feedback from (y,y_1) to both x_1 and x_2, then
$C = \max_{p(x_1,x_2)} \min\{I(X_1,X_2;Y), I(X_1;Y,Y_1|X_2)\}$.

4) For a general relay channel,
$C \leq \max_{p(x_1,x_2)} \min\{I(X_1,X_2;Y), I(X_1;Y,Y_1|X_2)\}$.

Superposition block Markov encoding is used to show achievability of C, and converses are established. The capacities of the Gaussian relay channel and certain discrete relay channels are evaluated. Finally, an achievable lower bound to the capacity of the general relay channel is established.

Manuscript received December 1, 1977; revised September 28, 1978. This work was partially supported by the National Science Foundation under Grant ENG76-03684, JSEP N00016-67-A-0012-0601, and Stanford Research Institute Contract DAHC-15-C-0187. This work was presented at the International Symposium on Information Theory, Grigano, Italy, June 25–27, 1979.

T. M. Cover is with the Departments of Electrical Engineering and Statistics, Stanford University, Stanford, CA 94305.

A. A. El Gamal is with the Department of Electrical Engineering Systems, University of Southern California, University Park, Los Angeles, CA 90007.

I. INTRODUCTION

THE DISCRETE memoryless relay channel denoted by $(\mathcal{X}_1 \times \mathcal{X}_2, p(y,y_1|x_1,x_2), \mathcal{Y} \times \mathcal{Y}_1)$ consists of four finite sets: $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{Y}$, $\mathcal{Y}_1$, and a collection of probability distributions $p(\cdot,\cdot|x_1,x_2)$ on $\mathcal{Y} \times \mathcal{Y}_1$, one for each $(x_1,x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$. 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. 1. 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 [1], [2], [3, p. 7 and pp. 32–34], and has also been studied by Sato [4]. In [1] a timesharing approach was used to find inner bounds for C. Outer bounds were found in [1] and [4]. However, C was established only for relatively degenerate channels.

The model that motivates our investigation of degraded relay channels is perhaps best illustrated in the Gaussian case (see Fig. 3 and the example in Section IV). Suppose the transmitter x_1 has power P_1 and the relay transmitter has power P_2. The relay receiver y_1 sees x_1+z_1, $z_1 \sim N(0,N_1)$. The intended receiver y sees the sum of the relay signal x_2 and a corrupted version of y_1, i.e., $y = x_2 + y_1 + z_2$, $z_2 \sim N(0,N_2)$. How should x_2 use his knowledge of x_1 (obtained through y_1) to help y understand x_1? We shall

0018-9448/79/0900-0572$00.75

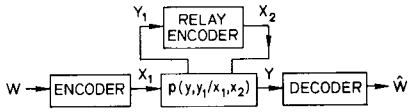

Fig.1. Relay channel.

show that the capacity is given by

$$C^* = \max_{0 \le \alpha \le 1} \min\left\{ C\left(\frac{P_1 + P_2 + 2\sqrt{\bar{\alpha} P_1 P_2}}{N_1 + N_2}\right), C\left(\frac{\alpha P}{N_1}\right)\right\}$$

where $C(x) = (1/2)\log(1+x)$. An interpretation consistent with achieving C^* in this example is that y_1 discovers x_1 perfectly, then x_2 and x_1 cooperate *coherently* in the next block to resolve the remaining y uncertainty about x_1. However, in this next block, fresh x_1 information is superimposed, thus resulting in a steady-state resolution of the past uncertainty and infusion of new information.

An (M,n) code for the relay channel consists of a set of integers

$$\mathcal{M} = \{1,2,\cdots,M\} \triangleq [1,M] \tag{1}$$

an encoding function

$$x_1 : \mathcal{M} \to \mathcal{X}_1^n \tag{2}$$

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

$$x_{2i} = f_i(Y_{11}, Y_{12}, \cdots, Y_{1i-1}), \qquad 1 \le i \le n, \tag{3}$$

and a decoding function

$$g : \mathcal{Y}^n \to \mathcal{M}. \tag{4}$$

For generality, the encoding functions $x_1(\cdot), f_i(\cdot)$ and decoding function $g(\cdot)$ are allowed to be stochastic functions.

Note that the allowed relay 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 $y_1^i = (y_{11}, y_{12}, \cdots, y_{1i-1})$. This is the definition used by van der Meulen [1]. The channel is memoryless in the sense that (y_i, y_{1i}) depends on the past (x_1^i, x_2^i) only through the current transmitted symbols (x_{1i}, x_{2i}). Thus, for any choice $p(w)$, $w \in M$, and code choice $x_1 : \mathcal{M} \to \mathcal{X}_1^n$ and relay functions $\{f_i\}_{i=1}^n$, the joint probability mass function on $\mathcal{M} \times \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}^n \times \mathcal{Y}_1^n$ is given by

$$p(w, \mathbf{x}_1, \mathbf{x}_2, \mathbf{y}, \mathbf{y}_1) = p(w) \prod_{i=1}^{n} p(x_{1i}|w) p(x_{2i}|y_{11}, y_{12}, \cdots, y_{1i-1}) \cdot p(y_i, y_{1i}|x_{1i}, x_{2i}). \tag{5}$$

Remark: Throughout this paper we shall use the notational convenience $p_{V|U}(v|u) = p(v|u)$, where the dropped subscripts will be obvious by inspection of the arguments of the function. We shall also write $X \sim p(x)$ to indicate that the random variable X is drawn according to the probability mass function $p(x)$.

If the message $w \in \mathcal{M}$ is sent, let

$$\lambda(w) = \Pr\{ g(Y) \neq w \} \tag{6}$$

denote the conditional probability of error. We define the *average probability of error* of the code to be

$$\bar{P}_n(e) = \frac{1}{M} \sum_w \lambda(w). \tag{7a}$$

The probability of error is calculated under a special distribution— the uniform distribution over the codewords $w \in [1,M]$. Finally, let

$$\lambda_n = \max_{w \in \mathcal{M}} \lambda(w) \tag{7b}$$

be the *maximal probability of error* for the (M,n) code.

The *rate* R of an (M,n) code is defined by

$$R = \frac{1}{n} \log M \qquad \text{bits/transmission}. \tag{8}$$

The rate R is said to be *achievable* by a relay channel if, for any $\epsilon > 0$ and for all n sufficiently large, there exists an (M,n) code with

$$M \geqslant 2^{nR} \tag{9}$$

such that $\lambda_n < \epsilon$. The *capacity* C of the relay channel is the supremum of the set of achievable rates.

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.

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

$$p(y,y_1|x_1,x_2) = p(y_1|x_1,x_2) p(y|y_1,x_2). \tag{10}$$

Equivalently, we see by inspection of (10) that a relay channel is degraded if $p(y|x_1,x_2,y_1) = p(y|x_2,y_1)$, i.e., $X_1 \to (X_2, Y_1) \to Y$ form a Markov chain. The previously discussed Gaussian channel is therefore degraded. For the reader familiar with the definition of the degraded broadcast channel, we observe that a degraded relay channel can be looked at as a family of *physically* degraded broadcast channels indexed by x_2. A weaker form of degradation (stochastic) can be defined for relay channels, but Theorem 1 below then becomes only an inner bound to the capacity. The case in which the relay y_1 is worse than y is less interesting (except for the converse) and is defined as follows.

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

$$p(y,y_1|x_1,x_2) = p(y|x_1,x_2) p(y_1|y,x_2). \tag{11}$$

The main contribution of this paper is summarized by the following three theorems.

Theorem 1: The capacity C of the degraded relay channel is given by

$$C = \sup_{p(x_1,x_2)} \min\{I(X_1,X_2;Y), I(X_1;Y_1|X_2)\} \tag{12}$$

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

Theorem 2: The capacity C_0 of the reversely degraded relay channel is given by

$$C_0 = \max_{x_2 \in \mathcal{X}_2} \max_{p(x_1)} I(X_1; Y|x_2). \tag{13}$$

Theorem 2 has a simple interpretation. Since the relay y_1 sees a corrupted version of what y sees, x_2 can contribute no new information to y—thus x_2 is set constantly at the symbol that "opens" the channel for the transmission of x_1 directly to y at rate $I(X_1; Y|x_2)$. The converse proves that one can do no better.

Theorem 1 has a more interesting interpretation. The first term in the brackets in (12) suggests that a rate $I(X_1, X_2; Y)$ can be achieved where $p(x_1, x_2)$ is arbitrary. However, this rate can only be achieved by complete cooperation of x_1 and x_2. To set up this cooperation x_2 must know x_1. Thus the x_1 rate of transmission should be less than $I(X_1; Y_1|X_2)$. (How they cooperate given these two conditions will be left to the proof.) Finally, both constraints lead to the minimum characterization in (12).

The obvious notion of an arbitrary relay channel with causal feedback (from both y and y_1 to x_1 and x_2) will be formalized in Section V. The following theorem can then be proved.

Theorem 3: The capacity of C_{FB} of an *arbitrary* relay channel with feedback is given by

$$C_{FB} = \sup_{p(x_1,x_2)} \min\{I(X_1,X_2;Y), I(X_1;Y,Y_1|X_2)\}. \quad (14)$$

Note that C_{FB} is the same as C except that Y_1 is replaced by (Y, Y_1) in $I(X_1; Y_1|X_2)$. The reason is that the feedback changes an arbitrary relay channel into a degraded relay channel in which x_1 transmits information to x_2 by way of y_1 *and* y. Clearly Y is a degraded form of (Y, Y_1).

Theorem 2 is included for reasons of completeness, but it can be shown to follow from a chain of remarks in van der Meulen [1] under slightly stronger conditions. Specifically in [1] see Equation (8.1), Lemma 4.1, Theorem 10.1, and Theorem 7.1.

