THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

Volume 54 | October 1975 | Number 8

The Wire-Tap Channel

By A. D. WYNER

(Manuscript received May 9, 1975)

We consider the situation in which digital data is to be reliably transmitted over a discrete, memoryless channel (DMC) that is subjected to a wire-tap at the receiver. We assume that the wire-tapper views the channel output via a second DMC. Encoding by the transmitter and decoding by the receiver are permitted. However, the code books used in these operations are assumed to be known by the wire-tapper. The designer attempts to build the encoder-decoder in such a way as to maximize the transmission rate R, and the equivocation d of the data as seen by the wire-tapper. In this paper, we find the trade-off curve between R and d, assuming essentially perfect ("error-free") transmission. In particular, if d is equal to H_S, the entropy of the data source, then we consider that the transmission is accomplished in perfect secrecy. Our results imply that there exists a $C_s > 0$, such that reliable transmission at rates up to C_s is possible in approximately perfect secrecy.

I. INTRODUCTION

In this paper we study a (perhaps noisy) communication system that is being wire-tapped via a second noisy channel. Our object is to encode the data in such a way that the wire-tapper's level of confusion will be as high as possible. To fix ideas, consider first the simple special case depicted in Fig. 1 (in which the main communication system is noiseless). The source emits a data sequence $S_1, S_2, \cdots$, which consists of independent copies of the binary random variable S, where $\Pr\{S = 0\} = \Pr\{S = 1\} = \frac{1}{2}$. The encoder examines the first K source bits $\mathbf{S}^K = (S_1, \cdots, S_K)$ and encodes $\mathbf{S}^K$ into a binary N vector

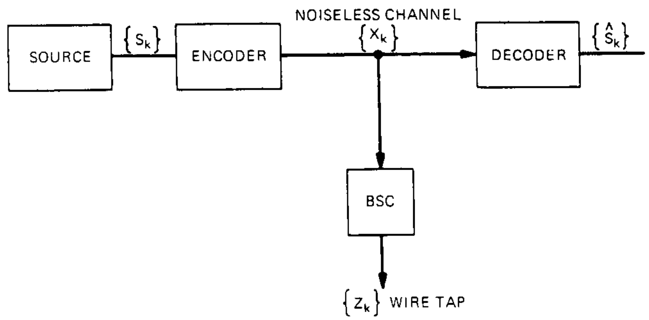

Fig. 1—Wire-tap channel (special case).

$\mathbf{X}^N = (X_1, \cdots, X_N)$. $\mathbf{X}^N$ in turn is transmitted perfectly to the decoder via the noiseless channel and is transformed into a binary data stream $\hat{\mathbf{S}}^K = (\hat{S}_1, \cdots, \hat{S}_K)$ for delivery to the destination. The "error probability" is defined as

$$P_e = \frac{1}{K} \sum_{k=1}^{K} \Pr\{S_k \neq \hat{S}_k\}. \tag{1}$$

The entire process is repeated on successive blocks of K source bits. The transmission rate is K/N bits per transmitted channel symbol.

The wire-tapper observes the encoded vector $\mathbf{X}^N$ through a (memoryless) binary symmetric channel (BSC) with crossover probability $p_0 (0 < p_0 \leqq \frac{1}{2})$. The corresponding output at the wire-tap is $\mathbf{Z}^N = (Z_1, \cdots, Z_N)$, so that for $x, z = 0, 1$ $(1 \leqq n \leqq N)$,

$$\Pr\{Z_n = z | X_n = x\} = (1 - p_0)\delta_{x,z} + p_0(1 - \delta_{x,z}).$$

We take the equivocation

$$\Delta \triangleq \frac{1}{K} H(\mathbf{S}^K | \mathbf{Z}^N) \tag{2}$$

as a measure of the degree to which the wire-tapper is confused. The logarithms in H are, as are all logarithms in this paper, taken to the base 2. The system designer would like to have P_e close to zero, with K/N and Δ as large as possible.

Consider the following schemes:

(*i*) Set $K = N = 1$, and let $X_1 \equiv S_1$. This results in $P_e = 0$, $K/N = 1$, and $\Delta = H(X_1|Z_1) = h(p_0)$, where

$$h(\lambda) = -\lambda \log \lambda - (1 - \lambda) \log (1 - \lambda), \quad 0 \leqq \lambda \leqq 1, \tag{3}$$

(take $0 \log 0 = 0$).

(*ii*) Set $K = 1$, and let N be arbitrary. Let C_0 be the subset of binary N space, $\{0, 1\}^N$, consisting of those N vectors with even parity (i.e., an even number of 1's). Let $C_1 \subseteq \{0, 1\}^N$ be the subset of vectors with odd parity. The encoder works as follows. When $S_1 = i$, $(i = 0, 1)$, the encoder output $\mathbf{X}^N$ is a randomly chosen vector in C_i. Thus, the encoder is a *channel* with transition probability

$$\Pr\{\mathbf{X}^N = \mathbf{x} | S_1 = i\} = \begin{cases} 2^{-(N-1)}, & \mathbf{x} \in C_i, \\ 0, & \mathbf{x} \notin C_i, \end{cases}$$

for $i = 0, 1$. Clearly, the decoder can recover S_1 from $\mathbf{X}^N$ perfectly, so that $P_e = 0$. We now turn to the wire-tapper who observes $\mathbf{Z}^N$, the output of the BSC corresponding to the input $\mathbf{X}^N$. Let $\mathbf{z} \in \{0, 1\}^N$ be a vector of, say, even parity. Then

$$\begin{aligned} \Pr\{S_1 = 0 | \mathbf{Z}^N = \mathbf{z}\} &= \Pr\left\{\begin{matrix}\text{the BSC makes an} \\ \text{even number of errors}\end{matrix}\right\} \\ &= \sum_{\substack{j=0 \\ j \text{ even}}}^{N} \binom{N}{j} p_0^j (1 - p_0)^{N-j} = \tfrac{1}{2} + \tfrac{1}{2}(1 - 2p_0)^N. \end{aligned}$$

The last equality can be verified by applying the binomial formula to

$$[(1 - p_0) \pm x p_0]^N = \sum_{j=0}^{N} \binom{N}{j} p_0^j (1 - p_0)^{N-j} (\pm x)^j.$$

Then

$$\begin{aligned} 2 \sum_{j \text{ even}} \binom{N}{j} p_0^j (1 - p_0)^{N-j} &= (1 - p_0 + 1 \cdot p_0)^N + (1 - p_0 - 1 \cdot p_0)^N \\ &= 1 + (1 - 2p_0)^N \end{aligned}$$

(S. P. Lloyd). Similarly, for $\mathbf{z} \in \{0, 1\}^N$ of odd parity,

$$\begin{aligned} \Pr\{S_1 = 0 | \mathbf{Z}^N = \mathbf{z}\} &= \Pr\left\{\begin{matrix}\text{the BSC makes an} \\ \text{odd number of errors}\end{matrix}\right\} \\ &= \tfrac{1}{2} - \tfrac{1}{2}(1 - 2p_0)^N. \end{aligned}$$

Therefore, for all $\mathbf{z} \in \{0, 1\}^N$,

$$H(S_1 | \mathbf{Z}^N = \mathbf{z}) = h[\tfrac{1}{2} - \tfrac{1}{2}(1 - 2p_0)^N],$$

so that

$$\begin{aligned} \Delta = H(S_1 | \mathbf{Z}^N) &= h[\tfrac{1}{2} - \tfrac{1}{2}(1 - 2p_0)^N] \\ &\rightarrow 1 = H(S_1), \quad \text{as } N \rightarrow \infty. \end{aligned}$$

Thus, as $N \rightarrow \infty$, the equivocation at the wire-tap approaches the unconditional source entropy, so that communication is accomplished in perfect secrecy. The "catch" is that, as $N \rightarrow \infty$, the transmission rate $K/N = 1/N \rightarrow 0$.

A central question to which this paper is addressed is whether or not it is possible to transmit at a rate bounded away from zero, and yet achieve approximately perfect secrecy, i.e., $\Delta \approx H(S_1)$. Before giving the answer to this question, we shall describe the more general problem that is addressed in the sequel.

Refer to Fig. 2. The source is discrete and memoryless with entropy H_S. The "main channel" and the "wire-tap channel" are discrete memoryless channels with transition probabilities $Q_M(\cdot|\cdot)$ and $Q_W(\cdot|\cdot)$, respectively. The source and the transition probabilities Q_M and Q_W are given and fixed. The encoder, as in the above example, is a channel with the K vector $\mathbf{S}^K$ as input and the N vector $\mathbf{X}^N$ as output. The vector $\mathbf{X}^N$ is in turn the input to the main channel. The main channel output and the wire-tap channel input is $\mathbf{Y}^N$. The wire-tap channel output is $\mathbf{Z}^N$. The decoder associates a K vector $\hat{\mathbf{S}}^K$ with $\mathbf{Y}^N$ and the error probability P_e is given by (1). The equivocation Δ is given by (2), and the transmission rate is KH_S/N source bits per channel input symbol. Roughly speaking, a pair (R, d) is achievable if it is possible to find an encoder-decoder with arbitrarily small P_e and KH_S/N about R, and Δ about d (with perhaps N and K very large). Our main problem is the characterization of the family of achievable (R, d) pairs, and such a characterization is given in Theorem 2. It turns out (Theorem 3) that, in nearly every case, there exists a "secrecy capacity," $C_s > 0$, such that (C_s, H_S) is achievable [while for $R > C_s$, (R, H_S) is not achievable]. Thus, it is possible to reliably transmit information at the positive rate C_s in essentially perfect secrecy.

For the special case of our introductory example ($H_S = 1$, Q_M corresponding to a noiseless channel and Q_W to a BSC), the conclusion of Theorem 2 specializes to the assertion that (R, d) is achievable if and only if $0 \leqq R \leqq 1$, $0 \leqq d \leqq 1$, and $Rd \leqq h(p_0)$. Note that scheme (*i*) suggested above for this special case asserts that $R = 1$, $d = h(p_0)$

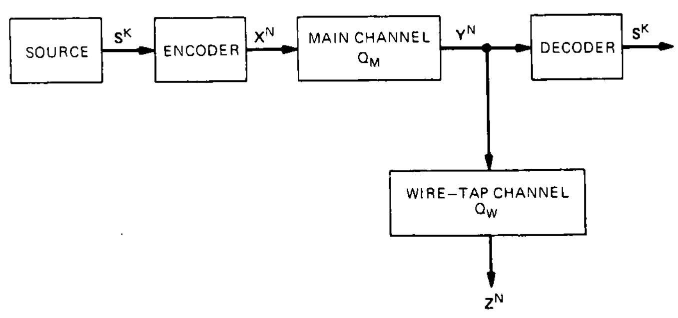

Fig. 2—Wire-tap channel (general case).

is achievable. From Theorem 2, this value of $d = h(p_0)$ is the maximum achievable d, if $R = 1$. Scheme (*ii*) above asserts that $R = 0$, $d = 1$ is achievable, but this is distinctly suboptimal since from Theorem 2, $R = h(p_0)$, $d = 1$ is achievable. Thus, reliable transmission at a rate $h(p_0)$ is possible with perfect secrecy, and $C_s = h(p_0)$.

An outline of the remainder of this paper now follows. In Section II, we give a formal statement of the problem and state the main results (Theorems 2 and 3). In Section III we give a proof of Theorem 2 for the special case discussed above (main channel noiseless, wire-tap channel a BSC). In Section IV, we prove the converse half of Theorem 2, and in Section V the direct half of that theorem.

II. FORMAL STATEMENT OF THE PROBLEM AND SUMMARY OF RESULTS

In this section we give a precise statement of the problem that we stated informally in Section I. We then summarize our results.

First, a word about notation. Let $\mathcal{U}$ be an arbitrary finite set. Denote its cardinality by $|\mathcal{U}|$. Consider $\mathcal{U}^N$, the set of N vectors with components in $\mathcal{U}$. The members of $\mathcal{U}^N$ will be written as

$$\mathbf{u}^N = (u_1, u_2, \cdots, u_N),$$

where subscripted letters denote the components and boldface superscripted letters denote vectors. A similar convention applies to random vectors and random variables, which are denoted by upper-case letters. When the dimension N of a vector is clear from the context, we omit the superscript.

