A Note on Wyner's Wiretap Channel

AYDANO B. CARLEIAL, MEMBER, IEEE, AND
MARTIN E. HELLMAN, MEMBER, IEEE

***Abstract*—Wyner recently introduced the concept of a wiretap channel and showed that by transmitting at a rate less than capacity on the main link it was possible to keep the wiretapper's information about the entire message equal to zero. It is shown that it is possible to send at capacity on the main link and still keep the wiretapper's information equal to zero on many, large, arbitrary portions of the message.**

I. Introduction

Wyner [1] introduced the wiretap channel shown in Fig. 1. The message $\boldsymbol{u}$ is to be reliably conveyed to the legitimate receiver as rapidly as possible, while keeping the wiretapper is ignorant of $\boldsymbol{u}$ as possible. There is hope of doing this because the wiretapper's received data $\boldsymbol{z}$ is a noisy version of the legitimate receiver's received data $\boldsymbol{y}$. It is seen that this network defines a degraded broadcast channel [3], but the goal here is quite different from the usual one. Usually, one seeks to maximize the flow of information to both receivers, whereas here one seeks to minimize the information gained by the wiretapper and to maximize the information gained by the legitimate receiver.

For simplicity, we consider only the special case of the wiretap channel shown in Fig. 2. The vector $\boldsymbol{u}$ is a totally random binary sequence of length k; the main channel is a noiseless binary channel; the wiretap channel is a binary symmetric channel (BSC) with bit error rate p. We thus have

$$\boldsymbol{y} = \boldsymbol{x}$$
$$\boldsymbol{z} = \boldsymbol{x} \oplus \boldsymbol{e} \tag{1}$$

where $\boldsymbol{x}$, an n bit vector, is the coded version of the message $\boldsymbol{u}$, and $\boldsymbol{e}$ is the error sequence on the wiretap channel. The rate of the code is k/n. It is assumed that the wiretapper has full knowledge of the code being used and that his confusion results only from his uncertainty about $\boldsymbol{e}$.

Letting $H(\boldsymbol{U}|\boldsymbol{Z})$ denote the wiretapper's equivocation, Wyner defines an (R,d) pair to be *achievable* if, for all $\epsilon > 0$, there exists an encoder-decoder with parameters n and k such that

$$k/n \geq R - \epsilon,$$
$$(1/k)H(U|Z) \geq d - \epsilon,$$

and

$$P_e \leq \epsilon,$$

where P_e denotes the decoded bit error rate at the legitimate receiver.

Since the main channel considered here is noiseless, $P_e = 0$ if and only if the encoding mapping is invertible. Wyner used a randomized, but invertible, encoding to confuse the wiretapper. He established both a forward and a converse theorem which, for the special case of Fig. 2, shows that a pair (R,d) is achievable if

Manuscript received October 7, 1975; revised July 10, 1976. The work of A. B. Carleial was supported in part by the United States Air Force Office of Scientific Research under Contract F44620-73-C-0065. The work of M. E. Hellman was supported in part by the National Science Foundation under Grants GK-33250 and ENG-10173. Paper previously presented at the Fourth International Symposium on Information Theory, Leningrad, U.S.S.R., June 15–19, 1976.

A. B. Carleial is with the Brazilian Space Research Institute (INPE), São Jose dos Campos, São Paulo, Brazil. He was with the Department of Electrical Engineering, Stanford University, Stanford, CA.

M. E. Hellman is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305.

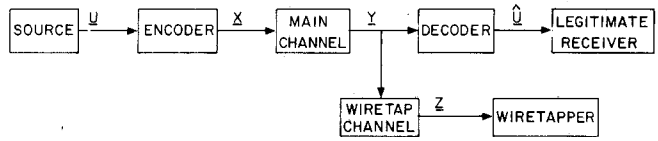

Fig. 1. Wiretap channel.

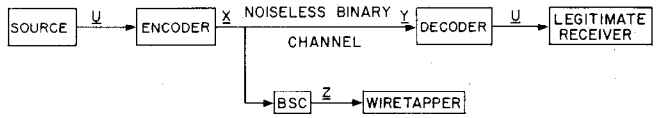

Fig. 2. Simple wiretap channel.