Before proving the theorems, we apply the result (12) to a simple example introduced by Sato [4]. The channel as shown in Fig. 2 has $\mathcal{X}_1 = \mathcal{Y} = \mathcal{Y}_1 = \{0,1,2\}$, $\mathcal{X}_2 = \{0,1\}$, and the conditional probability $p(y, y_1|x_1, x_2)$ satisfies (10). Specifically, the channel operation is

$$y_1 \equiv x_1 \quad (15a)$$

and

$$p(y|y_1, x_2=0) = \begin{matrix} \\ y_1=0 \\ y_1=1 \\ y_1=2 \end{matrix} \begin{matrix} y=0 & y=1 & y=2 \\ \end{matrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 \end{bmatrix} \quad (15b)$$

$$p(y|y_1, x_2=1) = \begin{matrix} \\ y_1=0 \\ y_1=1 \\ y_1=2 \end{matrix} \begin{matrix} y=0 & y=1 & y=2 \\ \end{matrix} \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (15c)$$

Sato calculated a cooperative upper bound to the capacity of the channel, $R_{UG} = \max_{p(x_1,x_2)} I(X_1, X_2; Y) = 1.170$. By restricting the relay encoding functions to

i) $X_{2i} = f_i(y_{11}, \cdots, y_{1i-1}) = f(y_{1i-1}) \quad 1 < i \leqslant n,$

ii) $X_{2i} = f_i(y_{11}, \cdots, y_{1i-1}) = f(y_{1i-2}, y_{1i-1}) \quad 1 < i \leqslant n.$

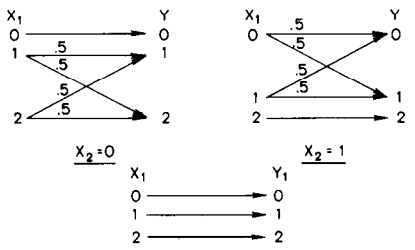

Fig. 2. Ternary relay channel.

TABLE I
OPTIMAL $p(x_1, x_2)$ FOR THE TERNARY RELAY CHANNEL.

	$x_1 = 0$	$x_1 = 1$	$x_1 = 2$
$x_2 = 0$	.35431	.072845	.072845
$x_2 = 1$	.072845	.072845	.35431

Sato calculated two lower bounds to the capacity of the channel:

i) $R_1 = 1.0437$

ii) $R_2 = 1.0549.$

From Theorem 1 we obtain the true capacity $C = 1.161878\cdots$. The optimal joint distribution on $\mathcal{X}_1 \times \mathcal{X}_2$ is given in Table I.

We shall see that instead of letting the encoding functions of the relay depend only on a finite number of previous y_1 transmissions, we can achieve C by allowing block Markovian dependence of x_2 and y_1 in a manner similar to [6].

II. ACHIEVABILITY OF C IN THEOREMS 1,2,3

The achievability of $C_O = \sup_{p(x_1)} \max_{x_2} I(X_1; Y|x_2)$ in Theorem 2 follows immediately from Shannon's basic result [5] if we set $X_{2i} = x_2$, $i = 1, 2, \cdots$. Also, the achievability of C_{FB} in Theorem 3 is a simple corollary of Theorem 1, when it is realized that the feedback relay channel is a degraded relay channel. The converses will be delayed until Section III.

We are left only with the proof of Theorem 1—the achievability of C for the degraded relay channel. We begin with a brief outline of the proof. We consider B blocks, each of n symbols. A sequence of $B-1$ messages $w_i \in [1, 2^{nR}]$, $i = 1, 2, \cdots, B-1$ will be sent over the channel in nB tranmissions. (Note that as $B \to \infty$, for fixed n, 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

$$\mathcal{C} = \{x_1(w_b|s_b), x_2(s_b)\}; \quad w_b \in [1, 2^{nR}], \; s_b \in [1, 2^{nR_0}], \quad x_1(\cdot|\cdot) \in \mathcal{X}_1^n, \; x_2(\cdot) \in \mathcal{X}_2^n. \quad (16)$$

We shall also need a partition

$$\mathcal{S} = \{S_1, S_2, \cdots, S_{2^{nR_0}}\} \text{ of } \mathcal{M} = \{1, 2, \cdots, 2^{nR}\} \text{ into } 2^{nR_0} \text{ cells, } S_i \cap S_j = \varnothing, \; i \neq j, \; \cup S_i = \mathcal{M}. \quad (17)$$

The partition $\mathcal{S}$ will allow us to send information to the receiver using the random binning proof of the source coding theorem of Seplian and Wolf [7].

The choice of $\mathcal{C}$ and $\mathcal{S}$ achieving C will be random, but the description of the random code and partition will be delayed until the use of the code is described. For the time being, the code should be assumed fixed.

We pick up the story in block $i-1$. First, let us assume that the receiver y knows w_{i-2} and s_{i-1} at the end of block $i-1$. Let us also assume that the relay receiver knows w_{i-1}. We shall show that a good choice of $\{\mathcal{C}, \mathcal{S}\}$ will allow the receiver to know (w_{i-1}, s_i) and the relay receiver to know w_i at the end of block i (with probability of error $\leqslant \epsilon$). Thus the information state (w_{i-1}, s_i) of the receiver propagates forward, and a recursive calculation of the probability of error can be made, yielding probability of error $\leqslant B\epsilon$.

We summarize the use of the code as follows.

Transmission in block $i-1$: $x_1(w_{i-1}|s_{i-1})$, $x_2(s_{i-1})$.

Received signals in block $i-1$: $Y_1(i-1)$, $Y(i-1)$.

Computation at the end of block $i-1$: the relay receiver $Y_1(i-1)$ is assumed to know w_{i-1}. The integer w_{i-1} falls in some cell of the partition $\mathcal{S}$. Call the index of this cell s_i. Then the relay is prepared to send $x_2(s_i)$ in block i. Transmitter x_1 also computes s_i from w_{i-1}. Thus s_i will furnish the basis for cooperative resolution of the y uncertainty about w_{i-1}.

Remark: In the first block, the relay has no information s_1 necessary for cooperation. However any good sequence x_2 will allow the block Markov scheme to get started, and the slight loss in rate in the first block becomes asymptotically negligible as the number of blocks $B \to \infty$.

Transmission in block i: $x_1(w_i|s_i)$, $x_2(s_i)$.

Received signals in block i: $y_1(i)$, $y(i)$.

Computation at end of block i: 1) The relay calculates w_i from $y_1(i)$. 2) The unique jointly typical $x_2(s_i)$ with the received $y(i)$ is calculated. Thus s_i is known to the receiver. 3) The receiver calculates his ambiguity set $\mathcal{L}(y(i-1))$, i.e., the set of all w_{i-1} such that $(x_1(w_{i-1}|s_{i-1}), x_2(s_{i-1}), y(i-1))$ are jointly ϵ-typical.

The receiver then intersects $\mathcal{L}(y(i-1))$ and the cell S_{s_i}. By controlling the size of $\mathcal{L}$, we shall $(1-\epsilon)$-guarantee that this intersection has one and only one member—the correct value w_{i-1}. We conclude that the receiver $y(i)$ has correctly computed (w_{i-1}, s_i) from (w_{i-2}, s_{i-1}) and $(y(i-1), y(i))$.

Proof of achievability of C in Theorem 1: We shall use the code as outlined previously in this section. It is important to note that, although Theorem 1 treats degraded relay channels, the proof of achievability of C and all constructions in this section apply without change to arbitrary relay channels. It is only in the converse that degradedness is needed to establish that the achievable rate C is indeed the capacity. The converse is proved in Section III.

We shall now describe the random codes. Fix a probability mass function $p(x_1, x_2)$.

Random Coding: First generate at random $M_0 = 2^{nR_0}$ independent identically distributed n-sequences in $\mathcal{X}_2^n$, 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 $M = 2^{nR}$ 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 code book $\mathcal{C} = \{x_1(w|s), x_2(s)\}$.

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 cell S_s. We shall use the functional notation $s(w)$ to denote the index of the cell in which w lies.

Typical Sequences: We recall some basic results concerning typical sequences. Let $\{X^{(1)}, X^{(2)}, \cdots, X^{(k)}\}$ denote a finite collection of discrete random variables with some fixed joint distribution $p(x^{(1)}, x^{(2)}, \cdots, x^{(k)})$, for $(x^{(1)}, x^{(2)}, \cdots, x^{(k)}) \in \mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \cdots \times \mathcal{X}^{(k)}$. Let S denote an ordered subset of these random variables, and consider n independent copies of S. Thus

$$\Pr\{\boldsymbol{S} = \boldsymbol{s}\} = \prod_{i=1}^{n} \Pr\{S_i = s_i\}, \qquad \boldsymbol{s} \in \mathcal{S}^n. \tag{18}$$

Let $N(s; \boldsymbol{s})$ be the number of indices $i \in \{1, 2, \cdots, n\}$ such that $S_i = s$. By the law of large numbers, for any subset S of random variables and for all $s \in \mathcal{S}$,

$$\frac{1}{n} N(s; \boldsymbol{s}) \to p(s). \tag{19}$$

Also

$$-1/n \log p(s_1, s_2, \cdots, s_n) = -1/n \sum_{i=1}^{n} \log p(s_i) \to H(S). \tag{20}$$

Convergence in (19) and (20) takes place simultaneously with probability one for all 2^k subsets

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

Consider the following definition of joint typicality.

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

$$\begin{aligned} A_\epsilon(X^{(1)}, X^{(2)}, \cdots, X^{(k)}) = A_\epsilon = \Big\{ & (\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(k)}): \\ & \Big| \frac{1}{n} N(x^{(1)}, x^{(2)}, \cdots, x^{(k)}; \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(k)}) \\ & - p(x^{(1)}, x^{(2)}, \cdots, x^{(k)}) \Big| \\ & < \epsilon \| \mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \cdots \times \mathcal{X}^{(k)} \|, \text{ for } (x^{(1)}, x^{(2)}, \cdots, x^{(k)}) \\ & \in \mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \cdots \times \mathcal{X}^{(k)} \Big\} \end{aligned} \tag{21}$$

where $\|\mathcal{A}\|$ is cardinality of the set $\mathcal{A}$.

Remark: This definition of typicality, sometimes called strong typicality, can be found in the work of Wolfowitz [13] and Berger [12]. Strong typicality implies (weak) typicality used in [8], [14]. The distinction is not needed until the proof of Theorem 6 in Section VI of this paper.

The following is a version of the asymptotic equipartition property involving simultaneous constraints [12], [14].