For random variables X, Y, Z, etc., the notation $H(X)$, $H(X|Y)$, $I(X;Y)$, $I(X;Y|Z)$, etc., denotes the standard information quantities as defined in Gallager.[1] The logarithms in these quantities are, as are all logarithms in this paper, taken to the base 2. Finally, for $n = 3, 4, 5, \cdots$, we say that the sequence of random variables $\{X_i\}_{i=1}^n$ is a "Markov chain" if $(X_1, X_2, \cdots, X_{j-1})$ and $(X_{j+1}, \cdots, X_n)$ are conditionally independent, given $X_j (1 < j < n)$. We make repeated use of the fact that, if X_1, X_2, X_3 is a Markov chain, then

$$H(X_3|X_1, X_2) = H(X_3|X_2). \tag{4}$$

At this point we call attention to Appendix A, in which the data-processing theorem and Fano's inequality are given in several forms.

We now turn to the description of the communication system. We assume that the system designer is given a source and two channels that are defined as follows.

(*i*) The *source* is defined by the sequence $\{S_k\}_1^\infty$, where the S_k are independent, identically distributed random variables that take values in the finite set $\mathcal{S}$. We assume that the probability law that defines the $\{S_k\}$ is known. Let the entropy $H(S_k) = H_S$. In Appendix C we show how to extend the results of this paper to arbitrary stationary finite alphabet ergodic sources.

(*ii*) The *main channel* is a discrete memoryless channel with finite input alphabet $\mathcal{X}$, finite output alphabet $\mathcal{Y}$, and transition probability $Q_M(y|x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. Since the channel is memoryless, the transition probability for N vectors is

$$Q_M^{(N)}(\mathbf{y}|\mathbf{x}) = \prod_{n=1}^{N} Q_M(y_n|x_n). \quad (5)$$

Denote the channel capacity of the main channel by C_M.

(*iii*) The *wire-tap channel* is also a discrete memoryless channel with input alphabet $\mathcal{Y}$, finite output alphabet $\mathcal{Z}$, and transition probability $Q_W(z|y)$, $y \in \mathcal{Y}$, $z \in \mathcal{Z}$. The cascade of the main channel and the wire-tap channel is another memoryless channel with transition probability

$$Q_{MW}(z|x) = \sum_{y \in \mathcal{Y}} Q_W(z|y) Q_M(y|x). \quad (6)$$

Occasionally, when there is no ambiguity, we use the transition probability of a channel to denote the channel itself. Let C_{MW} be the capacity of channel Q_{MW}.

With the source statistics and channels Q_M and Q_W given, the designer must specify an encoder and a decoder, defined as follows:

(*iv*) The *encoder* with parameters (K, N) is another channel with input alphabet $\mathcal{S}^K$, output alphabet $\mathcal{X}^N$, and transition probability $q_E(\mathbf{x}|\mathbf{s})$, $\mathbf{s} \in \mathcal{S}^K$, $\mathbf{x} \in \mathcal{X}^N$. When the K source variables $\mathbf{S}^K = (S_1, \cdots, S_K)$ are the input to the encoder, the output is the random vector $\mathbf{X}^N$. Let $\mathbf{Y}^N$ and $\mathbf{Z}^N$ be the output of channels $Q_M^{(N)}$ and $Q_{MW}^{(N)}$, respectively, when the input is $\mathbf{X}^N$. The equivocation of the source at the output of the wire-tap channel (corresponding to a particular encoder) is

$$\Delta \triangleq \frac{1}{K} H(\mathbf{S}^K|\mathbf{Z}^N). \quad (7)$$

We take Δ as our criterion of the wire-tapper's confusion. From the system designer's point of view, it is, of course, desirable to make Δ large.

(*v*) The *decoder* is a mapping

$$f_D: \mathcal{Y}^N \to \mathcal{S}^K. \quad (8a)$$

Let $\hat{\mathbf{S}} = (\hat{S}_1, \cdots, \hat{S}_K) = f_D(\mathbf{Y})$. Corresponding to a given encoder and decoder, the *error-rate* is

$$P_e = \frac{1}{K} \sum_{k=1}^{N} \Pr\{S_k \neq \hat{S}_k\}. \quad (8b)$$

We refer to the above as an encoder-decoder (K, N, Δ, P_e).* The applicability of the above to the system in Fig. 2 should be obvious.

Next, we say that the pair (R, d) (where $R, d > 0$) is *achievable* if, for all $\epsilon > 0$, there exists an encoder-decoder (N, K, Δ, P_e) for which

$$\frac{(H_S K)}{N} \geqq R - \epsilon, \quad (9a)$$

$$\Delta \geqq d - \epsilon, \quad (9b)$$

$$P_e \leqq \epsilon. \quad (9c)$$

Our problem is to characterize the set $\mathcal{R}$ of achievable (R, d) pairs. Let us remark here that it follows immediately from the definition that $\mathcal{R}$ is a closed subset of the first quadrant of the (R, d) plane. Before stating our characterization of $\mathcal{R}$, we digress to discuss a certain information-theoretic quantity that plays a crucial role in our solution.

Consider the channels Q_M, Q_W, and Q_{MW} defined above. Let $p_X(x)$, $x \in \mathcal{X}$, be a probability mass function and let X be the random variable defined by

$$\Pr\{X = x\} = p_X(x), \quad x \in \mathcal{X}.$$

Let Y, Z be the outputs of channels Q_M and Q_{MW}, respectively, when X is the input. For $R \geqq 0$, let $\mathcal{P}(R)$ be the set of p_X such that $I(X;Y) \geqq R$. Of course, $\mathcal{P}(R)$ is empty for $R > C_M$, the capacity of channel Q_M. Finally, for $0 \leqq R \leqq C_M$, define

$$\Gamma(R) \triangleq \sup_{p_X \in \mathcal{P}(R)} I(X;Y|Z). \quad (10)$$

We remark here that, for any distribution p_X on $\mathcal{X}$, the corresponding X, Y, Z forms a Markov chain, so that the definition of mutual information and (4) yield

$$\begin{aligned} I(X;Y|Z) &= H(X|Z) - H(X|Y,Z) \\ &= H(X|Z) - H(X|Y) = I(X;Y) - I(X;Z). \quad (11) \end{aligned}$$

Thus, we can write (10) as

$$\Gamma(R) = \sup_{p_X \in \mathcal{P}(R)} I(X;Y|Z) = \sup_{p_X \in \mathcal{P}(R)} [I(X;Y) - I(X;Z)]. \quad (12)$$

* This should be read as "... an encoder-decoder with parameters (K, N, Δ, P_e)."

As an example, suppose that $\mathfrak{X} = \mathfrak{Y} = \mathfrak{Z} = \{0, 1\}$. Let Q_M be a noiseless (binary) channel, and let Q_W be a binary symmetric channel (BSC) with crossover probability p_0. Then for arbitrary p_X,

$$\begin{aligned} I(X;Y) - I(X;Z) &= H(X) - [H(Z) - H(Z|X)] \\ &= h(p_0) + H(X) - H(Z) \leqq h(p_0), \end{aligned}$$

where $h(\cdot)$ is defined in (3). The inequality follows from the well-known fact (see, for example, Ref. 2) that the entropy of the output of a BSC, i.e., $H(Z)$, is not less than the entropy of the input, $H(X)$. Further, $H(X) = H(Z)$ if and only if $p_X(0) = p_X(1) = \frac{1}{2}$. Since this distribution belongs to $\mathcal{P}(R)$, for all R, $0 \leqq R \leqq C_M = 1$, we conclude that, in this case,

$$\Gamma(R) = h(p_0), \quad 0 \leqq R \leqq C_M. \tag{13}$$

In Appendix B, we establish the following lemma concerning $\Gamma(R)$.

Lemma 1: The quantity $\Gamma(R)$, $0 \leqq R \leqq C_M$, satisfies the following:

- (*i*) *The "supremum" in the definition of Γ[(10) or (12)] is, in fact, a maximum—i.e., for each R, there exists a $p_X \in \mathcal{P}(R)$ such that $I(X;Y|Z) = \Gamma(R)$.*
- (*ii*) *$\Gamma(R)$ is a concave function of R.*
- (*iii*) *$\Gamma(R)$ is nonincreasing in R.*
- (*iv*) *$\Gamma(R)$ is continuous in R.*
- (*v*) *$C_M \geqq \Gamma(R) \geqq C_M - C_{MW}$, where C_M and C_{MW} are the capacities of channels Q_M and Q_{MW}, respectively.*

We can now state our main result, the proof of which is given in the remaining sections.

Theorem 2: The set $\mathcal{R}$, as defined above, is equal to $\overline{\mathcal{R}}$, where

$$\overline{\mathcal{R}} \triangleq \{(R, d): \ 0 \leqq R \leqq C_M, \quad 0 \leqq d \leqq H_S, \quad Rd \leqq H_S\Gamma(R)\}. \tag{14}$$

Remarks:

(1) A sketch of a typical region $\overline{\mathcal{R}}$ is given in Fig. 3. In the above example (Q_M noiseless and Q_W a BSC), $\Gamma(R) = h(p_0)$, a constant, so that the curve $Rd = H_S\Gamma(R)$ is a hyperbola. Observe that in this case the region $\overline{\mathcal{R}}$ is not convex. This is in contrast to the up-to-now essentially universal situation in multiple-user Shannon theory problems, where the solution is nearly always a convex region. Whether or not $\Gamma(R)/R$ is always convex, as it appears in Fig. 3, is an open question.

(2) The points in $\overline{\mathcal{R}}$ for which $R = C_M$ correspond to data rates of about the capacity of Q_M. This is clearly the maximum rate at which reliable transmission over Q_M is possible. An equivocation at the wire-tap of about $H_S\Gamma(C_M)/C_M$ is achievable at this rate. An increase in equivocation requires a reduction of transmission rate.

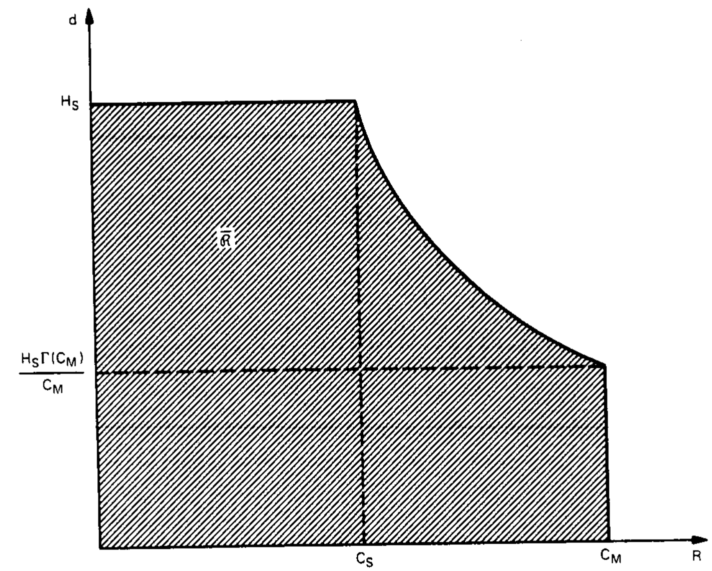

Fig. 3—Region $\overline{\mathcal{R}}$.

(3) The points in $\overline{\mathcal{R}}$ for which $d = H_S$ are of considerable interest. These correspond to an equivocation for the wire-tapper of about H_S—i.e., perfect secrecy. A transmission rate of

$$C_s = \max_{(R, H_S) \in \overline{\mathcal{R}}} R$$

is therefore achievable in perfect secrecy. We call C_s the "secrecy capacity" of the channel pair (Q_M, Q_W). The following theorem clarifies this remark.

Theorem 3: If $C_M > C_{MW}$, there exists a unique solution C_s of

$$C_s = \Gamma(C_s). \tag{15}$$

Further, C_s satisfies

$$0 < C_M - C_{MW} \leqq \Gamma(C_M) \leqq C_s \leqq C_M, \tag{16}$$

and C_s is the maximum R such that $(R, H_S) \in \mathcal{R}$.

Proof: Define $G(R) = \Gamma(R) - R$, $0 \leqq R \leqq C_M$. From Lemma 1 (*v*),

$$G(C_M) = \Gamma(C_M) - C_M \leqq 0,$$

and

$$G(0) = \Gamma(0) \geqq C_M - C_{MW} > 0.$$

Since by Lemma 1, (*iii*) and (*iv*), $G(R)$ is continuous and strictly decreasing in R, a unique $C_s \in (0, C_M]$ exists such that $G(C_s) = \Gamma(C_s) - C_s = 0$. This is the unique solution to (15). Inequality (16) follows from $C_s \in (0, C_M]$ and Lemma 1, (*iii*) and (*v*). Finally, from (15) and (16) we have $(C_s, H_S) \in \bar{\mathcal{R}} = \mathcal{R}$. Also, if $(R_1, H_S) \in \mathcal{R}$, then $H_S R_1 \leqq H_S \Gamma(R_1)$ so that $G(R_1) \geqq 0$. Since $G(R)$ is strictly decreasing in R, we conclude that $R_1 \leqq C_s$. Thus, C_s is the maximum of those R for which $(R_1, H_S) \in \mathcal{R}$, completing the proof.