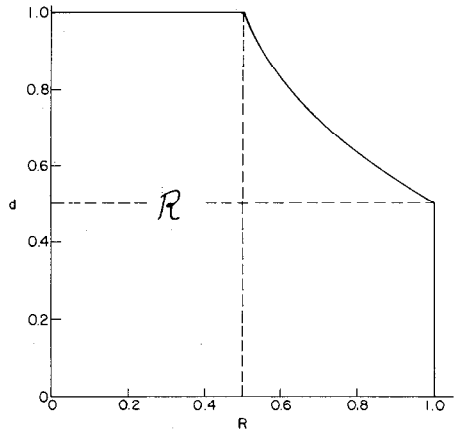

Fig. 3. Region $\mathcal{R}$ of achievable (R,d) points when $h(p) = 0.5$.

and only if

$$Rd \leq h(p) \equiv -p \log p - (1-p) \log (1-p)$$
$$0 \leq R \leq 1 \tag{2}$$
$$0 \leq d \leq 1.$$

(All logarithms are to the base 2.) The set $\mathcal{R}$ of (R,d) pairs that are achievable in the above sense is shown in Fig. 3 for $h(p) = 0.5$ which corresponds to $p = 0.110027864\cdots$. It is seen that as long as $R \leq 0.5$, the equivocation d can be kept total through use of appropriate coding. Wyner [1] found the region $\mathcal{R}$ for general wiretap channels.

When used on the binary wiretap channel with $h(p) \geq 0.5$, Wyner's rate ½ code allows $\boldsymbol{u}$ to be of length $k = n/2$ and d to equal 1. Another $n/2$ randomization bits $\boldsymbol{r}$ are used to confuse the wiretapper, but are not used to convey information. In this correspondence we show that these $n/2$ randomization bits can also be used to convey information in a secret fashion (i.e., $H(\boldsymbol{R}|\boldsymbol{Z}) = n/2$), although of course $H(\boldsymbol{U},\boldsymbol{R}|\boldsymbol{Z})$ must be strictly less than n by Wyner's converse theorem [1]. The wiretapper cannot be kept totally ignorant of both messages jointly, but can be kept totally ignorant of each individually. The penalty paid for this increased rate is that if the wiretapper learns one message through separate means, he can then also determine the other message. Wyner has independently reached much the same conclusion [4] and will discuss his result later in a somewhat different context.

Since in this correspondence we are concerned with rate 1 codes, we shall change notation slightly by amalgamating what have been called $\boldsymbol{u}$ and $\boldsymbol{r}$ into a single information vector $\boldsymbol{u}$ of length n. Our code is thus an invertible mapping from $\boldsymbol{u}$ to $\boldsymbol{x}$. We

may then divide $\boldsymbol{u}$ into two halves $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ (corresponding to $\boldsymbol{u}$ and $\boldsymbol{r}$ in our previous notation), each of which can be totally protected (i.e., $H(\boldsymbol{S}_1|\boldsymbol{Z}) = H(\boldsymbol{S}_2|\boldsymbol{Z})$ $n/2$) on an individual basis.

If $h(p) < \frac{1}{2}$, then we cannot totally protect $\boldsymbol{s}_1$ or $\boldsymbol{s}_2$. If, however, $h(p) \geq \frac{1}{3}$, we will show that $\boldsymbol{u}$ can be divided into thirds, $\boldsymbol{s}_1$, $\boldsymbol{s}_2$, and $\boldsymbol{s}_3$, each of length $n/3$, and each of which can be totally protected on an individual basis. The following theorem summarizes the goal of this note. Its proof occupies the next section.

Theorem 1: Let $\boldsymbol{s}$ be an m-dimensional projection of $\boldsymbol{u}$ given by

$$\boldsymbol{s} = \boldsymbol{u}P, \tag{3}$$

where P is an arbitrary n by m matrix of rank m, and where $\boldsymbol{u}$ is a totally random binary n-vector. With P fixed, consider the ensemble of linear, rate 1 codes with encoding and decoding functions

$$\boldsymbol{x} = \boldsymbol{u}A^{-1} \tag{4}$$

$$\boldsymbol{u} = \boldsymbol{x}A, \tag{5}$$

where A is chosen randomly and uniformly from the set of n by n invertible binary matrices. Over this ensemble of codes,

$$\Pr\{A: H(\boldsymbol{S}|\boldsymbol{Z}) \geq m(1-\Delta)\} = 1 - \delta(n), \tag{6}$$

for any $\Delta > 0$, provided that the wiretap channel is noisy enough that

$$h(p) \geq m/n, \tag{7}$$

where $\delta(n)$ here and hereafter denotes any quantity which tends to 0 as $n \to \infty$.