Lemma 1: For any $\epsilon>0$, there exists an integer n such that $A_\epsilon(S)$ satisfies

$$\text{i)}\quad \Pr\{A_\epsilon(S)\}\geqslant 1-\epsilon, \qquad \text{for all } S\subseteq\{X^{(1)},\cdots,X^{(k)}\}$$

$$\text{ii)}\quad s\in A_\epsilon(S)\Rightarrow\left|-\frac{1}{n}\log p(s)-H(S)\right|<\epsilon$$

$$\text{iii)}\quad (1-\epsilon)2^{n(H(S)-\epsilon)}\leqslant\|A_\epsilon(S)\|\leqslant 2^{n(H(S)+\epsilon)}. \tag{22}$$

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 $X^{(1)},X^{(2)},\cdots,X^{(k)}$. Let S_1', S_2' be conditionally independent given S_3, with the marginals

$$p(s_1|s_3)=\sum_{s_2}p(s_1,s_2,s_3)/p(s_3)$$

$$p(s_2|s_3)=\sum_{s_1}p(s_1,s_2,s_3)/p(s_3). \tag{23}$$

The following lemma is proved in [14].

Lemma 2: Let $(\boldsymbol{S}_1,\boldsymbol{S}_2,\boldsymbol{S}_3)\sim\prod_{i=1}^n p(s_{1i},s_{2i},s_{3i})$ and $(\boldsymbol{S}_1',\boldsymbol{S}_2',\boldsymbol{S}_3)\sim\prod_{i=1}^n p(s_{1i}|s_{3i})p(s_{2i}|s_{3i})p(s_{3i})$. Then, for n such that $P\{A_\epsilon(S_1,S_2,S_3)\}\geqslant 1-\epsilon$,

$$(1-\epsilon)2^{-n(I(S_1;S_2|S_3)+7\epsilon)}\leqslant P\{(\boldsymbol{S}_1',\boldsymbol{S}_2',\boldsymbol{S}_3)\in A_\epsilon(S_1,S_2,S_3)\}\leqslant 2^{-n(I(S_1;S_2|S_3)-7\epsilon)}. \tag{24}$$

Encoding: Let $w_i\in[1,2^{nR}]$ be the new index to be sent in block i, and assume that $w_{i-1}\in S_{s_i}$. The encoder then sends $\boldsymbol{x}_1(w_i|s_i)$. 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 $\boldsymbol{x}_2(\hat{s}_i)$ 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 $(s_1,s_2,\cdots,s_i)$.

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

1) Knowing s_i, and upon receiving $\boldsymbol{y}_1(i)$, the *relay* receiver estimates the message of the transmitter $\hat{w}_i=w$ iff there exists a *unique* w such that $(\boldsymbol{x}_1(w|s_i),\boldsymbol{x}_2(s_i),\boldsymbol{y}_1(i))$ are jointly ϵ-typical. Using Lemma 2, it can be shown that $\hat{w}_i=w_i$ with arbitrarily small probability of error if

$$R<I(X_1;Y_1|X_2) \tag{25}$$

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 $(\boldsymbol{x}_2(s),\boldsymbol{Y}(i))$ are jointly ϵ-typical. From Lemma 2 we know that s_i can be decoded with arbitrarily small probability of error if s_i takes on less than $2^{nI(X_2;Y)}$ values, i.e., if

$$R_0<I(X_2;Y) \tag{26}$$

and n is sufficiently large.

3) Assuming that s_i is decoded successfully at the receiver, then $\hat{w}_{i-1}=w$ is declared to be the index sent in block $i-1$ iff there is a unique $w\in S_{s_i}\cap\mathfrak{L}(\boldsymbol{y}(i-1))$. It will be shown that if n is sufficeintly large and if

$$R<I(X_1;Y|X_2)+R_0 \tag{27}$$

then $\hat{w}_{i-1}=w_{i-1}$ with arbitrarily small probability of error.

Thus combining (26) and (27) yields $R<I(X_1,X_2;Y)$, the first term in the capacity expression in Theorem 1. The second term is given by constraint (25).

Calculation of Probability of Error: For the above scheme, we will declare an error in block i if ine or more of the following events occurs.

E_{0i} $(\boldsymbol{x}_1(w_i|s_i),\boldsymbol{x}_2(s_i),\boldsymbol{y}_1(i),\boldsymbol{y}(i))$ is not jointly ϵ-typical.
E_{1i} in decoding step 1, there exists $\tilde{w}\neq w_i$ such that $(\boldsymbol{x}_1(\tilde{w}|s_i),\boldsymbol{x}_2(s_i),\boldsymbol{Y}_1(i))$ is jointly typical.
E_{2i} in decoding step 2, there exists $\tilde{s}\neq s_i$ such that $(\boldsymbol{x}_2(\tilde{s}),\boldsymbol{y}(i))$ is jointly typical.
E_{3i} decoding step 3 fails. Let $E_{3i}=E_{3i}'\cup E_{3i}''$, where
E_{3i}' $w_{i-1}\notin S_{s_i}\cap\mathfrak{L}(\boldsymbol{y}(i-1))$, and
E_{3i}'' there exists $\tilde{w}\in[1,2^{nR}]$, $\tilde{w}\neq w_{i-1}$, such that $\tilde{w}\in S_{s_i}\cap\mathfrak{L}(\boldsymbol{y}(i-1))$.

Now we bound the probability of error over the B n-blocks. Let $\boldsymbol{W}=(W_1,W_2,\cdots,W_{B-1},\varnothing)$ be the transmitted sequence of indices. We assume the indices W_i are independent identically distributed random variables uniformly distributed on $[1,2^{nR}]$. The relay estimates $\boldsymbol{W}$ to be $\hat{\hat{\boldsymbol{W}}}=(\hat{\hat{W}}_1,\hat{\hat{W}}_2,\cdots,\hat{\hat{W}}_{B-1},\varnothing)$. The receiver estimates $\hat{S}=(\varnothing,\hat{S}_2,\hat{S}_3,\cdots,\hat{S}_B)$ and $\hat{\boldsymbol{W}}=(\varnothing,\hat{W}_1,\hat{W}_2,\cdots,\hat{W}_{B-1})$. Define the error events F_i for decoding errors in block i by

$$F_i=\left\{\hat{\hat{W}}_i\neq W_i \text{ or } \hat{W}_{i-1}\neq W_{i-1} \text{ or } \hat{S}_i\neq S_i\right\}=\bigcup_{k=0}^{3}E_{ki}. \tag{28}$$

We have argued in encoding–decoding steps 1) and 2) that

$$P(E_{0i}|F_{i-1}^c)\leqslant\epsilon/4B \tag{29}$$

$$P(E_{1i}\cap E_{0i}^c|F_{i-1}^c)\leqslant\epsilon/4B \tag{30}$$

$$P(E_{2i}\cap E_{0i}^c|F_{i-1}^c)\leqslant\epsilon/4B. \tag{31}$$

We now show that $P(E_{3i}\cap E_{2i}^c\cap E_{0i}^c|F_{i-1}^c)$ can be made small.

Lemma 3: If

$$R<I(X_1;Y|X_2)+R_0-7\epsilon,$$

then for sufficiently large n

$$P(E_{3i}\cap E_{2i}^c\cap E_{0i}^c|F_{i-1}^c)\leqslant\epsilon/4B. \tag{32}$$

Proof: First we bound $E\{\|\mathfrak{L}(\boldsymbol{Y}(i-1)\|\|F_{i-1}^c\}$, where $\|\mathfrak{L}\|$ denotes the number of elements in $\mathfrak{L}$. Let

$$\psi(w|\boldsymbol{y}(i-1))=\begin{cases}1, & (\boldsymbol{x}_1(w|s_{i-1}),\boldsymbol{x}_2(s_{i-1}),\boldsymbol{y}(i-1))\\ & \text{is jointly typical,}\\ 0, & \text{otherwise.}\end{cases} \tag{33}$$

The cardinality of $\mathfrak{L}(\boldsymbol{y}(i-1))$ is the random variable

$$\|\mathfrak{L}(\boldsymbol{y}(i-1))\|=\sum_w\psi(w|\boldsymbol{y}(i-1)) \tag{34}$$

and

$$E\{\|\mathfrak{L}(\boldsymbol{y}(i-1))\|\|F_{i-1}^c\}=E\{\psi(w_{i-1}|\boldsymbol{y}(i-1))|F_{i-1}^c\}+\sum_{w\neq w_{i-1}}E\{\psi(w|\boldsymbol{y}(i-1))|F_{i-1}^c\}.$$

From Lemma 2 we obtain, for each $w \in [1, M]$,

$$E\{\psi(w|y(i-1))|F_{i-1}^c\} \leq 2^{-n(I(X_1;Y|X_2)-7\epsilon)}, \qquad w \neq w_{i-1}. \tag{35}$$

Therefore

$$E\{\|\mathcal{L}(y(i-1))\||F_{i-1}^c\} \leq 1+(2^{nR}-1)(2^{-n(I(X_1;Y|X_2)-7\epsilon)}) \leq 1+2^{n(R-I(X_1;Y|X_2)+7\epsilon)}. \tag{36}$$

The event F_{i-1}^c implies that $w_{i-1} \in \mathcal{L}(y(i-1))$. Also $E_{2i}^c \Rightarrow \hat{s}_i = s_i \Rightarrow w_{i-1} \in S_{\hat{s}_i}$. Thus

$$P(E_{3i}' \cap E_{2i}^c \cap E_{0i}^c | F_{i-1}^c) = 0. \tag{37}$$

Hence

$$\begin{aligned} &P(E_{3i} \cap E_{2i}^c \cap E_{0i}^c | F_{i-1}^c) \\ &= P(E_{3i}'' \cap E_{2i}^c \cap E_{0i}^c | F_{i-1}^c) \\ &\leq P\{\text{there exists } w \neq W_{i-1} \text{ such that } w \in \mathcal{L}(y(i-1)) \cap S_{s_i} | F_{i-1}^c\} \\ &\leq E\left\{ \sum_{\substack{w \neq W_{i-1} \\ w \in \mathcal{L}(y(i-1))}} P(w \in S_{s_i}) | F_{i-1}^c \right\} \\ &\leq E\{\|\mathcal{L}(y(i-1))\| 2^{-nR_0} | F_{i-1}^c\} \\ &\leq 2^{-nR_0}(1+2^{n(R-I(X_1;Y|X_2))+7\epsilon}). \end{aligned}$$