(4) It is clear that the source statistics enter into the solution only via the source entropy H_S. We also remind the reader that the fairly simple extension of Theorems 2 and 3 to a stationary, ergodic source is given in Appendix C.

(5) If we define P_{ew}, the "wire-tapper's" error probability, as the error rate at a decoder built by the wire-tapper [defined analogously to (8)], then it follows from Fano's inequality (see Appendix A) that

$$\Delta \leqq h(P_{ew}) + P_{ew} \log |\mathcal{S}|.$$

Thus, a large value of the equivocation Δ implies a large value of P_{ew} (which the system designer will find desirable).

III. PROOF OF THEOREM 2 FOR A SPECIAL CASE

In this section we prove Theorem 2 for the very special case discussed in Section I. All alphabets $\mathcal{S}$, $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ are equal to $\{0, 1\}$. The source $\{S_k\}$ satisfies $\Pr\{S_k = 0\} = \Pr\{S_k = 1\} = \frac{1}{2}$. Channel Q_M is noiseless, i.e., $Q_M(y|x) = \delta_{x,y}$; and channel Q_W is a BSC with crossover probability p_0 $(0 \leqq p_0 \leqq \frac{1}{2})$, i.e.,

$$Q_W(z|y) = (1 - p_0)\delta_{y,z} + p_0(1 - \delta_{y,z}). \tag{17}$$

We show here that (R, d) is achievable if and only if

$$R \leqq C_M = 1, \quad d \leqq H_S = 1, \quad Rd \leqq h(p_0). \tag{18}$$

Since, for this case, $\Gamma(R) = h(p_0)$, this result is a special case of the as-yet-unproven Theorem 2. We begin with the converse ("only if") part of the result. Let $\mathbf{S}^K$, $\mathbf{X}^N$, $\mathbf{Z}^N$ correspond to an encoder-decoder (N, K, Δ, P_e) (note that $\mathbf{Y}^N = \mathbf{X}^N$). Then, making repeated use of the identity $H(U, V) = H(U) + H(V|U)$, we can write (dropping the superscript on vectors)

$$\begin{aligned} K\Delta &= H(\mathbf{S}^K|\mathbf{Z}^N) = H(\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\ &= H(\mathbf{S}, \mathbf{X}, \mathbf{Z}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\ &= H(\mathbf{Z}|\mathbf{X}, \mathbf{S}) + H(\mathbf{X}, \mathbf{S}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\ &\overset{(a)}{=} H(\mathbf{Z}|\mathbf{X}) + H(\mathbf{S}|\mathbf{X}) + H(\mathbf{X}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\ &\overset{(b)}{=} Nh(p_0) + H(\mathbf{S}|\mathbf{X}) + [H(\mathbf{X}) - H(\mathbf{Z})] - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}). \end{aligned} \tag{19}$$

These steps are justified as follows.

(a) From the fact that $(\mathbf{S}, \mathbf{X}, \mathbf{Z})$ is a Markov chain and (4), so that $H(\mathbf{Z}|\mathbf{X}, \mathbf{S}) = H(\mathbf{Z}|\mathbf{X})$.

(b) Since $\mathbf{X}$, $\mathbf{Z}$ are the input and output, respectively, of a BSC, $H(\mathbf{Z}|\mathbf{X}) = Nh(p_0)$, regardless of the distribution for $\mathbf{X}$.

Now from Fano's inequality [use ineq. (78) with $V = \mathbf{X}$], we have $H(\mathbf{S}|\mathbf{X}) \leqq Kh(P_e)$. Further, the entropy of the output of a BSC $\geqq$ the entropy of the input [this follows from Mrs. Gerber's lemma (Ref. 2, Theorem 1)], so that $H(\mathbf{X}) - H(\mathbf{Z}) \leqq 0$. Finally, $H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) \geqq 0$. Thus, (19) yields for any encoder-decoder (K, N, Δ, P_e),

$$K\Delta \leqq Nh(p_0) + Kh(P_e),$$

or

$$\frac{K}{N}[\Delta - h(P_e)] \leqq h(p_0). \tag{20}$$

Now suppose that (R, d) is achievable. It follows from the ordinary converse to the coding theorem (Ref. 1, Th. 4.3.4, p. 81) that $R \leqq C_M = 1$. Further, since $\Delta \leqq H_S = 1$, we conclude that $d \leqq 1$. Finally, if we apply (20) to an encoder-decoder (N, K, Δ, P_e) that satisfies (9) with $\epsilon > 0$ arbitrary, we have

$$(R - \epsilon)[(d - \epsilon) - h(\epsilon)] \leqq h(p_0).$$

Letting $\epsilon \to 0$ yields $Rd \leqq h(p_0)$. Thus, we have established the converse of Theorem 2, i.e., that an achievable (R, d) must satisfy (18).

We begin the proof of the direct half of Theorem 2 with a digression about group codes for the BSC. Let $G \subseteq \{0, 1\}^N$ be a group code (i.e., a parity check code) as defined for example in Ref. 1, Chapter 6, or Ref. 3, Chapter 4. The group code G has $M = 2^N/|G|$ cosets. Denote the cosets by $C_0 = G, C_1, C_2, \cdots, C_{M-1}$. Of course, the cosets are disjoint and

$$\bigcup_{i=0}^{M-1} C_i = \{0, 1\}^N.$$

Let λ be the word error probability when group code G (or for any of the cosets) is used on a BSC with crossover probability p_0, with maximum-likelihood (minimum distance) decoding. Thus, for each coset C_i, $0 \leqq i \leqq M - 1$, there exists a decoder mapping $D_i: \{0, 1\}^N \to C_i$, such that if $\mathbf{X}^N$ is the input to a BSC with crossover probability p_0, and $\mathbf{Z}^N$ is the corresponding output, then for all $\mathbf{x} \in C_i$, $0 \leqq i \leqq M - 1$,

$$\Pr\{D_i(\mathbf{Z}^N) \neq \mathbf{X}^N | \mathbf{X}^N = \mathbf{x}\} = \lambda.$$

Thus, regardless of the probability distribution for $\mathbf{X}^N$,

$$\Pr\{D_i(\mathbf{Z}^N) \neq \mathbf{X}^N | \mathbf{X}^N \in C_i\} = \lambda.$$

Letting $\psi(\mathbf{x}) = i$, for $\mathbf{x} \in C_i$, $0 \leqq i \leqq M - 1$, we have, from Fano's inequality [use ineq. (76) with $U = \mathbf{X}^N$, $V = \mathbf{Z}^N$, $\hat{U} = D_i(Z^N)$],

$$H(\mathbf{X}^N | \mathbf{Z}^N, \psi = i) \leqq h(\lambda) + \lambda \log |C_i|.$$

Therefore, for any $\mathbf{X}$ distribution (which induces a distribution of ψ),

$$H(\mathbf{X}^N | \mathbf{Z}^N, \psi) \leqq h(\lambda) + \lambda \log |G|. \tag{21}$$

We conclude this digression by stating as a lemma the well-known result of Elias that there exists a group code for transmitting reliably over a BSC at any rate up to capacity. A proof of this result can be found in Ref. 1, Section 6.2.

Lemma 4: Let $\epsilon_1 > 0$, $r < 1 - h(p_0)$ be arbitrary. Then, provided N is sufficiently large, there exists a group code G of block length N with $|G| \geqq 2^{Nr}$, such that, on the BSC *with crossover probability p_0, the error probability $\lambda \leqq \epsilon_1$.*

We now prove the direct half of Theorem 2 for our special case by showing that any (R, d), where R is rational, which satisfies

$$R \cdot d = h(p_0), \tag{22a}$$

$$0 \leqq d < 1, \tag{22b}$$

$$0 \leqq R \leqq 1 \tag{22c}$$

is achievable. Thus, for (R, d) satisfying (22), and arbitrary $\epsilon > 0$, we must show the existence of an encoder-decoder (N, K, Δ, P_e) that satisfies (9). We now proceed to this task.

Let K, N satisfy

$$\frac{K}{N} = R. \tag{23}$$

Let G be a binary group code with block length N and with $|G| = 2^{(N-K)}$. Thus, G has $M = 2^K$ cosets $\{C_i\}_{i=0}^{M}$. We can assume that the set $\mathcal{S}^K = \{0, 1\}^K$ is the set of integers $\{0, 1, \cdots, M - 1\}$. We construct the encoder such that when the source vector $\mathbf{S}^K = i$,* the encoder output $\mathbf{X}^N$ is a randomly chosen member of coset C_i—i.e.,

$$\Pr\{\mathbf{X}^N = \mathbf{x} | \mathbf{S} = i\} = \begin{cases} \dfrac{1}{|C_i|} = \dfrac{1}{|G|} = 2^{-(N-K)}, & \text{for } \mathbf{x} \in C_i, \\ 0, & x \notin C_i, \end{cases}$$

$0 \leqq i \leqq M - 1$. Since $\mathbf{S}^K$ is uniformly distributed on $\{0, 1, \cdots, M - 1\}$, $\mathbf{X}^N$ is uniformly distributed on $\mathfrak{X}^N = \{0, 1\}^N$. Thus, in particular,

$$H(\mathbf{X}^N) = H(\mathbf{Z}^N) = N, \tag{24}$$

where, as always, $\mathbf{Z}^N$ is the output of the wire-tap channel when $\mathbf{X}^N$ is the input. Also let us observe here that the quantity $\psi(\mathbf{X}^N)$, defined in the above digression, is identical to $\mathbf{S}^K$. Thus, (21) yields

$$H(\mathbf{X}^N | \mathbf{Z}^N, \mathbf{S}^K) \leqq h(\lambda) + \lambda(N - K), \tag{25}$$

where λ is the error probability for the group code G.

We now turn to the decoder. Letting $D(\mathbf{y}) = i$, when $\mathbf{y} \in C_i$, we conclude (since the channel Q_M is noiseless) that

$$P_e = 0. \tag{26}$$

Since (23) and (26) imply (9a) and (9c), it remains to show that a G exists such that the resulting encoder-decoder will satisfy (9b).

We now invoke (19), which is valid for any encoder-decoder. Substituting (24) and (25) into (19), and invoking (26), which implies $H(\mathbf{S}|\mathbf{X}) = 0$, we obtain

$$\Delta \geqq \left(\frac{N}{K}\right) h(p_0) - \frac{h(\lambda)}{K} - \lambda\left(\frac{N}{K} - 1\right). \tag{27}$$

Now, from (22a) and (23), we have

$$\frac{N}{K} h(p_0) = \frac{h(p_0)}{R} = d,$$

and from (23),

$$\lambda\left(\frac{N}{K} - 1\right) = \lambda\left(\frac{1}{R} - 1\right).$$

Thus, (27) yields

$$\Delta \geqq d - \left[\frac{h(\lambda)}{K} + \lambda\left(\frac{1}{R} - 1\right)\right]. \tag{28}$$

Finally, since from (23) and (22a) we have

$$|G| = 2^{N-K} \leqq 2^{N[1-h(p_0)/d]},$$

we can invoke Lemma 4 with $r = 1 - h(p_0)/d < 1 - h(p_0)$ [from (22b)] to assert the existence of a group code G with λ sufficiently small to make the term in brackets in (28) $\leqq \epsilon$. Then $\Delta \geqq d - \epsilon$, which is (9b). This completes the proof of the direct half.

IV. CONVERSE THEOREM

In this section, we establish the converse theorem that the family of achievable rates $\mathfrak{R}$ is contained in $\overline{\mathfrak{R}}$ as defined in (14). Suppose that

* This is an abuse of notation. A more precise statement is that $\mathbf{S}^K$ is a binary representation of i.

$(R, d) \in \mathcal{R}$. That $R \leqq C_M$ follows from the ordinary converse to the coding theorem (Ref. 1, Theorem 4.3.4, p. 81). That $d \leqq H_S$ follows from

$$\Delta = \frac{1}{K} H(\mathbf{S}^K | \mathbf{Z}^N) \leqq \frac{1}{K} H(\mathbf{S}^K) = H_S.$$

Thus, it remains to show that $Rd \leqq H_S \Gamma(R)$. We do this via a lemma, the proof of which is given at the conclusion of this section.