Remarks

1) If we have several m-dimensional projection operators P_1, $P_2, \cdots, P_M$ with

$$\boldsymbol{s}_i = \boldsymbol{u}P_i, \tag{8}$$

then by the union bound,

$$\Pr\{A: H(S_i|Z) \geq m(1-\Delta), \quad \text{for } i = 1,2,\cdots,M\} \geq 1 - M\delta_1(n) = 1 - \delta_2(n). \tag{9}$$

If $mM = n$, then we can take $\boldsymbol{s}_1$ to be the first m bits of $\boldsymbol{u}$, $\boldsymbol{s}_2$ to be the second m bits of $\boldsymbol{u}$, etc. Theorem 1 therefore includes direct partitioning as a special case.

2) Note that $H(\boldsymbol{S}|\boldsymbol{Z})$ in (6) is tacitly conditioned on A. We cannot deal with the ensemble entropy of $\boldsymbol{S}$ given $\boldsymbol{Z}$, because entropy is increased by mixing. Just showing that the ensemble entropy is large would not guarantee the existence of a code (equivalently, a matrix A) for which $H(\boldsymbol{S}|\boldsymbol{Z})$ was also large.

3) An intuitive interpretation can be gained by rewriting (7) as

$$nh(p) \leq m. \tag{10}$$

At least as many bits of noise entropy $nh(p)$ are needed as there are bits of information m to be covered by confusion.

II. Random Coding Argument

Our choice of the notation $\boldsymbol{s}$ was deliberate since we shall soon identify $\boldsymbol{s}$ with a syndrome vector. Letting

$$H = AP, \tag{11}$$

we find from (3) and (5) that

$$\boldsymbol{s} = \boldsymbol{x}AP = \boldsymbol{x}H. \tag{12}$$

The wiretapper receives $\boldsymbol{z} = \boldsymbol{x} \oplus \boldsymbol{e}$, and hence knows that

$$\boldsymbol{s} = (\boldsymbol{z} \oplus \boldsymbol{e})H = \boldsymbol{z}H \oplus \boldsymbol{e}H, \tag{13}$$

where $\boldsymbol{e}$ has the distribution appropriate to a BSC with bit error rate p. Because all 2^n values of $\boldsymbol{x}$ are equally likely, $\boldsymbol{z}$ and $\boldsymbol{e}$ are independent. The distribution on $\boldsymbol{s}$ given $\boldsymbol{z}$ is therefore just a translate of the distribution on $\boldsymbol{e}H$. We therefore have

$$H(\boldsymbol{S}|\boldsymbol{Z}) = H(\boldsymbol{E}H). \tag{14}$$

The following argument, leading to (15), is somewhat nonrigorous. Lemmas 1, 2, and 3 below establish Theorem 1 in a more precise manner. Regarding H as the parity check matrix for an $(n, n-m)$ linear code, we can consider $\boldsymbol{e}H$ as the syndrome vector. If this linear code is a "good" error correcting code, then, for $R < C = 1 - h(p)$ or equivalently for $h(p) < 1 - R = m/n$, "almost all" $\boldsymbol{e}$ will yield different syndromes, and thus

$$H(\boldsymbol{E}H) \doteq H(\boldsymbol{E}) = nh(p). \tag{15}$$

Then, by invoking the data processing theorem [2, p. 80] for degraded channels (increasing p results in a degraded channel), we can say that for $h(p) \geq m/n$, $H(\boldsymbol{S}|\boldsymbol{Z}) \doteq m$, and that

$$\begin{aligned} I(\boldsymbol{S};\boldsymbol{Z}) &= H(\boldsymbol{S}) - H(\boldsymbol{S}|\boldsymbol{Z}) \\ &= m - H(\boldsymbol{S}|\boldsymbol{Z}) \\ &\doteq 0. \end{aligned} \tag{16}$$

We now make these arguments more precise.