Thus, if

$$R_0 > R - I(X_1; Y|X_2) + 7\epsilon$$

then for sufficiently large n,

$$P\{E_{3i} \cap E_{2i}^c \cap E_{0i}^c | F_{i-1}^c\} \leq \epsilon/4B.$$

But, from (26), we required

$$R_0 < I(X_2; Y).$$

Combining the two constraints, R_0 drops out, leaving

$$R < I(X_1; Y|X_2) + I(X_2; Y) - 7\epsilon = I(X_1, X_2; Y) - 7\epsilon. \tag{38}$$

The probability of error is given by

$$\begin{aligned} P(W \neq \hat{W}) &\leq P\left\{ \bigcup_{i=1}^{B} F_i \right\} \qquad (39) \\ &= P\left\{ \bigcup_{i=1}^{B} \left\{ F_i - \bigcup_{j=1}^{i-1} F_j \right\} \right\} \\ &= \sum_{i=1}^{B} P\{F_i \cap F_1^c \cap F_2^c \cdots \cap F_{i-1}^c\} \\ &\leq \sum_{i=1}^{B} P\{F_i \cap F_{i-1}^c\}. \qquad (40) \end{aligned}$$

But

$$\{F_i \cap F_{i-1}^c\} = \bigcup_{k=0}^{3} E_{ki} \cap F_{i-1}^c. \tag{41}$$

Thus,

$$\begin{aligned} P\{F_i \cap F_{i-1}^c\} &\leq \sum_{k=0}^{3} P\left(\left\{ E_{ki} - \bigcup_{m=0}^{k-1} E_m \right\} \cap F_{i-1}^c \right) \\ &\leq \sum_{k=0}^{3} P(\{E_{ki} \cap E_{0i}^c \cdots \cap E_{k-1i}^c\} | F_{i-1}^c). \end{aligned}$$

The conditional probabilities of error are bounded by

$$P(E_{ki} \cap E_{0i}^c \cdots \cap E_{k-1i}^c | F_{i-1}^c) \leq \frac{\epsilon}{4B}.$$

Thus,

$$P(\hat{W} \neq W) \leq \epsilon. \tag{42}$$

This concludes the proof of Lemma 3.

It is now standard procedure to argue that there exists a code $\mathcal{C}^*$ such that $P(W \neq \hat{W} | \mathcal{C}^*) \leq \epsilon$. Finally, by throwing away the worst half of the w in $\{1, \cdots, 2^{nR}\}^{B-1}$ and reindexing them, we have the maximal error

$$P(\hat{W} \neq w_i | \mathcal{C}^*, w_i) \leq 2\epsilon, \qquad \text{for } i \in [1, 2^{nR(B-1)-1}]. \tag{43}$$

Thus, for $\epsilon > 0$, and n sufficiently large,

$$\lambda_n \leq 2\epsilon \text{ for rates } \tilde{R} = \frac{nR(B-1)-1}{nB} \text{ where } R < C. \tag{44}$$

First letting $n \to \infty$, then $B \to \infty$, and finally $\epsilon \to 0$, we see that $R < C$ is achievable. Thus the achievability of C in Theorem 1 is proved.

III. Converse

First we show that for the *general* (not necessarily degraded) relay channel an upper bound to the capacity C is given by

$$C \leq \sup_{p(x_1, x_2)} \min\{I(X_1, X_2; Y), I(X_1; Y, Y_1 | X_2)\}. \tag{45}$$

Theorem 4: If

$$R > \sup_{p(x_1, x_2)} \min\{I(X_1, X_2; Y), I(X_1; Y, Y_1 | X_2)\}$$

then there exists $\lambda > 0$ such that $\bar{P}_n(e) > \lambda$ for all n.

Before proving Theorem 4, we note that this theorem can be specialized to give the converses to Theorems 1, 2, and 3.

Corollary 1: (Converse to Theorem 1). For the degraded relay channel

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

Proof: It follows from the degradedness assumption (10) that

$$I(X_1; Y, Y_1 | X_2) = I(X_1; Y_1 | X_2)$$

thus rendering the upper bound in Theorem 4 and the bound in Corollary 1 equal.

Corollary 2: (Converse to Theorem 2). The reversely degraded relay channel has capacity

$$C_0 \leq \max_{p(x_1)} \max_{x_2} \{I(X_1; Y | x_2)\}.$$

Proof: Reverse degradation (11) implies in Theorem 4 that

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

Also, the second term in the brackets is always less than the first:

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

Finally,

$$C_0 \leqslant \sup_{p(x_1,x_2)} I(X_1; Y|X_2) = \max_{x_2} \max_{p(x_1)} I(X_1; Y|x_2)$$

since I is linear in $p(x_2)$, and $p_{x_2}(\cdot)$ takes values in a simplex. Thus I is maximized at an extreme point. Q.E.D.

Proof of Theorem 4: Given any (M,n) code for the relay channel, the probability mass function on the joint ensemble W, $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, $\boldsymbol{Y}$, $\boldsymbol{Y}_1$ is given by

$$p(w,\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{y},\boldsymbol{y}_1) = \frac{1}{M} p(\boldsymbol{x}_1|w) \prod_{i=1}^{n} p(x_{2i}|y_{11},\cdots,y_{1i-1}) \cdot p(y_i,y_{1i}|x_{1i},x_{2i}). \quad (46)$$

Consider the identity

$$nR = H(W) = I(W;\boldsymbol{Y}) + H(W|\boldsymbol{Y}). \quad (47)$$

By Fano's inequality

$$H(W|\boldsymbol{Y}) \leqslant \bar{P}_n(e) nR + 1 \triangleq n\delta_n. \quad (48)$$

Thus

$$nR \leqslant I(W;\boldsymbol{Y}) + n\delta_n.$$

We now upper bound $I(W;\boldsymbol{Y})$ in a lemma that is essentially similar to Theorem 10.1 in van der Meulen [1, p. 152].

Lemma 4:

i) $$I(W;\boldsymbol{Y}) \leqslant \sum_{i=1}^{n} I(X_{1i},X_{2i};Y_i) \quad (49a)$$

ii) $$I(W;\boldsymbol{Y}) \leqslant \sum_{i=1}^{n} I(X_{1i};Y_{1i},Y_i|X_{2i}). \quad (49b)$$

Proof: To simplify notation, we shall use $Y^i = (Y_1, Y_2, \cdots, Y_i)$ throughout the rest of this paper. First considering i), we apply the chain rule to obtain

$$\begin{aligned} I(W;\boldsymbol{Y}) &= \sum_{i=1}^{n} I(W;Y_i|Y^{i-1}) \\ &= \sum_{i=1}^{n} \left(H(Y_i|Y^{i-1}) - H(Y_i|W,Y^{i-1})\right) \\ &\leqslant \sum_{i=1}^{n} \left(H(Y_i) - H(Y_i|W,Y^{i-1})\right) \\ &\leqslant \sum_{i=1}^{n} \left(H(Y_i) - H(Y_i|X_{1i},X_{2i},W,Y^{i-1})\right). \end{aligned} \quad (50)$$

By discrete memorylessness of the channel, Y_i and (W,Y^{i-1}) are conditionally independent given (X_{1i},X_{2i}). Thus the conditioning can be dropped in the last term, yielding

$$I(W;\boldsymbol{Y}) \leqslant \sum_{i=1}^{n} I(X_{1i},X_{2i};Y_i).$$

Considering ii) we have

$$\begin{aligned} I(W;\boldsymbol{Y}) &\leqslant I(W;\boldsymbol{Y},\boldsymbol{Y}_1) \\ &= \sum_{i=1}^{n} I(W;Y_i,Y_{1i}|Y^{i-1},Y_1^{i-1}) \\ &= \sum_{i=1}^{n} H(W|Y^{i-1},Y_1^{i-1}) - H(W|Y^{i-1},Y_1^{i-1}). \end{aligned} \quad (51)$$

It is easy to see that W and X_{2i} are conditionally independent given (Y^{i-1},Y_1^{i-1}). Hence

$$H(W|Y^{i-1},Y_1^{i-1},X_{2i}) = H(W|Y^{i-1},Y_1^{i-1}) \quad (52)$$

and continuing the sequence of upper bounds in (51), we have

$$\begin{aligned} I(W;\boldsymbol{Y}) &\leqslant \sum_{i=1}^{n} \left(H(W|Y^{i-1},Y_1^{i-1},X_{2i}) - H(W|Y^{i},Y_1^{i},X_{2i})\right) \\ &= \sum_{i=1}^{n} I(W;Y_{1i},Y_i|Y^{i-1},Y_1^{i-1},X_{2i}) \\ &= \sum_{i=1}^{n} \left(H(Y_{1i},Y_i|Y^{i-1},Y_1^{i-1},X_{2i}) - H(Y_{1i},Y_i|W,Y^{i-1},Y_1^{i-1},X_{2i})\right) \\ &\leqslant \sum_{i=1}^{n} \left(H(Y_{1i},Y_i|X_{2i}) - H(Y_{1i},Y_i|X_{1i},X_{2i})\right) \\ &= \sum_{i=1}^{n} I(X_{1i};Y_i,Y_{1i}|X_{2i}) \end{aligned}$$

and Lemma 4 is proved.

From (48) and Lemma 4 it follows that

$$R \leqslant \min\left\{\frac{1}{n}\sum_{i=1}^{n} I(X_{1i},X_{2i};Y_i), \frac{1}{n}\sum_{i=1}^{n} I(X_{1i};Y_i,Y_{1i}|X_{2i})\right\} + \delta_n. \quad (53)$$

We now eliminate the variable n by a simple artifice. Let Z be a random variable independent of $\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y},\boldsymbol{Y}_1$ taking values in the set $\{1,\cdots,n\}$ with probability

$$p(Z=i) = \frac{1}{n} \qquad 1 \leqslant i \leqslant n. \quad (54)$$

Set

$$X_1 \triangleq X_{1Z}, \quad X_2 \triangleq X_{2Z}, \quad Y \triangleq Y_Z, \quad Y_1 \triangleq Y_{1Z}.$$

Then

$$\frac{1}{n}\sum_{i=1}^{n} I(X_{1i},X_{2i};Y_i) = I(X_1,X_2;Y|Z) \leqslant I(X_1,X_2;Y)$$

by the Markovian relation $Z \to (X_1,X_2) \to (Y,Y_1)$ induced by the channel and the code. Similarly

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} I(X_{1i};Y_i,Y_{1i}|X_{2i}) &= I(X_1;Y,Y_1|X_2,Z) \\ &\leqslant I(X_1;Y,Y_1|X_2). \end{aligned}$$

Thus

$$R \leqslant \min\{I(X_1,X_2;Y), I(X_1;Y_1,Y|X_2)\} + \delta_n \quad (55)$$

and Theorem 4 is proved.