Lemma 5: Let $\mathbf{S}^K$, $\mathbf{X}^N$, $\mathbf{Y}^N$, $\mathbf{Z}^N$ *correspond to an encoder-decoder* (N, K, Δ, P_e). *Then*

$$(i) \quad \frac{K}{N}[\Delta - \delta(P_e)] \leqq \frac{1}{N} \sum_{n=1}^{N} I(X_n; Y_n | Z_n, \mathbf{Y}^{n-1}), \tag{29a}$$

$$(ii) \quad \frac{K}{N}[H_S - \delta(P_e)] \leqq \frac{1}{N} \sum_{n=1}^{N} I(X_n; Y_n | \mathbf{Y}^{n-1}), \tag{29b}$$

where

$$\delta(P_e) = h(P_e) + P_e \log |\mathcal{S}|, \tag{29c}$$

and where the $n = 1$ *term in the summations of (29a, b) is given the obvious interpretation—i.e., that* $I(X_1; Y_1 | Z_1, \mathbf{Y}^0) = I(X_1; Y_1 | Z_1)$, *etc.*

Now for $n = 2, 3, \cdots, N$, any $\mathbf{y} \in \mathcal{Y}^{n-1}$, set

$$\alpha_n(\mathbf{y}) = I(X_n; Y_n | \mathbf{Y}^{n-1} = \mathbf{y}). \tag{30a}$$

Also let

$$\alpha_1 = I(X_1; Y_1). \tag{30b}$$

It follows from the definition of $\mathcal{P}(R)$ in Section II that the distribution p_1, defined by

$$p_1(x) \triangleq \Pr\{X_1 = x\}, \quad x \in \mathfrak{X},$$

belongs to $\mathcal{P}(\alpha_1)$. Similarly, for $2 \leqq n \leqq N$, with $\mathbf{y} \in \mathcal{Y}^{n-1}$ fixed, define

$$p_{n,\mathbf{y}}(x) \triangleq \Pr\{X_n = x | \mathbf{Y}^{n-1} = \mathbf{y}\}, \quad x \in \mathfrak{X}.$$

Then $p_{n,\mathbf{y}} \in \mathcal{P}[\alpha_n(\mathbf{y})]$. Thus, from (10) and the fact that channels $Q_M^{(N)}$ and $Q_W^{(N)}$ are memoryless,

$$\Gamma(\alpha_1) \geqq I(X_1; Y_1 | Z_1), \tag{31a}$$

and for $2 \leqq n \leqq N$, $\mathbf{y} \in \mathcal{Y}^{n-1}$,

$$\Gamma[\alpha_n(\mathbf{y})] \geqq I(X_n; Y_n | Z_n, \mathbf{Y}^{n-1} = \mathbf{y}). \tag{31b}$$

It follows that the right member of (29a) is (giving the $n = 1$ term the obvious interpretation)

$$\begin{aligned}
\frac{1}{N} \sum_{n=1}^{N} I(X_n; Y_n | Z_n, \mathbf{Y}^{n-1})
&= \frac{1}{N} \sum_{n=1}^{N} \sum_{\mathbf{y} \in \mathcal{Y}^{n-1}} \Pr\{\mathbf{Y}^{n-1} = \mathbf{y}\} I(X_n; Y_n | Z_n, \mathbf{Y}^{n-1} = \mathbf{y}) \\
&\overset{(a)}{\leqq} \frac{1}{N} \sum_n \sum_{\mathbf{y}} \Pr\{\mathbf{Y}^{n-1} = \mathbf{y}\} \Gamma[\alpha_n(\mathbf{y})] \\
&\overset{(b)}{\leqq} \Gamma\left[\frac{1}{N} \sum_n \sum_{\mathbf{y}} \Pr\{\mathbf{Y}^{n-1} = \mathbf{y}\} \alpha_n(\mathbf{y})\right] \\
&\overset{(c)}{=} \Gamma\left(\frac{1}{N} \sum_n I(X_n Y_n | \mathbf{Y}^{n-1})\right) \\
&\overset{(d)}{\leqq} \Gamma\left(\frac{K}{N} H_S - \delta(P_e)\right).
\end{aligned} \tag{32}$$

Step (a) follows from (31), step (b) from the concavity of Γ [Lemma 1(*ii*)], step (c) from the definition of α_n, and step (*d*) from (29b) and the monotonicity of Γ [Lemma 1(*iii*)]. Applying (29a) to (32) yields

Corollary 6: For any encoder-decoder (N, K, Δ, P_e),

$$\frac{K}{N}[\Delta - \delta(P_e)] \leqq \Gamma\left[\frac{K}{N} H_S - \delta(P_e)\right]. \tag{33}$$

We now show that, if $(R, d) \in \mathcal{R}$, then $Rd \leqq H_S \Gamma(R)$. Let $(R, d) \in \mathcal{R}$, and let $\epsilon > 0$ be arbitrary. Apply Corollary 6 to the encoder-decoder (N, K, Δ, P_e) that satisfies (9). Inequalities (33) and (9) yield

$$(R - \epsilon)[(d - \epsilon) - \delta(\epsilon)] \leqq H_S \Gamma[(R - \epsilon) - \delta(\epsilon)]. \tag{34}$$

Letting $\epsilon \to 0$ and invoking the continuity of Γ [Lemma 1(*iv*)] yield $Rd \leqq H_S \Gamma(R)$, completing the proof of the converse. It remains to prove Lemma 5.

Proof of Lemma 5:

(*i*) Let $\mathbf{S}^K$, $\mathbf{X}^N$, $\mathbf{Y}^N$, $\mathbf{Z}^N$ correspond to an encoder-decoder (N, K, Δ, P_e). First observe that

$$\begin{aligned}
\frac{1}{K} H(\mathbf{S}^K | \mathbf{Z}^N, \mathbf{Y}^N) &\leqq \frac{1}{K} H(\mathbf{S}^K | \mathbf{Y}^N) \\
&\overset{(a)}{\leqq} h(P_e) + P_e \log(|\mathcal{S}| - 1) = \delta(P_e).
\end{aligned} \tag{35}$$

Inequality (a) follows from Fano's inequality [use (78) with $V = \mathbf{Y}^N$].

Next, using the definition of Δ (7) and (35), write

$$\begin{aligned}
K\Delta = H(\mathbf{S}^K | \mathbf{Z}^N) &\leqq H(\mathbf{S}^K | \mathbf{Z}^N) - H(\mathbf{S}^K | \mathbf{Z}^N, \mathbf{Y}^N) + K\delta(P_e) \\
&= I(\mathbf{S}^K; \mathbf{Y}^N | \mathbf{Z}^N) + K\delta(P_e) \\
&\leqq I(\mathbf{X}^K; \mathbf{Y}^N | \mathbf{Z}^N) + K\delta(P_e).
\end{aligned} \tag{36}$$

The last inequality in (36) follows from the data-processing theorem, since given $\mathbf{Z}^N = \mathbf{z}$, $(\mathbf{Y}^N, \mathbf{X}^N, \mathbf{S}^K)$ is a Markov chain (Appendix A). Transposing the $K\delta(P_e)$ term in (36) and continuing:

$$
\begin{aligned}
K[\Delta - \delta(P_e)] &\leqq I(\mathbf{X}^N; \mathbf{Y}^N | \mathbf{Z}^N) \\
&= H(\mathbf{X}^N | \mathbf{Z}^N) - H(\mathbf{X}^N | \mathbf{Z}^N, \mathbf{Y}^N) \\
&\overset{(a)}{=} H(\mathbf{X}^N | \mathbf{Z}^N) - H(\mathbf{X}^N | \mathbf{Y}^N) \\
&= I(\mathbf{X}^N; \mathbf{Y}^N) - I(\mathbf{X}^N; \mathbf{Z}^N) \\
&= H(\mathbf{Y}^N) - H(\mathbf{Z}^N) + H(\mathbf{Z}^N | \mathbf{X}^N) - H(\mathbf{Y}^N | \mathbf{X}^N) \\
&\overset{(b)}{=} \sum_{n=1}^{N} [H(Y_n | \mathbf{Y}^{n-1}) - H(Z_n | \mathbf{Z}^{n-1}) \\
&\qquad\qquad + H(Z_n | X_n) - H(Y_n | X_n)] \\
&\overset{(c)}{\leqq} \sum_{n=1}^{N} [H(Y_n | \mathbf{Y}^{n-1}) - H(Z_n | \mathbf{Z}^{n-1}, \mathbf{Y}^{n-1}) \\
&\qquad\qquad + H(Z_n | X_n) - H(Y_n | X_n)] \\
&\overset{(d)}{=} \sum_{n=1}^{N} [H(Y_n | \mathbf{Y}^{n-1}) - H(Z_n | \mathbf{Y}^{n-1}) + H(Z_n | X_n, \mathbf{Y}^{n-1}) \\
&\qquad\qquad + H(Y_n | X_n, \mathbf{Y}^{n-1})] \\
&= \sum_{n=1}^{N} [I(X_n, Y_n | \mathbf{Y}^{n-1}) - I(X_n; Z_n | \mathbf{Y}^{n-1})] \\
&= \sum_{n=1}^{N} [H(X_n | Z_n, \mathbf{Y}^{n-1}) - H(X_n | Y_n, \mathbf{Y}^{n-1})] \\
&\overset{(e)}{=} \sum_{n=1}^{N} [H(X_n | Z_n, \mathbf{Y}^{n-1}) - H(X_n | Y_n, Z_n, \mathbf{Y}^{n-1})] \\
&= \sum_{n=1}^{N} I(X_n; Y_n | Z_n, \mathbf{Y}^{n-1}).
\end{aligned} \tag{37}
$$

The steps in (37) that require explanation are:

(a) that follows from the fact that $\mathbf{X}^N$, $\mathbf{Y}^N$, $\mathbf{Z}^N$ is a Markov chain and (4);

(b) that follows from the standard identity

$$H(\mathbf{U}^N) = \sum_{n=1}^{N} H(U_n | \mathbf{U}^{n-1}),$$

and the fact that channels $Q_M^{(N)}$ and $Q_{MW}^{(N)}$ are memoryless;

(c) that follows from the fact that conditioning decreases entropy;

(d) that follows on applying (4) to the Markov chains $(\mathbf{Z}^{n-1}, \mathbf{Y}^{n-1}, Z_n)$, $(\mathbf{Y}^{n-1}, X_n, Y_n, Z_n)$;

(e) that follows from the fact that, given $\mathbf{Y}^{n-1}$, (X_n, Y_n, Z_n) is a Markov chain.

Since (37) is (29a), we have established part (*i*) of Lemma 5.

(*ii*) With $\mathbf{S}^K$, $\mathbf{X}^N$, $\mathbf{Y}^N$, $\mathbf{Z}^N$, as in part (*i*) write

$$
\begin{aligned}
H(\mathbf{S}^K) &= I(\mathbf{S}^K; \mathbf{Y}^N) + H(\mathbf{S}^K | \mathbf{Y}^N) \\
&\leqq I(\mathbf{X}^N; \mathbf{Y}^N) + K\delta(P_e),
\end{aligned} \tag{38}
$$

where the inequality follows from the data-processing theorem (since $\mathbf{S}^K$, $\mathbf{X}^N$, $\mathbf{Y}^N$, is a Markov chain) and from Fano's inequality as in (35). Since $H(\mathbf{S}^K) = KH_S$, (38) yields

$$
\begin{aligned}
K[H_S - \delta(P_e)] &\leqq I(\mathbf{X}^N; \mathbf{Y}^N) \\
&\overset{(a)}{=} \sum_{n=1}^{N} [H(Y_n | \mathbf{Y}^{n-1}) - H(Y_n | X_n)] \\
&\overset{(b)}{=} \sum_{n=1}^{N} [H(Y_n | \mathbf{Y}^{n-1}) - H(Y_n | X_n, \mathbf{Y}^{n-1})] \\
&= \sum_{n=1}^{N} I(X_n; Y_n | \mathbf{Y}^{n-1}).
\end{aligned} \tag{39}
$$

Step (a) follows on application of $H(\mathbf{Y}^N) = \sum_n H(Y_n | \mathbf{Y}^{n-1})$, and the memorylessness of channel $Q_M^{(N)}$, and step (b) from the fact that $\mathbf{Y}^{n-1}$, X_n, Y_n is a Markov chain. Inequality (39) is (29b), so that the proof of Lemma 5 is complete.