Lemma 1: Let

$$H = AP, \tag{17}$$

where P is a fixed n by m binary matrix of rank m. If A is chosen uniformly from the set of nonsingular, n by n binary matrices, then the distribution induced on H is the uniform distribution over all n by m binary matrices of rank m.

Lemma 2: Choosing an n by m parity check matrix H of rank m uniformly from all its possible values yields the same ensemble of codes as choosing an n-m by n generator matrix G of rank $n - m$ uniformly from all its possible values.

Lemma 3: If an $n - m$ by n generator matrix G of rank $n - m$ is chosen uniformly from all its possible values, and if H is an associated parity check matrix, then, for any $\Delta > 0$,

$$\Pr\{G: H(\boldsymbol{E}H) \geq n(h(p) - 2\Delta)\} = 1 - \delta(n), \tag{18}$$

provided m/n is kept constant and $h(p) < 1 - R = m/n$, where p is the bit error rate of $\boldsymbol{e}$.

Proof of Lemma 1: Lemma 1 is almost self-evident. However, we include a proof both for completeness and to introduce a counting argument which will be used in later proofs.

We first prove Lemma 1 for the special case

$$P = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & & 1 \\ 0 & 0 & & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & & 0 \end{bmatrix} \triangleq P^* \tag{19}$$

which corresponds to $\boldsymbol{s}$ being the first m components of $\boldsymbol{u}$. In this case, the columns of H are the first m columns of A in the same order (see (17)).

Note that the number of nonsingular n by n matrices A is

$$(2^n - 1)(2^n - 2)(2^n - 4)\cdots(2^n - 2^{n-1}) \tag{20}$$

because there are $2^n - 1$ choices for the first column of A (any nonzero n-vector will do); there are $2^n - 2$ choices for the second column of A once the first column is chosen (any n-vector not in the one-dimensional subspace spanned by the first column of A will do, and there are 2^k vectors in a k-dimensional subspace); there are $2^n - 2^2$ choices for the third column of A once the first

two columns are chosen (any n-vector not in the two-dimensional subspace spanned by the first two columns will do), etc.

Similarly, there are

$$(2^n - 1)(2^n - 2)(2^n - 4) \cdots (2^n - 2^{m-1}) \tag{21}$$

possible values for H since its columns are m linearly independent n-vectors. In the special case (19), the columns of H are the first m columns of A, thus there are an equal number

$$(2^n - 2^m)(2^n - 2^{m+1}) \cdots (2^n - 2^{n-1}) \tag{22}$$

of A mapping onto each H. This proves Lemma 1 for the special case (19).

Now let P be an arbitrary n by m, rank m matrix. There exists a (nonunique) nonsingular matrix B such that

$$P = BP^*. \tag{23}$$

(The first m columns of B must be the same as the columns of P). Then the ensemble of H is generated by

$$H = AP = (AB)P^*, \tag{24}$$

as A goes through all nonsingular matrices. But as A goes through this range so does AB. Therefore the ensemble of H is the same as when $P = P^*$. Q.E.D.

Proof of Lemma 2: For a given G matrix, the code generated is the $(n - m)$-dimensional subspace spanned by the rows of G. Since the rows of G are n-vectors, we can use a counting argument similar to that used in the proof of Lemma 1 to show that each $(n,n - m)$ code (i.e., each $(n - m)$-dimensional subspace) is generated by the same number

$$(2^{n-m} - 1)(2^{n-m} - 2)(2^{n-m} - 4) \cdots (2^{n-m} - 2^{n-m-1}) \tag{25}$$

of different G. (The first row of G can be any nonzero vector in the $(n - m)$-dimensional code space, etc.) Choosing G uniformly from all $(n - m)$ by n matrices of maximal rank is therefore equivalent to choosing a code uniformly from all $(n,n - m)$ linear codes.

Similarly, since the span of the columns of a parity check matrix determines the code, there are the same number

$$(2^m - 1)(2^m - 2)(2^m - 4) \cdots (2^m - 2^{m-1}) \tag{26}$$

of different H giving rise to each $(n,n - m)$ code. Therefore choosing H uniformly is also equivalent to choosing an $(n,n - m)$ code uniformly. Q.E.D.