IV. The Gaussian Degraded Relay Channel

Suppose a transmitter x_1 with power P_1 sends a signal intended for receiver y. However, this signal is also received by a relay y_1 that is perhaps physically closer to x_1 than is y. Transmissions are corrupted by additive Gaussian noise. How can the relay x_2 make good use of y_1 to send a signal at power P_2 that will add to the signal received by the ultimate receiver y?

First we define the model for discrete time additive white Gaussian noise degraded relay channel as shown in Fig. 3.

Let $\boldsymbol{Z}_1=(Z_{11},\cdots,Z_{1n})$ be a sequence of independent identically distributed (i.i.d.) normal random variables (r.v.'s) with mean zero and variance N_1, and let $\boldsymbol{Z}_2=(Z_{21},\cdots,Z_{2n})$ be i.i.d. normal r.v.'s independent of $\boldsymbol{Z}_1$ with mean zero and variance N_2. Define $N=N_1+N_2$. At the ith transmission the real numbers x_{1i} and x_{2i} are sent and

$$y_{1i}=x_{1i}+z_{1i}$$
$$y_i=x_{2i}+y_{1i}+z_i \tag{56}$$

are received. Thus the channel is degraded.

Let the message power constraints on the transmitted power be

$$\frac{1}{n}\sum_{i=1}^{n}x_{1i}^2(w)\leqslant P_1,\qquad w\in\{1,2,\cdots,M\} \tag{57a}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}x_{2i}^2(y_{11},y_{12},\cdots,y_{1i-1})\leqslant P_2,\qquad (y_{11},\cdots,y_{1n})\in\mathbb{R}^n \tag{57b}$$

for the transmitted signal $\boldsymbol{x}_1=(x_{11},\cdots,x_{1n})$ and the relay signal $\boldsymbol{x}_2=(x_{21},\cdots,x_{2n})$, respectively.

The definition of a code for this channel is the same as given in Section I with the additional constraints in (57).

Theorem 5: The capacity C^* of the Gaussian degraded relay channel is given by

$$C^*=\max_{0\leqslant\alpha\leqslant 1}\min\left\{C\left(\frac{P_1+P_2+2\sqrt{\bar{\alpha}P_1P_2}}{N}\right),C\left(\frac{\alpha P_1}{N_1}\right)\right\} \tag{58}$$

where $\bar{\alpha}=(1-\alpha)$ and

$$C(x)=\tfrac{1}{2}\log(1+x)\qquad x\geqslant 0.$$

Remarks: 1) If

$$P_2/N_2\geqslant P_1/N_1 \tag{59}$$

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 $C(P_1/N_1)$.

2) For $P_2/N_2<P_1/N_1$, it can be seen that the maximizing $\alpha=\alpha^*$ is strictly less than one, and is given by solving for α in

$$\frac{1}{2}\ln\left(1+\frac{P_1+P_2+2\sqrt{\bar{\alpha}P_1P_2}}{N_1+N_2}\right)=\frac{1}{2}\ln\left(1+\frac{\alpha P_1}{N_1}\right) \tag{61}$$

yielding $C^*=C(\alpha^*P_1/N_1)$.

Proof: We first sketch the achievability of C^* and the random code that achieves it. For $0\leqslant\alpha\leqslant 1$, let $X_2\sim N(0,P_2)$, $X_{10}\sim N(0,\alpha P_1)$, with X_{10},X_2 independent, and let

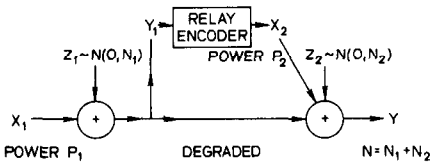

Fig. 3. Degraded Gaussian relay channel.

$X_1=\sqrt{\bar{\alpha}P_1/P_2}\,X_2+X_{10}$. Then, referring to Theorem 1, we evaluate

$$I(X_1,X_2;Y)=\frac{1}{2}\log\left(1+\frac{P_1+P_2+2\sqrt{\bar{\alpha}P_1P_2}}{N_1+N_2}\right)$$
$$I(X_1;Y_1|X_2)=\frac{1}{2}\log\left(1+\frac{\alpha P_1}{N_1}\right). \tag{62}$$

The assertion that this distribution $p(x_1,x_2)$ actually maximizes $\min\{I(X_1;Y_1|X_2),I(X_1,X_2;Y)\}$ will follow from the proof of the converse.

The random code book (Section II) associated with this distribution is then given by a random choice of

$$\tilde{\boldsymbol{X}}_1(w)\ \text{i.i.d.}\sim N_n(0,\alpha P_1 I)\qquad w\in[1,2^{nR}]$$
$$\boldsymbol{X}_2(s)\ \text{i.i.d.}\sim N_n(0,P_2 I)\qquad s\in[1,2^{nR_0}]$$

where $R_0=(1/2)\log(1+(\sqrt{P_2}+\sqrt{\bar{\alpha}P_1})^2/(\alpha P_1+N))-\epsilon$, and $N_n(0,I)$ denotes the n-variate normal distribution with identity covariance matrix I. The code book is given by

$$\boldsymbol{x}_1(w|s)=\tilde{\boldsymbol{x}}_1(w)+\sqrt{\frac{\bar{\alpha}P_1}{P_2}}\,\boldsymbol{x}_2(s)$$
$$\boldsymbol{x}_2(s),\quad w\in[1,2^{nR}],\quad s\in[1,2^{nR_0}].$$

The codewords so generated $(1-\epsilon)$-satisfy the power constraints with high probability, and thus the overall average probability of error can be shown to be small.

Now the converse. Any code for the channel specifies a joint probability distribution on $W,\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{Y}_1,\boldsymbol{Y}$. Lemma 4 gives

$$\text{ii)}\quad nR\leqslant\sum_{i=1}^{n}I(X_{1i};Y_{1i},Y_i|X_{2i})+n\delta_n$$
$$=\sum_{i=1}^{n}I(X_{1i};Y_{1i}|X_{2i})+n\delta_n \tag{63}$$

by degradedness.Thus,

$$nR\leqslant\sum_{i=1}^{n}\left[H(Y_{1i}|X_{2i})-H(Y_{1i}|X_{1i},X_{2i})\right]+n\delta_n$$
$$=\sum_{i=1}^{n}\left[H(Y_{1i}|X_{2i})-\tfrac{1}{2}\ln 2\pi eN_1\right]+n\delta_n. \tag{64}$$

Now, for any i,

$$H(Y_{1i}|X_{2i})=EH(Y_{1i}|x_{2i})$$
$$\leqslant E\tfrac{1}{2}\ln 2\pi e\,\text{var}(Y_{1i}|x_{2i})$$
$$\leqslant\tfrac{1}{2}\ln 2\pi eE\,\text{var}(Y_{1i}|x_{2i}). \tag{65}$$

The last step is an application of Jensen's inequality. Since

$Y_{1i} = X_{1i} + Z_{1i}$, then

$$E\,\mathrm{var}(Y_{1i}|X_{2i}) = E\,\mathrm{var}(X_{1i}|X_{2i}) + E(Z_{1i}^2) = A_i + N_1 \tag{66}$$

where $E\,\mathrm{var}(X_{1i}|X_{2i}) \triangleq A_i$, $1 \leqslant i \leqslant n$. Substituting in (64), we have

$$R \leqslant \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\ln\left(1+\frac{A_i}{N_1}\right)+\delta_n \leqslant \frac{1}{2}\ln\left(1+\frac{\frac{1}{n}\sum_{i=1}^{n}A_i}{N_1}\right)+\delta_n \tag{67}$$

again by Jensen's inequality. However

$$\frac{1}{n}\sum_{i=1}^{n}A_i = \frac{1}{n}\sum_{i=1}^{n}\left(E(X_{1i}^2) - E(E^2(X_{1i}|X_{2i}))\right) \leqslant P_1 - \frac{1}{n}\sum_{i=1}^{n}EE^2(X_{1i}|X_{2i})$$

by the power constraint on the code book.

Define

$$\bar{\alpha}P_1 = \frac{1}{n}\sum_{i=1}^{n}EE^2(X_{1i}|X_{2i}), \qquad \alpha \in [0,1]. \tag{68}$$

Thus (67) becomes

$$R \leqslant \frac{1}{2}\ln\left(1+\frac{\alpha P_1}{N_1}\right)+\delta_n.$$

Next consider Lemma 4i):

$$\begin{aligned} nR &\leqslant \sum_{i=1}^{n} I(X_{1i},X_{2i};Y_i)+n\delta_n \\ &\leqslant \sum_{i=1}^{n}\left[H(Y_i)-H(Y_i|X_{1i},X_{2i})\right]+n\delta_n \\ &= \sum_{i=1}^{n}\left[H(X_{1i}+X_{2i}+Z_{1i}+Z_i)-\tfrac{1}{2}\ln 2\pi eN\right]+n\delta_n. \end{aligned}$$

For any i,

$$H(X_{1i}+X_{2i}+Z_{1i}+Z_i) \leqslant \tfrac{1}{2}\ln 2\pi e\left(E(X_{1i}+X_{2i})^2+N\right). \tag{69}$$

Hence

$$R \leqslant \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\ln\left(1+\frac{E(X_{1i}+X_{2i})^2}{N}\right)+\delta_n \leqslant \frac{1}{2}\ln\left(1+\frac{\frac{1}{n}\sum_{i=1}^{n}E(X_{1i}+X_{2i})^2}{N}\right)+\delta_n.$$

Now

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}E(X_{1i}+X_{2i})^2 &= \frac{1}{n}\sum_{i=1}^{n}EX_{1i}^2 + \frac{1}{n}\sum_{i=1}^{n}EX_{2i}^2 + \frac{2}{n}\sum_{i=1}^{n}EX_{1i}X_{2i} \\ &\leqslant P_1+P_2+\frac{2}{n}\sum_{i=1}^{n}E\{X_{2i}E\{X_{1i}|X_{2i}\}\}. \end{aligned} \tag{70}$$

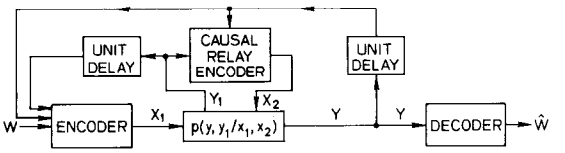

Fig. 4. Relay channel with feedback.