V. DIRECT HALF OF THEOREM 2

In this section we establish the direct (existence) part of Theorem 2, that is, $\bar{\mathcal{R}} \subseteq \mathcal{R}$. The first step is to establish two lemmas that are valid for any encoder-decoder as defined in Section II.

Lemma 7: Let $\mathbf{S}^K$, $\mathbf{X}^N$, $\mathbf{Y}^N$, $\mathbf{Z}^N$ *correspond to an arbitrary encoder-decoder* (N, K, Δ, P_e). *Then*

$$K\Delta \triangleq H(\mathbf{S}^K | \mathbf{Z}^N) = H(\mathbf{S}^K) + I(\mathbf{X}^N; \mathbf{Z}^N | \mathbf{S}^K) - I(\mathbf{X}^N; \mathbf{Z}^N). \tag{40}$$

Proof: By repeatedly using the identity $H(U, V) = H(U) + H(V | U)$, we obtain (we have omitted superscripts)

$$
\begin{aligned}
K\Delta &= H(\mathbf{S} | \mathbf{Z}) = H(\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\
&= H(\mathbf{S}, \mathbf{Z}, \mathbf{X}) - H(\mathbf{X} | \mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\
&= H(\mathbf{Z} | \mathbf{X}, \mathbf{S}) + H(\mathbf{X}, \mathbf{S}) - H(\mathbf{X} | \mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\
&= H(\mathbf{Z} | \mathbf{X}, \mathbf{S}) + H(\mathbf{S}) + [H(\mathbf{X} | \mathbf{S}) - H(\mathbf{X} | \mathbf{S}, \mathbf{Z})] - H(\mathbf{Z}) \\
&= H(\mathbf{S}) + I(\mathbf{X}; \mathbf{Z} | \mathbf{S}) - [H(\mathbf{Z}) - H(\mathbf{Z} | \mathbf{X}, \mathbf{S})].
\end{aligned} \tag{41}
$$

Now, since $\mathbf{S}$, $\mathbf{X}$, $\mathbf{Z}$ is a Markov chain, $H(\mathbf{Z}|\mathbf{X},\mathbf{S}) = H(\mathbf{Z}|\mathbf{X})$ [by (4)]. Thus, the term in brackets in the right member of (41) is $I(\mathbf{X};\mathbf{Z})$, completing the proof.

We now give some preliminaries for the second of the two lemmas. For the remainder of this section we take the finite set $\mathfrak{X}$ to be $\{1, 2, \cdots, A\}$. Let X^* be a random variable that takes values in $\mathfrak{X}$ with probability distribution

$$\Pr\{X^* = i\} = p_X^*(i), \quad 1 \leqq i \leqq A.$$

Let Y^* and Z^* be the output of channels Q_M, and Q_{MW}, respectively, when X^* is the input. As always, Q_{MW} is the cascade of Q_M and Q_W, so that X^*, Y^*, Z^* is a Markov chain. Next, for $1 \leqq i \leqq A$, and $\mathbf{x} \in \mathfrak{X}^N$ define

$$\begin{aligned}\#(i,\mathbf{x}) &\triangleq \text{card}\{n: x_n = i\}\\ &= \text{number of occurrences of the symbol } i \text{ in the } N\text{-vector } \mathbf{x}.\end{aligned} \tag{42}$$

For $N = 1, 2, \cdots$, define the set of "typical" X sequences as the set

$$T^* = T^*(N) = \left\{\mathbf{x} \in \mathfrak{X}^N : \left|\frac{\#(i,\mathbf{x})}{N} - p_X^*(i)\right| \leqq \delta_N, 1 \leqq i \leqq A\right\}, \tag{43a}$$

where

$$\delta_N \triangleq N^{-\frac{1}{3}}. \tag{43b}$$

Let us remark in passing that the random N-vector $\mathbf{X}^{*N}$ consisting of N independent copies of X^* satisfies $E\#(i,\mathbf{X}^{*N}) = Np_X^*(i)$, and $\text{Var}[\#(i,\mathbf{X}^{*N})] = Np_X^*(i)[1 - p_X^*(i)]$, for $1 \leqq i \leqq A$. Thus, by Chebyshev's inequality

$$\begin{aligned}\Pr\{\mathbf{X}^{*N} \notin T^*(N)\} &\leqq \sum_{i=1}^{A} \Pr\{|\#(i,\mathbf{X}^*) - Np_X^*(i)| > N\delta_N\}\\ &\leqq \sum_{i=1}^{A} \text{Var}[\#(i,\mathbf{X}^*)]/N^2\delta_N^2 = 0\left(\frac{1}{\sqrt{N}}\right) \to 0,\end{aligned} \tag{44}$$

as $N \to \infty$.

We can now state the second of our lemmas. We give the proof at the conclusion of this section.

Lemma 8: Let $\mathbf{X}^N$, $\mathbf{Z}^N$ correspond to an arbitrary encoder and let X^*, Z^*, T^* correspond to an arbitrary p_X^* as above. Then

$$\frac{1}{N} I(\mathbf{X}^N;\mathbf{Z}^N) \leqq I(X^*,Z^*) + (\log A)\Pr\{\mathbf{X}^N \notin T^*(N)\} + f_1(N),$$

where $f_1(N) \to 0$, as $N \to \infty$.

Lemma 8 implies that, if the encoder is such that with high probability $\mathbf{X}^N \in T^*$, then $(1/N)I(\mathbf{X}^N;\mathbf{Z}^N)$ cannot be much more than $I(X^*, Z^*)$.

Lemmas 7 and 8 hold for any encoder-decoder. Our next step is to describe a certain ad-hoc encoder-decoder and deduce several of its properties. We then show that when the parameters of the ad-hoc scheme are properly chosen, the direct half of Theorem 2 will follow easily.

We begin the discussion of the ad-hoc scheme by reviewing some facts about source coding. With the source given as in Section II, for $K = 1, 2, \cdots$, there exists a ("source encoder") mapping F_E: $\mathcal{S}^K \to \{1, 2, \cdots, M\}$, where

$$M = 2^{KH_S(1+\delta_K)}, \tag{45}$$

and $\delta_K = K^{-\frac{1}{3}}$. Let F_D: $\{1, 2, \cdots, M\} \to \mathcal{S}^K$ be a ("source decoder") mapping, and let

$$P_{es}^{(K)} = \Pr\{F_D \circ F_E(\mathbf{S}^K) \neq \mathbf{S}^K\}$$

be the resulting error probability. It is very well known that there exists (for each K) a pair (F_E, F_D) such that, as $K \to \infty$,

$$P_{es}^{(K)} = \Pr\{F_D(W) \neq \mathbf{S}^K\} \to 0, \tag{46a}$$

where

$$W = F_E(\mathbf{S}^K). \tag{46b}$$

We will design our system to transmit W using an (F_E, F_D) that satisfies (46).

We now turn to our ad-hoc system. (Refer to Fig. 4.) The source output is the vector $\mathbf{S}^K$, and the output of the source decoder is $W = F_E(\mathbf{S}^K)$. Let

$$q_i \triangleq \Pr\{W = F_E(\mathbf{S}^K) = i\}, \quad 1 \leqq i \leqq M. \tag{47}$$

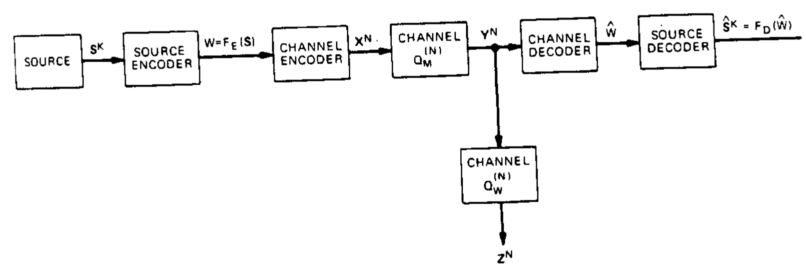

Fig. 4—Ad-hoc encoder-decoder.

Next, let $M_1 = M_2 M$ be a multiple of M to be specified later. Let

$$\{\mathbf{x}_m\}_1^{M_1}$$

be a subset of $\mathfrak{X}^N$. Clearly, $\{\mathbf{x}_m\}$ can be viewed as a channel code for channel $Q_M^{(N)}$ or channel $Q_{MW}^{(N)}$. The channel encoder and decoder in Fig. 4 work as follows. The channel encoder and decoder each contains a partition of $\{\mathbf{x}_m\}_1^{M_1}$ into M subcodes $C_1, C_2, \cdots, C_M$, each with cardinality M_2. Assume that

$$C_i = \{\mathbf{x}_{(i-1)M_2+1}, \cdots, \mathbf{x}_{iM_2}\}, \quad 1 \leqq i \leqq M. \tag{48}$$

When the random variable $W = i$, then the channel encoder output $\mathbf{X}^N$ is a (uniformly) randomly chosen member of the subcode C_i. Thus, for $1 \leqq i \leqq M$, $1 \leqq j \leqq M_2$,

$$\Pr\{\mathbf{X}^N = \mathbf{x}_{(i-1)M_2+j} \mid W = i\} = \frac{1}{M_2}, \tag{49a}$$

and

$$\Pr\{\mathbf{X}^N = \mathbf{x}_{(i-1)M_2+j}\} = \frac{q_i}{M_2}. \tag{49b}$$

Now the set $\{\mathbf{x}_m\}_1^{M_1}$ can be thought of as a channel code for channel $Q_M^{(N)}$ with prior probability distribution on the code words given by (49b). A decoder for the code is a mapping $G: \mathcal{Y}^N \to \{\mathbf{x}_m\}_1^{M_1}$ and the (word) error probability is

$$\lambda = \Pr\{G(\mathbf{Y}^N) \neq \mathbf{X}^N\}, \tag{50}$$

where $\mathbf{Y}^N$ is the output of $Q_M^{(N)}$, when the input $\mathbf{X}^N$ has distribution given by (49b). We assume that the channel decoder in Fig. 4 has stored the mapping G. When the channel output is $\mathbf{y} \in \mathcal{Y}^N$, the channel decoder computes $G(\mathbf{y})$. When $G(\mathbf{y}) \in C_i$, the channel decoder output is i, $1 \leqq i \leqq M$. Letting $\hat{W}$ be the output of the channel decoder, we have

$$\Pr\{W \neq \hat{W}\} \leqq \lambda.$$

The final step in the system of Fig. 4 is the emission by the source decoder of $\hat{\mathbf{S}}^K = F_D(\hat{W})$, where $F_D: \{1, 2, \cdots, M\} \to \mathcal{S}^K$ is chosen so that (46) holds. We have

$$\begin{aligned}\Pr\{\mathbf{S} = \hat{\mathbf{S}}\} &= \Pr\{\mathbf{S} = F_D(\hat{W})\} \\ &\geqq \Pr\{S = F_D(W); W = \hat{W}\}.\end{aligned}$$

Thus,

$$\begin{aligned}P_e \leqq \Pr\{\mathbf{S} \neq \hat{\mathbf{S}}\} &\leqq \Pr\{\mathbf{S} \neq F_D(\mathrm{W})\} \\ &\quad + \Pr\{W \neq \hat{W}\} \leqq P_{es}^{(K)} + \lambda. \end{aligned} \tag{51}$$

Next, let us observe that each of the subcodes C_i can be considered a code for channel $Q_{MW}^{(N)}$ with M_2 code words and uniform prior distribution on the code words. Let λ_i be the resulting (word) error probability for code C_i $(1 \leqq i \leqq M)$ with an optimal decoder, and let

$$\bar{\lambda} = \sum_{i=1}^{M} q_i \lambda_i. \tag{52}$$

We now establish

Lemma 9: For the ad-hoc encoder-decoder defined above

$$I(\mathbf{X}^N; \mathbf{Z}^N \mid \mathbf{S}^K) \geqq \log M_2 - [h(\bar{\lambda}) + \bar{\lambda} \log M_2].$$

Proof: Let $\mathbf{S}^K$ be such that $W = F_E(\mathbf{S}^K) = i$. Then the channel input $\mathbf{X}^N$ given $W = i$ has distribution given by (49a), i.e., $\mathbf{X}^N$ is a randomly chosen member of C_i. Since λ_i is the error probability for code C_i used on channel $Q_{MW}^{(N)}$, Fano's inequality [use (76) with $U = \mathbf{X}^N$, $V = \mathbf{Z}^N$, $\hat{U}$ = the decoded version of $\mathbf{Z}^N$ when code C_i is used] yields

$$H(\mathbf{X}^N \mid \mathbf{Z}^N, W = i) \leqq h(\lambda_i) + \lambda_i \log M_2,$$

and, since $H(\mathbf{X}^N \mid W = i) = \log M_2$, we have

$$I(\mathbf{X}^N; \mathbf{Z}^N \mid W = i) \geqq \log M_2 - h(\lambda_i) - \lambda_i \log M_2.$$

Averaging over i using the weighting $\{q_i\}$, and using the concavity of $h(\cdot)$, we have

$$I(\mathbf{X}^N; \mathbf{Z}^N \mid W) \geqq \log M_2 - [h(\bar{\lambda}) + \bar{\lambda} \log M_2]. \tag{53}$$

Finally, since $\mathbf{S}, W, \mathbf{X}, \mathbf{Z}$ is a Markov chain, (4) yields

$$\begin{aligned} I(\mathbf{X}^N; \mathbf{Z}^N \mid W) &= H(\mathbf{Z} \mid W) - H(\mathbf{Z} \mid \mathbf{X} W) \\ &= H(\mathbf{Z} \mid W, \mathbf{S}) - H(\mathbf{Z} \mid \mathbf{X}) \\ &= H(\mathbf{Z} \mid W, \mathbf{S}) - H(\mathbf{Z} \mid \mathbf{X}, \mathbf{S}) \\ &\leqq H(\mathbf{Z} \mid \mathbf{S}) - H(\mathbf{Z} \mid \mathbf{X}, \mathbf{S}) = I(\mathbf{X}^N; \mathbf{Z}^N \mid \mathbf{S}). \end{aligned} \tag{54}$$

Inequalities (53) and (54) imply Lemma 9.