Proof of Lemma 3: The proof of Lemma 3 is in two steps. First, we show that if G is chosen uniformly from the set of $n - m$ by n matrices of maximal rank, then with probability $1 - \delta(n)$, the code generated by G is "good" when used on a BSC with bit error rate p, provided $R < C = 1 - h(p)$. That is, the block error rate $P(e)$ of this code when used for error correction on such a channel is less than ϵ, where $\epsilon > 0$ is arbitrary. (Note that $P(e)$ is different from P_e. P_e is the bit error rate at the legitimate receiver and is 0 in our model since the main channel is noiseless. $P(e)$ has little physical significance since the code generated by G is not being used to correct errors on either channel; rather, it is used in a scrambling mode to confuse the wiretapper.)

If G were chosen uniformly from all $n - m$ by n matrices of arbitrary rank, we could invoke the coding theorem for parity check codes [2, p. 206] to say that $\Pr\{G: P(e) < \epsilon\} = 1 - \delta(n)$. Now, using the counting argument of Lemma 1, we can show that then $\Pr\{\rho(G) = n - m\} \geq 1 - 2^{-m} = 1 - \delta(n)$, where $\rho(G)$ is the rank of G. Therefore,

$$\Pr\{P(e) < \epsilon | \rho(G) = n - m\} \geq \Pr\{P(e) < \epsilon, \quad \rho(G) = n - m\} = 1 - \delta(n) \tag{27}$$

which is the desired result.

Now we will show that for a code with $P(e) < \epsilon$,

$$H(\boldsymbol{EH}) \geq \log(1 - \epsilon - \delta) + n(h(p) - \Delta)(1 - \epsilon - \delta), \tag{28}$$

where $\Delta > 0$ is also arbitrary, and $\delta \to 0$ as $n \to \infty$ for fixed Δ. Then, by letting $n \to \infty$ (equivalently $\delta \to 0$) and $\epsilon \to 0$, we obtain (18) as desired.

To establish (28), we first define the set of Δ-typical $\boldsymbol{e}$

$$T = \{\boldsymbol{e}: |(1/n) \log \Pr(\boldsymbol{e}) + h(p)| < \Delta\}. \tag{29}$$

By the asymptotic equipartition property, for any $\Delta > 0$,

$$\Pr(\boldsymbol{e} \in T) = 1 - \delta(n). \tag{30}$$

Now define C to be the set of coset leaders (of minimum weight) associated with the code. Since

$$P(e) = 1 - \Pr(\boldsymbol{e} \in C) < \epsilon \tag{31}$$

$$\Pr(\boldsymbol{e} \in C) \equiv \Pr(C) > 1 - \epsilon. \tag{32}$$

Therefore, the set of coset leaders which are also typical has

$$\Pr(\boldsymbol{e} \in T \cap C) \equiv \Pr(TC) \geq 1 - \epsilon - \delta(n). \tag{33}$$

We can now bound the quantity of interest to obtain (28)

$$\begin{aligned} H(\boldsymbol{EH}) &\geq H(\boldsymbol{EH} | \boldsymbol{E} \in TC) \Pr(TC) \\ &= \Pr(TC) \sum_{TC} \Pr(\boldsymbol{e}) / \Pr(TC) \log [\Pr(TC) / \Pr(\boldsymbol{e})] \\ &\geq \log \Pr(TC) + n(h(p) - \Delta) \sum_{TC} \Pr(\boldsymbol{e}) \\ &\geq \log(1 - \epsilon - \delta) + n(h(p) - \Delta)(1 - \epsilon - \delta). \end{aligned} \tag{34}$$

The proofs of these three lemmas also constitute a proof of Theorem 1.