Applying the Cauchy–Schwartz inequality to each term in the sum in (70), we obtain

$$\frac{1}{n}\sum_{i=1}^{n}E(X_{1i}+X_{2i})^2 \leqslant P_1+P_2+\frac{2}{n}\sum_{i=1}^{n}\sqrt{E\left\{E^2(X_{1i}|X_{2i})\right\}EX_{2i}^2}. \tag{71}$$

From (68) and the power constraints we know that

$$\frac{1}{n}\sum_{i=1}^{n}E\left\{E^2(X_{1i}|X_{2i})\right\} = \bar{\alpha}P_1, \quad \frac{1}{n}\sum_{i=1}^{n}EX_{2i}^2 \leqslant P_2.$$

Again applying the Cauchy–Schwartz inequality, we have

$$2\sum_{i=1}^{n}\left(\frac{E\left\{E^2\{X_{1i}|X_{2i}\}\right\}}{n}\right)^{1/2}\left(\frac{EX_{2i}^2}{n}\right)^{1/2} \leqslant 2\sqrt{(\bar{\alpha}P_1)P_2}$$

where the maximum occurs when $E\{E^2(X_{1i}|X_{2i})\} = \bar{\alpha}P_1$ and $EX_{2i}^2 = P_2$ for all i. Therefore

$$R \leqslant \frac{1}{2}\ln\left(1+\frac{P_1+P_2+2\sqrt{\bar{\alpha}P_1P_2}}{N}\right)+\delta_n. \tag{72}$$

The converse follows directly from (68) and (72).

V. The Capacity of the General Relay Channel with Feedback

Suppose we have a relay channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y, y_1|x_1,x_2), \mathcal{Y} \times \mathcal{Y}_1)$. No degradedness relation between y and y_1 will be assumed. Let there be feedback from (y,y_1) to x_1 and to x_2 as shown in Fig. 4.

To be precise the encoding functions in (2) and (3) now become

$$\begin{aligned} &x_{1i}(w,y_1,y_2,\cdots,y_{i-1},y_{11},y_{12},\cdots,y_{1i-1}) \\ &x_{2i}(y_1,y_2,\cdots,y_{i-1},y_{11},y_{12},\cdots,y_{1i-1}). \end{aligned} \tag{73}$$

Placing a distribution on $w \in [1, 2^{nR}]$ thus induces the joint probability mass function.

$$p(w,x_1,x_2,y,y_1) = p(w)\prod_{i=1}^{n}p(x_{1i}|w,y^{i-1},y_1^{i-1}) \cdot p(x_{2i}|y^{i-1},y_1^{i-1})p(y_i,y_{1i}|x_{1i},x_{2i}) \tag{74}$$

where $y^k = (y_1,y_2,\cdots,y_k)$. Theorem 3 states that the capacity C_{FB} of this channel is

$$C_{FB} = \max_{p(x_1,x_2)} \min\{I(X_1,X_2;Y), I(X_1;Y,Y_1|X_2)\}. \tag{75}$$

Proof of Theorem 3: The relay channel with feedback is an ordinary degraded relay channel under the substitution of (Y,Y_1) for Y_1. Thus the code and proof of the forward part of Theorem 1 apply, yielding the ϵ-achievability of C_{FB}. For the converse, inspection of the proof of Theorem 4 reveals that all the steps apply to the feedback channel—the crucial step being (52). Thus the proof of Theorem 3 is complete.

Remark: The code used in Theorem 1 will suffice for the feedback channel. However, this code can be simplified by eliminating the random partition $\mathcal{S}$, because the relay knows the y sequence through the feedback link. An enumeration encoding for the relay can be used instead (see [6]).

Corollary to Theorem 4: If the channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y,y_1|x_1,x_2), \mathcal{Y} \times \mathcal{Y}_1)$ is degraded or reversely degraded, then feedback does not increase the capacity.

Proof: If the channel is degraded, then

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

Thus $C_{FB} = C$. If the channel is reversely degraded, then

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

Thus $C_{FB} = C$.

VI. An Achievable Rate for the General Relay Channel

We are now in a position to discuss the nature of the capacity region for the general relay channel. First, if we have feedback, we know the capacity. Next, we note that the general relay channel will involve the idea of cooperation (for the degraded relay channel) and facilitation (for the reversely degraded relay channel). If y_1 is better than y, then the relay cooperates to send x_1; if y_1 is worse than y, then the relay facilitates the transmission of x_1 by sending the best x_2. Yet another consideration will undoubtedly be necessary—the idea of sending alternative information. This alternative information about x_1 is not zero, thus precluding simple facilitation, and is not perfect, thus precluding pure cooperation.

Finally, we note that the converse (Theorem 4) yields

$$C_{\text{general}} \leq \max_{p(x_1,x_2)} \min\{I(X_1,X_2; Y), I(X_1; Y, Y_1|X_2)\}$$

for the general channel. Moreover, the code construction for Theorem 1 shows that

$$C_{\text{general}} \geq \max_{p(x_1,x_2)} \min\{I(X_1,X_2; Y), I(X_1, Y_1|X_2)\}.$$

Also, from Theorem 2, we see

$$C_{\text{general}} \geq \max_{p(x_1)} \max_{x_2} I(X_1; Y|x_2).$$

If the relay channel is not degraded, cooperation may not be possible and facilitation can be improved upon. As an example, consider the general Gaussian relay channel shown in Fig. 5.

If we assume that $N_1 > N$, then cooperation in the sense of Section II cannot be realized, since every $(2^{nR}, n, \epsilon)$ code for y_1 will be a $(2^{nR}, n, \epsilon)$ code for y. However, the relay sequence $\boldsymbol{Y}_1$ is an "observation" of $\boldsymbol{X}_1$ that is independent of $\boldsymbol{Y}$. Thus sending $\boldsymbol{Y}_1$ to y will decrease the effective noise in the y observation of x_1. If the relay power P_1 is finite, and therefore we cannot send $\boldsymbol{Y}_1$ to y precisely, then we send an *estimate* $\hat{\boldsymbol{Y}}_1$ of $\boldsymbol{Y}_1$. The choice of the estimate $\hat{\boldsymbol{Y}}_1$ will be made clear in Theorem 6 for the discrete memoryless relay channel. Then in Theorem 7, we shall combine Theorems 1 and 6.

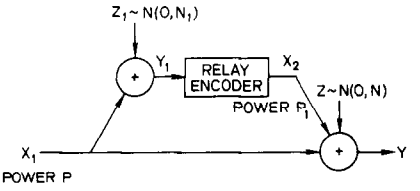

Fig. 5. General Gaussian relay channel.

Theorem 6: Let $(\mathcal{X}_1 \times \mathcal{X}_2, p(y,y_1|x_1,x_2), \mathcal{Y} \times \mathcal{Y}_1)$ be any discrete memoryless relay channel. Then the rate R_1^* is achievable, where

$$R_1^* = \sup I(X_1; Y, \hat{Y}_1|X_2) \tag{76a}$$

subject to the constraint

$$I(X_2; Y) \geq I(Y_1; \hat{Y}_1|X_2, Y) \tag{76b}$$

where the supremum is taken over all joint distributions on $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y} \times \mathcal{Y}_1 \times \hat{\mathcal{Y}}_1$ of the form

$$p(x_1,x_2,y,y_1,\hat{y}_1) = p(x_1)p(x_2)p(y,y_1|x_1,x_2)p(\hat{y}_1|y_1,x_2) \tag{77}$$

and $\hat{Y}_1$ has a finite range.

Outline of Proof: A block Markov encoding is used. At the end of any block i, the x_2 information is used to resolve the uncertainty of the receiver about w_{i-1}.

Random Coding:

1) Choose 2^{nR_1} i.i.d. $\boldsymbol{x}_1$ each with probability $p(\boldsymbol{x}_1) = \prod_{i=1}^n p(x_{1i})$. Label these $\boldsymbol{x}_1(w)$, $w \in [1, 2^{nR_1}]$.

2) Choose 2^{nR_0} i.i.d. $\boldsymbol{x}_2$ each with probability $p(\boldsymbol{x}) = \prod_{i=1}^n p(x_{2i})$. Label these $\boldsymbol{x}_2(s)$, $s \in [1, 2^{nR_0}]$.

3) Choose, for each $\boldsymbol{x}_2(s)$, $2^{n\hat{R}}$ i.i.d. $\hat{\boldsymbol{Y}}_1$ each with probability $p(\hat{\boldsymbol{y}}_1|\boldsymbol{x}_2(s)) = \prod_{i=1}^n p(\hat{y}_{1i}|x_{2i}(s))$, where, for $x_2 \in \mathcal{X}_2$, $\hat{y}_1 \in \hat{\mathcal{Y}}_1$, we define

$$p(\hat{y}_1|x_2) = \sum_{x_1,y,y_1} p(x_1)p(y,y_1|x_1,x_2)p(\hat{y}_1|y_1,x_2). \tag{78}$$

Label these $\hat{y}_1(z|s)$, $s \in [1, 2^{nR_0}]$, $z \in [1, 2^{n\hat{R}}]$.

4) Randomly partition the set $\{1, 2, \cdots, 2^{n\hat{R}}\}$ into 2^{nR_0} cells S_s, $s \in [1, 2^{nR_0}]$.

Encoding: Let w_i be the message to be sent in block i, and assume that $(\hat{\boldsymbol{Y}}_1(z_{i-1}|s_{i-1}), \boldsymbol{Y}_1(i-1), \boldsymbol{x}_2(s_{i-1}))$ are jointly ϵ-typical, and that $Z_{i-1} \in S_{s_i}$. Then the codeword pair $(\boldsymbol{x}_1(w_i), \boldsymbol{x}_2(s_i))$ will be transmitted in block i.