We are now ready to combine the above lemmas as:

Corollary 10: Let p_X^* be an arbitrary probability distribution on $\mathfrak{X}$, and let $T_X^*(N)$, X^*, Y^*, Z^* be as defined above (corresponding to p_X^*). Assume that $\mathbf{S}^K$, $\mathbf{X}^N$, $\mathbf{Y}^N$, $\mathbf{Z}^N$ correspond to the above ad-hoc encoder-decoder with parameters N, K, M, M_1, M_2, λ, $\bar{\lambda}$. Let P_e and Δ correspond to this ad-hoc scheme. Then

$$P_e \leqq P_{es}^{(K)} + \lambda \tag{55a}$$

and

$$\frac{K}{N}\Delta \geqq \frac{K}{N}H_S + \frac{1}{N}\log M_2 - I(X^*, Z^*) - \frac{h(\bar\lambda)}{N} - \frac{\bar\lambda \log M_2}{N} - (\log A)\Pr\{\mathbf{X}^N \notin T_X^*(N)\} - f_1(N), \tag{55b}$$

where $f_1(N) \to 0$ *as* $N \to \infty$.

Proof: Inequality (55a) is the same as (51). Inequality (55b) is obtained by substituting the results of Lemmas 8 and 9 into (40) and using $H(\mathbf{S}^K) = KH_S$.

Finally, we are ready to prove the direct half of Theorem 2. We do this by showing that any pair (R, d), which satisfies

$$R\cdot d = H_S\Gamma(R), \tag{56a}$$

$$0 \leqq R \leqq C_M, \tag{56b}$$

$$0 \leqq d \leqq H_S, \tag{56c}$$

is achievable. Thus, for (R, d) satisfying (56) and for arbitrary $\epsilon > 0$, we show that our ad-hoc scheme with appropriately chosen parameters satisfies (9). To begin with, choose K, N to satisfy

$$\frac{K}{N} = \frac{R}{H_S}. \tag{57}$$

(Assume that R/H_S is rational.) Note that (57) implies (9a). Also, let p_X^* be a distribution on $\mathfrak{X}$ that belongs to $\mathcal{P}(R)$ and achieves $\Gamma(R)$—that is,

$$I(X^*; Y^*) \geqq R,$$
$$I(X^*; Y^*) - I(X^*; Z^*) = I(X^*; Y^*|Z^*) = \Gamma(R), \tag{58}$$

where X^*, Y^*, Z^* correspond to p_X^*. We now assume that an encoder-decoder is constructed according to the above ad-hoc scheme with the parameter*

$$M_1 = \exp_2\left\{N\left[I(X^*; Y^*) - \frac{\epsilon R}{2H_S}\right]\right\}, \tag{59}$$

where X^*, Y^* correspond to the above choice of p_X^*. With this choice of M_1, and with M given by (45), we have

$$M_2 = \frac{M_1}{M} = \exp_2\left\{N\left[I(X^*; Y^*) - \frac{K}{N}H_S - \frac{K}{N}H_S\delta_K - \frac{\epsilon R}{2H_S}\right]\right\}. \tag{60}$$

Note that, from (57),

$$\begin{aligned}\frac{1}{N}\log M_2 &= I(X^*; Y^*) - \frac{K}{N}H_S - \frac{K}{N}H_S\delta_K - \frac{\epsilon R}{2H_S}\\ &\overset{(a)}{=} I(X^*; Y^*) - R - R\delta_K - \frac{\epsilon R}{2H_S}\\ &= I(X^*; Y^*) - \frac{(Rd/H_S)}{(d/H_S)} - R\delta_K - \frac{\epsilon R}{2H_S}\\ &\overset{(b)}{\leqq} I(X^*; Y^*) - \Gamma(R) - R\delta_K - \frac{\epsilon R}{2H_S}\\ &= I(X^*; Y^*) - I(X^*; Y^*|Z^*) - R\delta_K - \frac{\epsilon R}{2H_S}\\ &\overset{(c)}{=} I(X^*; Z^*) - R\delta_K - \frac{\epsilon R}{2H_S}.\end{aligned} \tag{61}$$

Step (a) follows from (57), step (b) from (56a) and (56c), and step (c) from the fact that X^*, Y^*, Z^* is a Markov chain—see (11).

Let us now apply Corollary 10 to the ad-hoc scheme with the above choice of M_1, M_2, and with the above choice of p_X^*. Inequality (55a) remains

$$P_e \leqq P_{es}^{(K)} + \lambda, \tag{62}$$

and substituting (60) into (55b) yields

$$\begin{aligned}(R\Delta)/H_S &\geqq I(X^*; Y^*) - I(X^*; Z^*) - f_2(N)\\ &= \Gamma(R) - f_2(N),\end{aligned} \tag{63a}$$

where

$$f_2(N) = \frac{\epsilon R}{2H_S} + R\delta_K + \frac{h(\bar\lambda)}{N} + \frac{\bar\lambda\log M_2}{N} + (\log A)\Pr\{\mathbf{X}^N \notin T^*(N)\} + f_1(N). \tag{63b}$$

Now observe $f_2(N)$ and $\bar\lambda$ depend on the choice of the set $\{\mathbf{x}_m\}_1^{M_1}$. The following lemma asserts the existence of a $\{x_m\}$ such that these quantities are small. Its proof is given at the end of this section.

Lemma 11: With p_X^* *and* M_1, M_2 *as given above, there exists for arbitrary* N *a set*

$$\{\mathbf{x}_m\}_{m=1}^{M_1}$$

such that

$$\left.\begin{matrix}\Pr\{\mathbf{X}^N \notin T^*(N)\},\\ \lambda,\\ \bar\lambda\end{matrix}\right\} \leqq f_3(N), \tag{64}$$

where $f_3(N) \to 0$, *as* $N \to \infty$.

* Assume that the right member of (59) is an integer. If not, a trivial modification of the sequel is necessary.

Now let the set $\{\mathbf{x}_m\}_1^{M_1}$ in the ad-hoc scheme be chosen to satisfy (64). Then, from (62) and (64) [using the fact that $P_{es}^{(K)} \to 0$, as $K \to \infty$ (46)], we can choose N (and $K = NR/H_S$) sufficiently large so that

$$P_e \leqq \epsilon,$$

this is (9c). It remains to establish (9b). But from (64) with N sufficiently large, we can make

$$R\delta_K + \frac{h(\bar{\lambda})}{N} + \frac{\bar{\lambda}\log M_2}{N} + (\log A)\Pr\{\mathbf{X}^N \notin T^*(N)\} + f_1(N) \leqq \frac{\epsilon R}{2H_S}.$$

Then (63) and (56a) yield

$$\Delta \geqq \frac{H_S\Gamma(R)}{R} - \epsilon = d - \epsilon,$$

which is (9b). Thus, (R, d) is achievable and the proof of the direct half of Theorem 2, i.e., $\bar{\mathcal{R}} \subseteq \mathcal{R}$, is complete. It remains to prove Lemmas 11 and 8.

Proof of Lemma 11: We begin with some notation. For $\mathbf{x} \in \mathfrak{X}^N$, let

$$\mu(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \notin T^*(N), \\ 0, & \text{otherwise.} \end{cases} \tag{65}$$

Also for a given set $\{\mathbf{x}_m\}_1^{M_1}$, let $\lambda^{(m)}(\mathbf{x}_1, \cdots, \mathbf{x}_{M_1})$ be the error probability that results when $\{\mathbf{x}_m\}$ is used as a channel code for channel $Q_M^{(N)}$ with prior probabilities (49b) when code word $\mathbf{x}_m$ is transmitted and when maximum liklihood decoding is used. Thus,

$$\lambda = \sum_{i=1}^{M} \sum_{m=(i-1)M_2+1}^{iM_2} \frac{q_i}{M_2} \lambda^{(m)}(\mathbf{x}_1, \cdots, \mathbf{x}_{M_1}).$$

Further, with λ_i defined as above as the error probability for code C_i on $Q_{MW}^{(N)}$, write $\lambda_i = \lambda_{MW}(\mathbf{x}_{(i-1)M_2+1}, \cdots, \mathbf{x}_{iM_2}) = \lambda_{MW}(C_i)$, so that the dependence of λ_i on C_i is explicit. We have

$$\bar{\lambda} = \sum_{i=1}^{M} q_i\lambda_i = \sum q_i\lambda_{MW}(C_i).$$

Finally, define

$$\begin{aligned}\Phi(\mathbf{x}_1, \cdots, \mathbf{x}_{M_1}) &\triangleq \Pr\{\mathbf{X}^N \notin T_X^*(N)\} + \lambda + \bar{\lambda} \\ &= \sum_{i=1}^{M} \sum_{m=(i-1)M_2+1}^{iM_2} \frac{q_i}{M_2}[\mu(\mathbf{x}_m) + \lambda^{(m)}(\mathbf{x}_1, \cdots, \mathbf{x}_{M_2})] \\ &\quad + \sum_{i=1}^{M} q_i\lambda_{MW}(C_i).\end{aligned} \tag{66}$$

Now suppose that the set $\{x_m\}_1^{M_1}$ is chosen at random, with each $\mathbf{x}_m$ chosen independently from $\mathfrak{X}^N$, with probability distribution $p_{X^*}^{(N)}(\mathbf{x}) = \prod_{n=1}^N p_X^*(x_n)$. We establish the lemma by showing that $E\Phi \leqq F_3(N)$. Now observe that, from (59), $(1/N)\log M_1$ is bounded below $I(X^*, Y^*)$. Also from (61), $(1/N)\log M_2$ is bound below $I(X^*; Z^*)$. It follows from the standard random channel-coding theorem (see, for example, Ref. 1, Theorem 5.6.2) that $E\lambda^{(m)}$, $E\lambda_{MW} \leqq f_4(N) \to 0$, as $N \to \infty$. Further, $E\mu = \Pr\{\mathbf{X}^* \notin T_X^*(N)\} \leqq f_5(N) \to 0$, by (44). Thus, $E\Phi \leqq 2f_4(N) + f_5(N) \triangleq f_3(N) \to 0$. Hence the lemma.