III. Discussion

In a discussion of [1] with its author, the question arose as to whether it was necessary for the wiretapper's equivocation per bit d to be close to 1 (i.e., total ignorance), or whether a nonzero value of d was sufficient in practice, provided that n was large enough so that 2^{nd}, the number of "typical" messages under the $\Pr(\boldsymbol{u}|\boldsymbol{z})$ distribution, was astronomical. Wyner pointed out that by letting $\boldsymbol{x} = \boldsymbol{u}$ (i.e., no encoding and operation at rate 1), d would equal $h(p) > 0$. Yet, if p were 0.001, the wiretapper would know $\boldsymbol{u}$ with only a 0.001 bit error rate by letting $\hat{\boldsymbol{u}} = \boldsymbol{z}$. This clearly is a bad situation even though $d = h(0.001) \doteq 0.01141$ so that a 10^5 bit message would give rise to $2^{1141} = 3 \times 10^{343}$ typical values of $\boldsymbol{u}$. There is too simple a description for this large set of $\boldsymbol{u}$ (e.g., {probable $\boldsymbol{u}$} = {$\boldsymbol{u}$: the Hamming weight of $\boldsymbol{u} \oplus \boldsymbol{z}$ is less than 120}). This simple description even allows a point estimate of $\boldsymbol{u}$, namely $\boldsymbol{z}$, with a small expected distortion. If, however, a suitable prescrambling operation is done prior to transmission, namely

$$\boldsymbol{x} = \boldsymbol{u}A^{-1}. \tag{35}$$

we have shown above that no simple description of the set of probable $\boldsymbol{u}$ exists, and that there is no point estimate of $\boldsymbol{u}$ which yields distortion (in the Hamming metric) much less than $\frac{1}{2}$.

There is a question of how to choose A. If A is to be chosen at random, the simplest technique might be to choose A uniformly from all n by n matrices by generating each of its n^2 entries independently and at random, and then to check if A is nonsingular. This can be done in at most n^3 operations (or n^2 if we count the exclusive-or of two binary n-vectors as one operation) which is not too much worse than the n^2 operations (or n if a vector exclusive-or counts as one operation) required to compute $\boldsymbol{x}$ from

$\boldsymbol{u}$. Since each such choice of A has probability

$$\prod_{i=1}^{n}(1-2^{-i}) > \prod_{i=1}^{\infty}(1-2^{-i}) = 0.2878809\cdots \tag{36}$$

of success, fewer than 3.5 attempts will be necessary on the average.

If A is to be chosen to ensure that a particular projection P (e.g., the first m components of $\boldsymbol{u}$) is to be protected, then A should be chosen so that $H = AP$ is the parity check matrix of a code that is known to be good. If it is desired to guarantee protection of a set of projections $P_1, P_2, \cdots, P_M$, then A should be chosen so that the matrices $H_i = AP_i$ are all parity check matrices of good codes. This can be done provided that $mM \leq n$ and that the columns of $H_1, H_2 \cdots, H_M$ are all independent.

However, the need to choose a code carefully is not as pressing here as it is in the case of an error correction. There, even though a randomly chosen code has excellent performance for large n, the need to compute the coset leader easily from the syndrome dictates the need for a structured code. Here there is no such need and the complexity of encoding and decoding are relatively independent of the choice of A.

The only real exception to the claim of irrelevancy of the choice of code would be if A allowed the use of a rate 1 convolutional, or similar, code which might be implemented more easily in hardware. However, the causal nature of a convolutional code prevents the early part of a message from being affected by later errors. Therefore $H(\boldsymbol{S}|\boldsymbol{Z})$ cannot take on its maximal value if $\boldsymbol{s}$ consists of the first m components of $\boldsymbol{u}$. It appears, however, that a suitable variation of convolutional coding exists [5]. First encode $\boldsymbol{u}$ with a randomly chosen rate 1 convolutional code of constraint length n; then invert the order of the resultant bits; and finally reencode the inverted bits with another, independently chosen rate 1 convolutional code of constraint length n. This removes the causal nature of the code, allowing essentially perfect double-sided error propagation [5]. It is the error-propagating, or mixing, nature of the mapping $\boldsymbol{u} = \boldsymbol{x}A$ which is responsible for the success of that technique when A is chosen uniformly from all invertible matrices; this allows errors which are sparsely distributed to concentrate their uncertainty in a small portion of the decoded message.