Decoding: At the end of block i we have the following.

i) The receiver estimates s_i by $\hat{s}_i$ by looking for the unique typical $\boldsymbol{x}_2(s_i)$ with $\boldsymbol{y}(i)$. If $R_0 < I(X_2; Y)$ and n is sufficiently large, then this decoding operation will incur small probability of error.

ii) The receiver calculates a set $L(\boldsymbol{y}(i-1))$ of z such that $z \in L(\boldsymbol{y}(i-1))$ if $(\hat{\boldsymbol{y}}_1(z|\hat{s}_{i-1}), \boldsymbol{x}_2(\hat{s}_{i-1}), \boldsymbol{y}(i-1))$ are jointly ϵ-typical. The receiver then declares that z_{i-1} was sent in block $i-1$ if

$$\hat{z}_{i-1} \in S_{\hat{s}_i} \cap L(\boldsymbol{y}(i-1)).$$

But, from an argument similar to that in Lemma 3, we see that $\hat{z}_{i-1} = z_{i-1}$ with arbitrarily high probability provided n is sufficiently large and

$$\hat{R} < I(\hat{Y}_1; Y|X_2) + R_0. \tag{79}$$

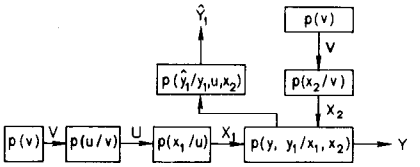

Fig. 6. Relationship of auxiliary variables.

iii) Using both $\hat{y}_1(\hat{z}_{i-1}|\hat{s}_{i-1})$ and $\boldsymbol{y}(i-1)$, the receiver finally declares that $\hat{w}_{i-1}$ was sent in block $i-1$ if $(\boldsymbol{x}_1(\hat{w}_{i-1}),\hat{y}_1(\hat{z}_{i-1}|\hat{s}_{i-1}),\boldsymbol{y}(i-1),\boldsymbol{x}_2(\hat{s}_{i-1}))$ are jointly ϵ-typical. Thus $\hat{w}_{i-1}=w_{i-1}$ with high probability if

$$R_1 < I(X_1; Y, \hat{Y}_1|X_2) \tag{80}$$

and sufficiently large.

iv) The relay, upon receiving $y_1(i)$, decides that z is "received" if $(\hat{y}_1(z|s_i),y_1(i),\boldsymbol{x}_2(s_i))$ are jointly ϵ-typical. There will exist such a z with high probability if

$$\hat{R} > I(\hat{Y}_1; Y_1|X_2) \tag{81}$$

and n is sufficiently large. (See [10], [11], [12] and especially the proof of Lemma 2.1.3 in [12].)

The decoding error calculations for steps i)–iii) are similar to those in Section II. The decoding error in step iv) follows from the theory of side information in [10], [11]. Let

$$\begin{aligned} R_0 &= I(X_2; Y) - \epsilon \\ \hat{R} &= I(\hat{Y}_1; Y_1|X_2) + \epsilon. \end{aligned} \tag{82}$$

This together with the constraint in (79) collapses to yield

$$I(X_2; Y) \geqslant I(\hat{Y}_1; Y_1|X_2, Y).$$

Thus we see that the rate R_1^* given in (76) is achievable.

Remarks:

1) Small values of $I(X_2; Y)$ will constrain $\hat{Y}_1$ to be a highly degraded version of Y_1 in order that (76b) be satisfied.

2) The reversely degraded relay capacity C_0 is always less than or equal to the achievable rate R_1^* given in (76).

We have seen in Section II that the rate of the channel from X_1 to Y can be increased through cooperation. Alternatively we have claimed in Theorem 6 that by transmitting an estimate $\hat{Y}_1$ of Y, the rate can also be increased. The obvious generalization of Theorems 1 and 6 is to superimpose the cooperation and the transmission of $\hat{Y}_1$.

Consider the probability structure of Fig. 6 on $\mathcal{V}\times\mathcal{U}\times\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{Y}\times\mathcal{Y}_1\times\hat{\mathcal{Y}}_1$ where $\mathcal{V}$, $\mathcal{U}$, and $\hat{\mathcal{Y}}_1$ are arbitrary sets. The auxiliary random variables have the following interpretation:

i) V will facilitate cooperation to resolve the residual uncertainty about U.

ii) U will be understood by Y_1 *but not* by Y. (Informally, U plays the role of X_1 in Theorem 1).

iii) $\hat{Y}_1$ is the estimate of Y_1 used in Theorem 6.

iv) X_2 is used to resolve Y uncertainty about Y_1.

Finally it can be shown that the following rate is achievable for any relay channel.

Theorem 7: For any relay channel $(\mathcal{X}_1\times\mathcal{X}_2, p(y, y_1|x_1,x_2), \mathcal{Y}\times\mathcal{Y}_1)$, the rate R^* is achievable, where

$$\begin{aligned} R^* = \sup_P \{\min\{ & I(X_1; Y, \hat{Y}_1|X_2, U) \\ & + I(U; Y_1|X_2, V), I(X_1, X_2; Y) \\ & - I(\hat{Y}_1; Y_1|X_2, X_1, U, Y)\}\} \end{aligned} \tag{83a}$$

where the supremum is taken over all joint probability mass functions of the form

$$\begin{aligned} p(u,v,x_1,x_2,y,y_1,\hat{y}_1) = & p(v)p(u|v)p(x_1|u) \\ & \cdot p(x_2|v)p(y,y_1|x_1,x_2)p(\hat{y}_1|x_2,y_1,u) \end{aligned} \tag{83b}$$

subject to the constraint

$$I(\hat{Y}_1; Y_1|Y, X_2, U) \leqslant I(X_2; Y|V). \tag{83c}$$

Remarks: The forward parts of Theorems 1, 2, and 6 are special cases of R^* in Theorem 7, as the following substitutions show:

1) (Degraded channel, Theorem 1) $R^* \geqslant C$. Choose $V\equiv X_2$, $U\equiv X_1$, and $\hat{Y}_1\equiv\phi$.

2) (Reversely degraded channel, Theorem 2) $R^* \geqslant C$. Choose $V\equiv\phi$, $U\equiv\phi$, and $\hat{Y}_1\equiv\phi$.

3) (Theorem 6) $R^* \geqslant R_1^*$. Choose $V\equiv\phi$, $U\equiv\phi$.

Proof of Theorem 7 (Outline): As in Theorems 1 and 6, a block Markov encoding scheme is used. At the end of block i, v is used to resolve the uncertainty of y about the past u, and x_2 is used to resolve the y uncertainty about $\hat{Y}_1$, thus enabling y to decode w_{i-1}.

Random Coding:

1) Generate $2^{n(I(V;Y)-\epsilon)}$ i.i.d. $\boldsymbol{v}$, each with probability $p(\boldsymbol{v})=\prod_{i=1}^n p(v_i)$. Label these $\boldsymbol{v}(m)$, $m\in[1,2^{n(I(V;Y)-\epsilon)}]$.

2) For every $\boldsymbol{v}(m)$, generate $2^{n(I(X_2;Y|V)-\epsilon)}$ i.i.d. $\boldsymbol{x}_2$, each with probability

$$p(\boldsymbol{x}_2|\boldsymbol{v}(m)) = \prod_{i=1}^n p(x_{2i}|v_i(m)).$$

Label these $\boldsymbol{x}_2(s|m)$, $s\in[1,2^{n(I(X_2;Y|V)-\epsilon)}]$.

3) For every $\boldsymbol{v}(m)$, generate 2^{nR_1} i.i.d. $\boldsymbol{u}$ each with probability

$$p(\boldsymbol{u}|\boldsymbol{v}(m)) = \prod_{i=1}^n p(u_i|v_i(m)).$$

Label these $\boldsymbol{u}(w'|m)$, $w'\in[1,2^{nR_1}]$.

4) For every $\boldsymbol{u}(w'|m)$ generate 2^{nR_2} i.i.d. $\boldsymbol{x}_1$, each with probability

$$p(\boldsymbol{x}_1|\boldsymbol{u}(w'|m)) = \prod_{i=1}^n p(x_{1i}|u_i(w'|m)).$$

Label these $\boldsymbol{x}_1(w''|m,w')$, $w''\in[1,2^{nR_2}]$.

5) For every $(\boldsymbol{x}_2(s|m), \boldsymbol{u}(w'|m))$, generate $2^{n(I(\hat{Y}_1;Y_1|X_2,U)+\epsilon)}$ i.i.d. $\hat{\boldsymbol{y}}_1$ each with probability

$$p(\hat{\boldsymbol{y}}_1|\boldsymbol{x}_2(s|m),\boldsymbol{u}(w'|m)) = \prod_{i=1}^n p(\hat{y}_{1i}|x_{2i}(s|m),u_i(w'|m))$$

where, for every $x_2\in\mathcal{X}_2$, $u\in\mathcal{U}$, we define

$$p(\hat{y}_1|x_2,u) = \frac{\sum_{v,x_1,y,y_1} p(v,u,x_1,x_2,y,y_1,\hat{y}_1)}{\sum_{v,x_1,y,y_1,\hat{y}_1} p(v,u,x_1,x_2,y,y_1,\hat{y}_1)} \tag{84}$$

and $p(v,u,x_1,x_2,y,y_1,\hat{y}_1)$ is defined in (83b). Label these $\hat{y}_1(z|w',s,m)$, $z\in[1,2^{n(I(\hat{Y}_1;Y_1|X_2,U)+\epsilon)}]$.

Random Partitions:

1) Randomly partition the set $\{1,2,\cdots,2^{nR_1}\}$ into $2^{n(I(V;Y)-\epsilon)}$ cells S_{1m}.

2) Randomly partition the set $\{1,2,\cdots,2^{n(I(\hat{Y}_1;Y_1|X_2,U)+\epsilon)}\}$ into $2^{n(I(X_2;Y|V)-\epsilon)}$ cells S_{2s}.

Encoding: Let $w_i=(w_i',w_i'')$ be the message to be sent in block i, and assume that

$$(\hat{y}_1(z_{i-1}|w'_{i-1},s_{i-1},m_{i-1}),y_1(i-1),\cdot u(w'_{i-1}|m_{i-1}),x_2(s_{i-1}|m_{i-1}))$$

are jointly ϵ-typical and that $w'_{i-1}\in S_{1m_i}$ and $z_{i-1}\in S_{2s_i}$. Then the codeword pair $(x_1(w_i''|m_i,w_i'),x_2(s_i|m_i))$ will be transmitted in block i.