Proof of Lemma 8: Here too we begin with some notation. Let p be a probability distribution on $\mathfrak{X}$, and let $\mathfrak{g}(p)$ be the mutual information between the input and output of channel Q_{MW} when the input has distribution p. It is known (Ref. 1, Theorem 4.4.2) that $\mathfrak{g}(p)$ is a concave function of p. Let $\mu(\mathbf{x})$ be as in (65), and write (for any encoder-decoder)

$$\begin{aligned}\frac{1}{N}I(\mathbf{X}^N; \mathbf{Z}^N) &= \frac{1}{N}I[\mathbf{X}^N, \mu(\mathbf{X}^N); \mathbf{Z}^N] \\ &= \frac{1}{N}I[\mathbf{X}^N; \mathbf{Z}^N | \mu(\mathbf{X}^N)] + \frac{1}{N}I[\mu(\mathbf{X}^N); \mathbf{Z}^N] \\ &= \frac{1}{N}\sum_{j=0}^{1}\Pr\{\mu(\mathbf{X}^N) = j\} I(\mathbf{X}^N; \mathbf{Z}^N | \mu(\mathbf{X}^N) = j) \\ &\quad + \frac{1}{N}I[\mu(\mathbf{X}^N); \mathbf{Z}^N].\end{aligned} \tag{67}$$

Now

$$\frac{1}{N}\Pr\{\mu(\mathbf{X}^N) = 1\} I[\mathbf{X}^N; \mathbf{Z}^N | \mu(\mathbf{X}^N) = 1] \leqq (\log A)\Pr\{\mathbf{X}^N \notin T^*(N)\}, \tag{68}$$

and

$$\frac{1}{N}I[\mu(\mathbf{X}^N); \mathbf{Z}^N] \leqq \frac{1}{N}H[\mu(\mathbf{X}^N)] \leqq \frac{1}{N}. \tag{69}$$

One term remains in (67). Using the memoryless property of channel $Q_{MW}^{(N)}$ (Ref. 1, Theorem 4.2.1), we have

$$\begin{aligned}\frac{1}{N}I(\mathbf{X}^N; \mathbf{Z}^N | \mu = 0) &\leqq \frac{1}{N}\sum_{n=1}^{N} I(X_n; Z_n | \mu = 0) \\ &= \frac{1}{N}\sum_{n=1}^{N}\mathfrak{g}(p_n) \leqq \mathfrak{g}\left(\frac{1}{N}\sum_{n=1}^{N}p_n\right),\end{aligned} \tag{70a}$$

where p_n is the probability distribution for X_n given $\mu = 0$, i.e., for $1 \leqq i \leqq A$,

$$p_n(i) = \sum_{\mathbf{x}\in T^*}\delta_{x_n,i}\Pr\{\mathbf{X}^N = \mathbf{x} | \mathbf{X}^N \in T^*\}. \tag{70b}$$

The last inequality in (70a) follows from the concavity of $\mathfrak{g}$. From

(70b),

$$\bar{p}(i) \triangleq \frac{1}{N} \sum_{n=1}^{N} p_n(i) = \sum_{\mathbf{x} \in T^*} \Pr\{\mathbf{X}^N = \mathbf{x} | \mathbf{X} \in T^*\} \frac{\#(i, \mathbf{x})}{N}. \quad (71)$$

The definition of T^* (43) and eq. (71) yields

$$|\bar{p}(i) - p_X^*(i)| \leq \delta_N \to 0, \quad \text{as } N \to \infty.$$

Since $\mathfrak{g}(p)$ is a continuous function of p, we have

$$|\mathfrak{g}(\bar{p}) - \mathfrak{g}(p_X^*)| \leq g(N) \to 0, \quad \text{as } N \to \infty. \quad (72)$$

Substituting (72) into (70a), we obtain

$$\frac{1}{N} \Pr\{\mu = 0\} I(\mathbf{X}^N; \mathbf{Z}^N | \mu = 0) \leq \mathfrak{g}(p_X^*) + g(N)$$
$$= I(X^*; Z^*) + g(N). \quad (73)$$

Finally, setting $f_1(N) = (1/N) + g(N)$, and substituting (68), (69), and (73) into (67) we have Lemma 8.

VI. ACKNOWLEDGMENTS

I would like to acknowledge helpful discussions with my colleagues D. Slepian, H. S. Witsenhausen, and C. Mallows that contributed to this paper. In particular, the problem was originally formulated in collaboration with Mr. Witsenhausen, and the coding scheme described above for the special case (main channel noiseless, wire-tap channel a BSC) is based on an idea of Mr. Mallows. I also wish to thank M. Hellman of Stanford University, whose recent paper[4] stimulated this research. Furthermore, the pioneering work of C. E. Shannon[5] on relating information theoretic ideas to cryptography should be noted.

APPENDIX A

The Data-Processing Theorem and Fano's Inequality

Let U, V, $\hat{U}$ be discrete random variables that form a Markov chain. Then the *data-processing theorem* can be stated as

$$H(U|V) \leq H(U|\hat{U}), \quad (74a)$$

or equivalently

$$I(U;V) \geq I(U;\hat{U}). \quad (74b)$$

Inequality (74a) follows on writing

$$H(U|V) \overset{(a)}{=} H(U|V,\hat{U}) \overset{(b)}{\leq} H(U|\hat{U}),$$

where step (a) follows from (4), and (b) from the fact that conditioning decreases entropy [Ref. 1, eq. (2.3.13)].

Next, let U, V, $\hat{U}$ be a Markov chain as above, but now assume that U, $\hat{U}$ take values in $\mathfrak{U}\,(|\mathfrak{U}| \leq \infty)$. Let

$$\lambda = \Pr\{U \neq \hat{U}\}. \quad (75)$$

Fano's inequality is

$$H(U|V) \leq h(\lambda) + \lambda \log(|\mathfrak{U}| - 1) \leq h(\lambda) + \lambda \log |\mathfrak{U}|. \quad (76)$$

To verify (76), define the random variable

$$\Phi(U, \hat{U}) = \begin{cases} 0, & U = \hat{U}, \\ 1, & U \neq \hat{U}, \end{cases}$$

and then write

$$\begin{aligned} H(U|V) &\overset{(a)}{\leq} H(U|\hat{U}) \leq H(U, \Phi|\hat{U}) \\ &= H(\Phi|\hat{U}) + H(U|\hat{U}, \Phi) \\ &\leq H(\Phi) + H(U|\hat{U}, \Phi) \\ &= H(\Phi) + \Pr\{\Phi = 0\} H(U|\hat{U}, \Phi = 0) \\ &\quad + \Pr\{\Phi = 1\} H(U|\hat{U}, \Phi = 1) \\ &\overset{(b)}{=} h(\lambda) + (1-\lambda)\cdot 0 + \lambda H(U|\hat{U}, \Phi = 1) \\ &\overset{(c)}{\leq} h(\lambda) + \lambda \log(|\mathfrak{U}| - 1) \leq h(\lambda) + \lambda \log |\mathfrak{U}|, \end{aligned}$$

which is (76). Step (a) is (74a), and step (b) follows from the fact that, given $\Phi = 0$, then $U = \hat{U}$, so that $H(U|\hat{U}, \Phi = 0) = 0$, and step (c) from the fact that, given $\Phi = 1$, U takes one of the $|\mathfrak{U}| - 1$ values in $\mathfrak{U}$ excluding $\hat{U}$.

A variation of Fano's inequality is the following. Let $\mathbf{S}^K$, V, $\hat{\mathbf{S}}^K$ be a Markov chain where the coordinates of $\mathbf{S}^K$ and $\hat{\mathbf{S}}^K$ take the values in the set $\mathcal{S}$. Let

$$P_{ek} = \Pr\{S_k \neq \hat{S}_k\} \quad (77a)$$

and

$$P_e = \frac{1}{K} \sum_{k=1}^{K} P_{ek}. \quad (77b)$$

We will show that Fano's inequality implies

$$\frac{1}{K} H(\mathbf{S}^K|V) \leq h(P_e) + P_e \log(|\mathcal{S}| - 1) \triangleq \delta(P_e). \quad (78)$$

To verify (78), write

$$\begin{aligned} \frac{1}{K} H(\mathbf{S}^K|V) &\overset{(a)}{\leq} \frac{1}{K} \sum_{k=1}^{N} H(S_k|V) \\ &\overset{(b)}{\leq} \frac{1}{K} \sum_{k=1}^{N} \delta(P_{ek}) \overset{(c)}{\leq} \delta(P_e), \end{aligned}$$

which is (78). Step (a) is a standard inequality, step (b) follows on applying (76) to the Markov chain S_k, V, $\hat{S}_k$, and step (c) from the concavity of $\delta(\cdot)$.

APPENDIX B

Proof of Lemma 1

(*i*) With no loss of generality, let $\mathfrak{X} = \{1, 2, \cdots, A\}$. Any probability distribution p_X can be thought of as an A-vector $\mathbf{p} = (p_1, p_2, \cdots, p_A)$. Since $I(X;Y)$ is a continuous function of p_X, the set $\mathcal{P}(R)$ is a compact subset of Euclidean A-space. Since $I(X;Y|Z)$ is also a continuous function of p_X, we conclude that $I(X;Y|Z)$ has a maximum on $\mathcal{P}(R)$. This is part (*i*).

(*ii*) Let $0 \leqq R_1, R_2 \leqq C_M$, and $0 \leqq \theta \leqq 1$. We must show that

$$\Gamma[\theta R_1 + (1-\theta)R_2] \geqq \theta\Gamma(R_1) + (1-\theta)\Gamma(R_2). \tag{79}$$

For $i = 1, 2$, let $\mathbf{p}_i \in \mathcal{P}(R_i)$ achieve $\Gamma(R_i)$. In other words, letting X_i, Y_i, Z_i correspond to $\mathbf{p}_i$, $i = 1, 2$, then

$$I(X_i, Y_i) \geqq R_i, \quad I(X_i, Y_i|Z_i) = \Gamma(R_i). \tag{80}$$

Now let the random variable X be defined as in Fig. 5. For $i = 1, 2$, the box labeled "$\mathbf{p}_i$" generates the random variable X_i that has probability distribution "$\mathbf{p}_i$." The switch takes upper position ("position 1") with probability θ and the lower position ("position 2") with probability $1-\theta$. Let V denote the switch position. In the figure, $V = 1$. Assume that V, X_1, X_2 are independent. As indicated in the figure, $X = X_i$, when $V = i$, $i = 1, 2$. Now

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \overset{(a)}{=} H(Y) - H(Y|X,V) \\ &\geqq H(Y|V) - H(Y|X,V) = I(X;Y|V) \\ &= \theta I(X;Y|V=1) + (1-\theta)I(X;Y|V=2) \\ &= \theta I(X_1;Y_1) + (1-\theta)I(X_2;Y_2) \\ &\overset{(b)}{\geqq} \theta R_1 + (1-\theta)R_2. \end{aligned} \tag{81}$$

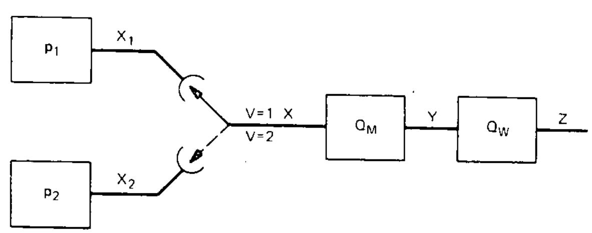

Fig. 5—Defining the random variable X.

Step (a) follows from the fact that V, X, Y is a Markov chain and (4). Step (b) follows from (80). Inequality (81) implies that the distribution defining X belongs to $\mathcal{P}[\theta R_1 + (1-\theta)R_2]$. Thus, from the definition of Γ,

$$\Gamma[\theta R_1 + (1-\theta)R_2] \geqq I(X;Y|Z). \tag{82}$$

Continuing (82) and paralleling (81), we have

$$\begin{aligned} \Gamma[\theta R_1 + (1-\theta)R_2] &\geqq H(Y|Z) - H(Y|XZ) \\ &= H(Y|Z) - H(Y|XZV) \\ &\geqq H(Y|ZV) - H(Y|XZV) \\ &= I(X;Y|ZV) = \theta I(X;Y|Z, V=1) \\ &\quad + (1-\theta)I(X;Y|Z, V=2) \\ &= \theta I(X_1;Y_1|Z_1) + (1-\theta)I(X_2;Y_2|Z_2) \\ &= \theta\Gamma(R_1) + (1-\theta)\Gamma(R_2), \end{aligned}$$

which is (79). This is part (*ii*).

(*iii*) This part follows immediately from the definition of $\Gamma(R)$ (10), since $\mathcal{P}(R)$ is a nonincreasing set.

(*iv*) Since $\Gamma(R)$ is concave on $[0, C_M]$, and nonincreasing, it must be continuous for $0 \leqq R < C_M$. Thus, we need only verify the continuity of $\Gamma(R)$ at $R = C_M$. Let $\mathbf{p}$ be a probability distribution on $\mathfrak{X}$ viewed as a vector in Euclidean A-space, as in the proof of part (*i*). Let $\mathfrak{g}(\mathbf{p})$ and $\hat{\mathfrak{g}}(\mathbf{p})$ be the values of $I(X;Y)$ and $I(X;Y|Z)$, respectively, which correspond to $\mathbf{p}$. $\mathfrak{g}(p)$ and $\hat{\mathfrak{g}}(p)$ are continuous functions of $\mathbf{p}$.