The technique discussed in this note provides at least a partial quantification of the concepts of diffusion and mixing transformations introduced by Shannon into cryptography [6, p. 708]. Shannon suggests using diffusion to dissipate the redundancy of a message into long-range statistics, i.e., into statistical structures involving long combinations of letters. Thus while 1000 characters of English text possess about 1500 bits of *a priori* uncertainty and any consecutive 100 characters possess *a priori* uncertainty only slightly greater than 150 bits, a good diffusion coding would cause the entropy of any consecutive 100 coded characters to be approximately 500 bits, assuming each letter is mapped into 5 bits. Any successive 300 coded characters would have entropy close to 1500 bits, which is the maximum possible. Generalizing to include arbitrary projections, as in this paper, is even better.

One last comment concerns our assumption that the n-vector $\boldsymbol{u}$ is totally random and therefore has no redundancy. If this assumption is violated, our encoding operation may actually allow the wiretapper to lower $H(\boldsymbol{U}|\boldsymbol{Z})$ below what it would be if $\boldsymbol{u}$ were transmitted directly. Indeed, if $\boldsymbol{u}$ is the output of a memoryless source and $H(U) < 1 - h(p)$, then the joint source and channel coding theorem [2, p. 534] implies that as $n \to \infty$ the wiretapper can reduce $H(\boldsymbol{U}|\boldsymbol{Z})$ essentially to 0 by using the code as a joint source and channel code. If $H(\boldsymbol{U}) > n[1 - h(p)]$, we conjecture that as $n \to \infty$ any m-dimensional projection $\boldsymbol{s}$ of u can be totally protected provided that

$$m < nh(p) - [n - H(\boldsymbol{U})]. \tag{37}$$

The term in brackets in (37) is the redundancy present in the message $\boldsymbol{u}$ and acts much as an information flow to the wiretapper. In particular, it reduces the maximal size projection that can be totally protected.

ACKNOWLEDGMENT

The authors wish to thank Dr. Aaron Wyner of Bell Laboratories for several valuable suggestions and discussions.

REFERENCES

[1] A. Wyner, "The wiretap channel," *Bell Syst Tech. J.*, vol. 54, pp. 1355–1387, Oct. 1975.
[2] R. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.
[3] P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 197–207, Mar. 1973.
[4] A. Wyner, paper in preparation.
[5] M. Hellman, "On using natural redundancy for error detection," *IEEE Trans. Comm.*, vol. COM-22, pp. 1690–1693, Oct. 1974.
[6] C. Shannon, "Communication theory of secrecy systems," *Bell Syst Tech. J.*, vol. 28, pp. 656–715, 1949.

Information Structure: Common and Private

G. HEXNER AND Y. C. HO, FELLOW, IEEE

***Abstract*—Two new definitions for the common and private information structures between two decisionmakers are introduced and are shown to satisfy various reasonable properties. The new definitions are framed in a decisionmaking context and do not require a probability assignment on the space of events.**

I. INTRODUCTION

We start by defining the space of all relevant events (the states of nature) as a measurable space $(\Omega, \mathcal{F})$ where $\mathcal{F}$ is a separable σ-field. The information structure of a decisionmaker A is then defined as a subfield $A \subset \mathcal{F}$.[1] Equivalently, we can consider decisionmaker A's information as a mapping h from $(\Omega, \mathcal{F})$ to $(Z_A, \mathcal{Z}_A)$ with $z_A = h(\omega)$. A is then the field induced on $\mathcal{F}$ by h. Roughly speaking, the information structure A specifies the extent to which the decisionmaker can determine the occurrence or nonoccurrence of the various states of nature through his observation z_A.

In a decision problem involving more than one decisionmaker, each with different information structure, the problem of cost-benefit analysis for the establishment of any communication among the decisionmakers naturally arises. Often, one would like to minimize the amount of information that has to be transmitted between decisionmakers. Consequently, one would like to be able to characterize "what we both know" (the common information structure) and "what I know that you don't know" (the private information structure). This is to be contrasted with Wyner's definition [1] of "common and private information" which are

Manuscript received October 2, 1975; revised August 8, 1976. This work was supported in part by the U.S. Office of Naval Research by the Joint Electronics Program under Contract N00014-75-C-0648 and in part by the National Science Foundation under Grant GK-31511.

The authors are with Harvard University, Cambridge, MA 02138.

[1] This is not the most general possible definition, but it is sufficient for our purposes.