Decoding: At the end of block i we have the following.

i) The receiver estimates m_i and s_i, by first looking for the unique ϵ-typical $v(m_i)$ with $y(i)$, then for the unique ϵ-typical $x_2(s_i|m_i)$ with $(y(i),v(m_i))$. For sufficiently large n this decoding step can be done with arbitrarily small probability of error. Let the estimates of s_i and m_i be $\hat{s}_i$, $\hat{m}_i$ respectively.

ii) The receiver calculates a set $L_1(y(i-1))$ of w' such that $w'\in L_1(y(i-1))$ if $(u(w'|m_{i-1}),y(i-1))$ are jointly ϵ-typical. The receiver then declares that $\hat{w}'_{i-1}$ was sent in block $i-1$ if

$$\hat{w}_{i-1}\in S_{1m_i}\cap L_1(y(i-1)). \tag{85}$$

From Lemma 3, we see that $\hat{w}'_{i-1}=w'_{i-1}$ with arbitrarily high probability provided n is sufficiently large and

$$R_1<I(V;Y)+I(U;Y|X_2,V)-\epsilon. \tag{86}$$

iii) The receiver calculates a set $L_2(y(i-1))$ of z such that $z\in L_2(y(i-1))$ if $(\hat{y}_1(z|\hat{w}'_{i-1},\hat{s}_{i-1},\hat{m}_{i-1}),x_2(\hat{s}_{i-1}|\hat{m}_{i-1}),y(i-1))$ are jointly ϵ-typical. The receiver declares that $\hat{z}_{i-1}$ was sent in block $i-1$ if

$$\hat{z}_{i-1}\in S_{2s_i}\cap L_2(y(i-1)). \tag{87}$$

From [12] we see that $\hat{z}_{i-1}=z_{i-1}$ with arbitrarily small probability of error if n is sufficiently large and

$$I(\hat{Y}_1;Y_1|X_2,U)+\epsilon<I(\hat{Y}_1;Y|X_2,U)+I(X_2;Y|V)-\epsilon$$

i.e.,

$$I(X_2;Y|V)>I(\hat{Y}_1;Y_1|X_2,U)-I(\hat{Y}_1;Y|X_2,U)+2\epsilon. \tag{88}$$

But since

$$I(\hat{Y}_1;Y_1,Y|X_2,U)=I(\hat{Y}_1;Y_1|X_2,U)$$

then condition (88) becomes

$$I(X_2;Y|V)>I(\hat{Y}_1;Y_1|Y,X_2,U)+2\epsilon$$

which as $\epsilon\to 0$ gives condition (83c) in Theorem 7.

iv) Using both $\hat{y}_1(\hat{z}_{i-1}|\hat{w}'_{i-1},\hat{s}_{i-1},\hat{m}_{i-1})$ and $y(i-1)$, the receiver finally declares that $\hat{w}''_{i-1}$ was sent in block $i-1$ if $(x_1(\hat{w}''_{i-1}|\hat{m}_{i-1},\hat{w}'_{i-1}),\hat{y}_1(\hat{z}_{i-1}|\hat{w}'_{i-1},\hat{s}_{i-1},\hat{m}_{i-1}),y(i-1))$ are jointly ϵ-typical. $\hat{w}''_{i-1}=w''_{i-1}$ with high probability if

$$R_2=I(X_1;Y,\hat{Y}_1|X_2,U)-\epsilon \tag{89}$$

and n is sufficiently large.

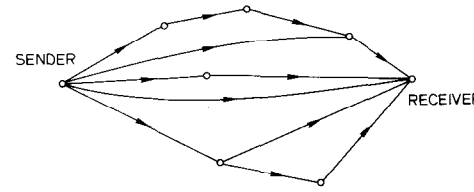

Fig. 7. Single sender single receiver network.

v) The relay upon receiving $y_1(i)$ declares that $\hat{w}'$ was received if $(u(\hat{w}'|m_i),y_1(i),x_2(s_i|m_i))$ are jointly ϵ-typical. $w_i'=\hat{w}'$ with high probability if

$$R_1<I(U;Y_1|X_2,V) \tag{90}$$

and n is sufficiently large. Thus, the relay knows that $w_i'\in S_{1m_{i+1}}$.

vi) The relay also estimates z_i such that $(\hat{y}_1(z_i|w_i',s_i,m_i),y_1(i),x_2(s_i|m_i))$ are jointly ϵ-typical. Such a z_i will exist with high probability for large n, therefore the relay knows that $z_i\in S_{2s_{i+1}}$.

From (86), (89), and (90), we obtain

$$R_1<I(V;Y)+I(U;Y|X_2,V)-\epsilon$$

$$R_1<I(U;Y_1|X_2,V)$$

$$R_2=I(X_1;Y,\hat{Y}_1|X_2,U)-\epsilon.$$

Therefore, the rate of transmission from X_1 to Y is bounded by

$$R<I(U;Y_1|X_2,V)+I(X_1;Y,\hat{Y}_1|X_2,V)-\epsilon$$

$$R<I(V;Y)+I(U;Y|X_2,V)+I(X_1;Y,\hat{Y}_1|X_2,U)-2\epsilon. \tag{91}$$

Substituting from (88) we obtain

$$\begin{aligned}R&<I(X_2,V;Y)-I(\hat{Y}_1;Y_1|Y,X_2,U)+I(U;Y|X_2,V)\\&\quad+I(X_1;Y,\hat{Y}_1|X_2,U)-4\epsilon\\&=I(X_1,X_2;Y)-I(\hat{Y}_1;Y_1|X_2,X_1,U,Y)-4\epsilon\end{aligned} \tag{92}$$

which establishes Theorem 7.

VII. Concluding Remarks

Theorems 1, 2, and 3 establish capacity for degraded, reversely degraded, and feedback relay channels. The full understanding of the relay channel may yield the capacity of the single sender single receiver network in Fig. 7. This will be the information theoretic generalization of the well-known maximum flow–minimum cut theorem [9].

Acknowledgment

The authors are extremely grateful for the very extensive and helpful remarks of the referees. The authors also wish to thank J. Wolfowitz for his helpful comments.

REFERENCES

[1] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.
[2] ——, "Transmission of information in a T-terminal discrete memoryless channel," Ph.D. dissertation, Dep. of Statistics, University of California, Berkeley, 1968.
[3] ——, "A survey of multi-way channels in information theory: 1961–1976," *IEEE Trans. Inform. Theory*, vol. IT-23, no. 2, January 1977.
[4] H. Sato, "Information transmission through a channel with relay," The Aloha System, University of Hawaii, Honolulu, Tech. Rep. B76-7, Mar. 1976.
[5] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, July 1948.
[6] T. M. Cover and S. K. Leung-Yan-Cheong, "A rate region for the multiple-access channel with feedback," submitted to *IEEE Trans. Inform. Theory*, 1976.
[7] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, July, 1973.
[8] 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.
[9] L. R. Ford and D. R. Fulkerson, *Flows in Networks*, Princeton, NJ: Princeton Univ. Press, 1962.
[10] R. Ahlswede and J. Körner, "Source coding with side information and converse for the degraded broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 629–637, Nov. 1975.
[11] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 294–300, May 1975.
[12] T. Berger, "Multiterminal source coding," Lecture notes presented at the 1977 CISM Summer School, Udine, Italy, July 18–20, 1977, Springer–Verlag.
[13] J. Wolfowitz, *Coding Theorems of Information Theory*, New York: Springer–Verlag, Third Ed., 1964.
[14] T. Cover, "An achievable rate region for the broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 399–404, Jul. 1975.

Computational Moments for Sequential Decoding of Convolutional Codes

TAKESHI HASHIMOTO AND SUGURU ARIMOTO, MEMBER, IEEE

***Abstract*—The long standing conjecture is established that, for a discrete memoryless channel, there exists a linear convolutional code with infinite constraint length such that the ρth ($\rho \geqslant 1$) moment of the number of F-hypotheses in the Fano sequential decoding algorithm is bounded, provided that the transmission rate R is less than $E_0(\rho,r)/\rho$, where $r(x)$ is a distribution over the channel input alphabet. A new concept of independence for a finite set of message sequences plays an essential role in averaging a product of likelihood ratios over an ensemble of code sequences in a code tree. A simpler version of the method can be applied to the proof of the conjecture for general tree codes.**

I. INTRODUCTION

ONE OF THE most important problems associated with sequential decoding for a class of discrete memoryless channels is to estimate *a priori* upper bounds on the moments of the number of computation steps or F-hypotheses in decoding. It has long been conjectured that for any $\rho > 0$, the ρth moment of the number of F-hypotheses is bounded for optimal tree codes or linear convolutional codes with rate R, provided that

$$R < \frac{1}{\rho} E_0(\rho, r),$$

where $r(x)$ is a probability distribution over the channel inputs,

$$E_0(\rho,r) = -\ln \sum_y \left[\sum_x P(y|x)^{1/(1+\rho)} r(x) \right]^{1+\rho},$$

and $P(y|x)$ is the transmission probability function of the discrete memoryless channel.

This conjecture was first proved by Falconer [1] for $\rho \in [0,1]$, by Savage [2] for integer values of ρ, and by Jelinek [3] for all $\rho \geqslant 1$. However, the results by Savage and Jelinek do not necessarily imply the validity of the conjecture for the ensemble of linear convolutional codes, although Falconer's result is valid for both tree codes and linear convolutional codes. In addition, Savage's method [2], which is also used by Jelinek [3], is quite involved.

In this paper we prove the correctness of the conjecture for any $\rho \geqslant 1$ in the case of infinite constraint-length linear convolutional codes. To prove this on the basis of random coding techniques we introduce a new concept of independence, called N-independence, for a finite set of message sequences with equal length. This is a generalization of the pairwise independence of Gallager [4] and Massey [6], and plays an essential role in averaging a function of likelihood ratios over an ensemble of code sequences in a code tree. A simpler version of the method gives rise to a self-contained proof of the conjecture for the case of general tree codes.

Throughout the paper we will assume that the reader is familiar with the Fano algorithm for sequential decoding

Manuscript received August 1, 1978; revised January 28, 1979.
The authors are with the Faculty of Engineering Science, Osaka University, Toyonaka, Osaka, Japan.

0018-9448/79/0900-0584$00.75