Now let $\{R_j\}_1^\infty$ be a monotone increasing sequence such that $R_j \to C_M$, and $R_j \leqq C_M$. We must show that, as $j \to \infty$,

$$\Gamma(R_j) \to \Gamma(C_M). \tag{83}$$

Now from the monotonicity of $\Gamma(R)$, $\lim_{j\to\infty} \Gamma(R_j)$ exists and

$$\lim_{j\to\infty} \Gamma(R_j) \geqq \Gamma(C_M). \tag{84}$$

It remains to verify the reverse of ineq. (84). Let $\{\mathbf{p}_j\}_1^\infty$ satisfy

$$\mathfrak{g}(\mathbf{p}_j) \geqq R_j, \quad \hat{\mathfrak{g}}(\mathbf{p}_j) = \Gamma(R_j), \tag{85}$$

for $1 \leqq j < \infty$. Since the set of probability A-vectors is compact, there exists a probability distribution $\mathbf{p}^*$ on $\mathfrak{X}$ such that for some subsequence $\{\mathbf{p}_{j_k}\}_{k=1}^\infty$

$$\lim_{k\to\infty} \mathbf{p}_{j_k} = \mathbf{p}^*.$$

It follows from the continuity of $\mathcal{g}(\cdot)$, and (85) that $\mathcal{g}(\mathbf{p}^*) \geqq C_M$, so that $\mathbf{p}^* \in \mathcal{P}(C_M)$. Therefore, from the continuity of $\hat{\mathcal{g}}(\cdot)$, and (85), we have

$$\lim_{j\to\infty} \Gamma(R_j) = \lim_{k\to\infty} \Gamma(R_{jk}) = \lim_{k\to\infty} \hat{\mathcal{g}}(\mathbf{p}_{jk}) = \hat{\mathcal{g}}(\mathbf{p}^*) \overset{(a)}{\leqq} \Gamma(C_M), \tag{86}$$

where step (a) follows from $\mathbf{p}^* \in \mathcal{P}(C_M)$. Inequalities (84) and (86) yield (83) and part (*iv*).

(*v*) From (12),

$$\begin{aligned}\Gamma(R) &= \sup_{p_X \in \mathcal{P}(R)} [I(X;Y) - I(X;Z)] \\ &\leqq \sup_{p_X \in \mathcal{P}(R)} I(X;Y) \leqq C_M,\end{aligned}$$

which is the first inequality in part (*v*). Also, using (12),

$$\begin{aligned}\Gamma(C_M) &= \sup_{p_X \in \mathcal{P}(C_M)} [I(X;Y) - I(X;Z)] \\ &\geqq \sup_{p_X \in \mathcal{P}(C_M)} [I(X;Y) - C_{MW}] = C_M - C_{MW}.\end{aligned} \tag{87}$$

Since $\Gamma(R)$ is nonincreasing, (87) yields $\Gamma(R) \geqq \Gamma(C_M) \geqq C_M - C_{MW}$, completing the proof of part (*v*).

APPENDIX C

Source with Memory

In this appendix, we show how to modify our definitions and results for a source with memory. We will take the source output sequence $\{S_k\}$ to be a stationary, ergodic sequence (where S_k takes values in $\mathcal{S}$) with entropy (as defined in Ref. 1, Section 3.5) of H_S. As in Section II, we continue to assume that $|\mathcal{S}| < \infty$, and that the source statistics are known.

The channels Q_M and Q_W remain as in Section II, as does the definition of an encoder-decoder with parameters N and K. The definition of P_e also remains unchanged, but a new definition for Δ is necessary. To see this, let us suppose that the source was binary, i.e., $\mathcal{S} = \{0, 1\}$, with entropy H_S, and with $H(S_1) > H_S$. Suppose also that the channel Q_M is a noiseless binary channel, and that Q_W has zero capacity. A possible encoder-decoder has $K = N = 1$ and takes $X_1 = S_1$. Such a scheme has $P_e = 0$, but with Δ as defined in (7) given by $\Delta = H(S_1) > H_S$. Using (9), this would lead us to accept the pair $[H_S, H(S_1)]$ as achievable, which would not be reasonable. Accordingly, we give a new definition of Δ.

Let $\mathbf{S}^K$, $\mathbf{Z}^N$ correspond to an encoder with parameters K, N as defined in Section II. Let $\mathbf{S}^K(j)$, $\mathbf{Z}^N(j)$, $j = 1, 2, \cdots, \nu$, correspond to the ν successive repetitions of the encoding process. Then define the equivocation at the wire-tap as

$$\begin{aligned}\Delta &= \lim_{\nu\to\infty} \frac{1}{K\nu} H[\mathbf{S}^K(1), \cdots, \mathbf{S}^K(\nu) | \mathbf{Z}^N(1), \cdots, \mathbf{Z}^N(\nu)] \\ &= \lim_{\nu\to\infty} \frac{1}{K\nu} H(\mathbf{S}^{K\nu} | \mathbf{Z}^{N\nu}).\end{aligned} \tag{88}$$

With Δ as defined by (88), we define the sets $\mathcal{R}$ and $\bar{\mathcal{R}}$ as in Section II. We claim that Theorem 2 remains valid.

The proof of the converse-half of Theorem 2 given in Section IV goes over to the case where the source has memory with only trivial changes. Further, the results in Section V are all valid exactly for the source with memory. They yield that, if (R, d) satisfies (56), then we can for $\epsilon > 0$ arbitrary find an encoder-decoder with parameters N, K, and P_e which satisfies

$$\frac{KH_S}{N} \geqq R - \epsilon, \tag{89a}$$

$$P_e \leqq \epsilon, \tag{89b}$$

$$\frac{1}{K} H(\mathbf{S}^K | \mathbf{Z}^N) \geqq d - \epsilon. \tag{89c}$$

Further, we can do this for arbitrarily large K. We show below that there exists a function $f(K)$, $K = 1, 2, \cdots$, such that for any code with parameters K, N

$$\Delta = \lim_{\nu\to\infty} \frac{1}{K\nu} H(\mathbf{S}^{K\nu} | \mathbf{Z}^{N\nu}) \geqq \frac{1}{K} H(\mathbf{S}^K | \mathbf{Z}^N) - f(K), \tag{90}$$

where $\lim_{K\to\infty} f(K) = 0$, and $f(K)$ depends only on the source statistics. Combining (90) with (89c), we have

$$\Delta \geqq d - \epsilon - f(K).$$

Since $f(K) \to 0$, we conclude that (R, d) is achievable. This is the direct half of Theorem 2. It remains to verify (90).

First, imagine that the encoder-decoder begins operation infinitely far in the past. Let $[\mathbf{S}(j), \mathbf{Z}(j)]$ be the $(\mathbf{S}^K, \mathbf{Z}^K)$ corresponding to the jth encoding operation, $-\infty < j < \infty$. Thus, $\mathbf{S}^{K\nu} = (S_1, \cdots, S_{K\nu}) = [\mathbf{S}(1), \cdots, \mathbf{S}(\nu)]$ and $Z^{K\nu} = [\mathbf{Z}(1), \cdots, \mathbf{Z}(\nu)]$, $\nu = 1, 2, \cdots$. Let $\mathbf{Z}^* = [\cdots, Z(-1), Z(0), Z(+1), \cdots]$. Of course,

$$H(\mathbf{S}^{K\nu} | \mathbf{Z}^{N\nu}) \geqq H(\mathbf{S}^{K\nu} | \mathbf{Z}^*). \tag{91}$$

Further,

$$
\begin{aligned}
H(\mathbf{S}^{K\nu}|\mathbf{Z}^*) &= H[\mathbf{S}(1), \cdots, \mathbf{S}(\nu)|Z^*] \\
&\overset{(a)}{=} \sum_{j=1}^{\nu} H[\mathbf{S}(j)|\mathbf{Z}^*, \mathbf{S}(j+1), \cdots, \mathbf{S}(\nu)] \\
&\overset{(b)}{=} \sum_{j=1}^{\nu} H[\mathbf{S}(1)|\mathbf{Z}^*, \mathbf{S}(2), \cdots, \mathbf{S}(j)] \\
&\overset{(c)}{\geqq} \nu H[\mathbf{S}(1)|\mathbf{Z}^*, \mathbf{S}(2), \cdots, \mathbf{S}(\nu)] \geqq \nu H[\mathbf{S}(1)|\mathbf{Z}^*, \mathbf{S}'],
\end{aligned}
\tag{92}
$$

where $\mathbf{S}' = [\mathbf{S}(2), \mathbf{S}(3), \cdots]$. Step (a) is a standard identity, step (b) follows from the stationarity of the sequence $\{S_k\}$ and the memorylessness of the channel Q_{MW}, and step (c) follows from the fact that conditioning decreases entropy. Now, let

$$
\begin{aligned}
\mathbf{S} &= \mathbf{S}^K = \mathbf{S}(1), \quad \mathbf{S}' = [\mathbf{S}(2), \mathbf{S}(3), \cdots], \\
\mathbf{Z} &= \mathbf{Z}^N = \mathbf{Z}(1), \quad \mathbf{Z}' = [\cdots, \mathbf{Z}(-1), \mathbf{Z}(0), \mathbf{Z}(+2), \cdots].
\end{aligned}
$$

Thus, (91) and (92) become

$$
\begin{aligned}
\frac{1}{K\nu} H(\mathbf{S}^{K\nu}|\mathbf{Z}^{N\nu}) &\geqq \frac{1}{K} H(\mathbf{S}|\mathbf{Z}, \mathbf{Z}', \mathbf{S}') \\
&= \frac{1}{K}[H(\mathbf{SZ}|\mathbf{Z}'\mathbf{S}) - H(\mathbf{Z}|\mathbf{Z}'\mathbf{S}')] \\
&= \frac{1}{K}[H(\mathbf{S}|\mathbf{Z}'\mathbf{S}') + H(\mathbf{Z}|\mathbf{SZ}'\mathbf{S}') - H(\mathbf{Z}|\mathbf{Z}'\mathbf{S}')] \\
&\overset{(a)}{=} \frac{1}{K}[H(\mathbf{S}|\mathbf{S}') + H(\mathbf{Z}|\mathbf{S}) - H(\mathbf{Z}|\mathbf{Z}'\mathbf{S}')] \\
&\geqq \frac{1}{K}[H(\mathbf{S}|\mathbf{S}') + H(\mathbf{Z}|\mathbf{S}) - H(\mathbf{Z})].
\end{aligned}
\tag{93}
$$

Step (a) follows from the fact that $\mathbf{Z}', \mathbf{S}', \mathbf{S}$ and $(\mathbf{S}', \mathbf{Z}'), \mathbf{S}, \mathbf{Z}$ are Markov chains, and (4). Now

$$
\begin{aligned}
\frac{1}{K} H(\mathbf{S}|\mathbf{S}') &= \frac{1}{K} \sum_{k=1}^{K} H(S_k|\mathbf{S}', S_{k+1}, \cdots, S_K) \\
&= \frac{1}{K} \sum_{k=1}^{K} H_S = H_S.
\end{aligned}
\tag{94}
$$

Also,

$$
\left|\frac{1}{K} H(\mathbf{S}) - H_S\right| \leqq f(K) \to 0, \quad \text{as } K \to \infty. \tag{95}
$$

Substituting (95) and (94) into (93), we have

$$
\begin{aligned}
\frac{1}{K\nu} H(\mathbf{S}^{K\nu}|\mathbf{Z}^{N\nu}) &\geqq \frac{1}{K}[H(\mathbf{S}) + H(\mathbf{Z}|\mathbf{S}) - H(\mathbf{Z})] - f(K) \\
&= \frac{1}{K} H(\mathbf{S}|\mathbf{Z}) - f(K),
\end{aligned}
$$

which is (90).

REFERENCES

1. R. G. Gallager, *Information Theory and Reliable Communication*, New York: John Wiley, 1968.
2. A. D. Wyner and J. Ziv, "A Theorem on the Entropy of Certain Binary Sequences and Applications: Part I," IEEE Transactions on Information Theory, *IT-19* (Nov. 1973), pp. 769–772.
3. R. B. Ash, *Information Theory*, New York: Interscience, 1965.
4. Martin E. Hellman, "The Information Theoretic Approach to Cryptography," Stanford University, Center for Systems Research, April 1974.
5. C. E. Shannon, "Communication Theory of Secrecy Systems," B.S.T.J., *28*, No. 4 (October 1949), pp. 656–715.