THE CAPACITIES OF CERTAIN CHANNEL CLASSES UNDER RANDOM CODING[1]

By David Blackwell, Leo Breiman, and A. J. Thomasian

University of California, Berkeley

1. Introduction and Summary. For any two finite sets U, V, a Markov matrix s with row set U and column set V will be called a U, V *channel*. Thus a U, V channel is any nonnegative function s, defined for all pairs (u, v), $u \,\varepsilon\, U$, $v \,\varepsilon\, V$, for which

$$\sum_v s(u, v) = 1 \qquad \text{for all } u.$$

The sets U, V will be called the *input* and *output* sets, respectively, of the channel. We shall denote by $M(U, V)$ the set of all U, V channels. A channel s may be thought of as a random device which, on being given an input element $u \,\varepsilon\, U$, produces an output element $v \,\varepsilon\, V$, with the probability of a particular output v given by $s(u, v)$.

A U, V channel s may be used as a means of communication from one person, the sender, to another person, the receiver. There is given in advance a finite set D of messages, exactly one of which will be presented to the sender for transmission. The sender encodes the message by an *encoding channel* $s_1 \,\varepsilon\, M(D, U)$, with $s_1(d, u)$ being the probability that input u is given to channel s when message d is presented to the sender for transmission. When the receiver observes the output v of the transmission channel s, he decodes it by a *decoding channel* $s_2 \,\varepsilon\, M(V, D)$, with $s_2(v, d)$ being the probability that, on receiving the transmission channel output v, the receiver will decide that message d is intended. The pair (s_1, s_2) will be called a (D, U, V) code. For a U, V channel s and a (D, U, V) code $c = (s_1, s_2)$, the matrix $\epsilon(s, c) = s_1 s s_2$, which is an element of $M(D, D)$ will be called the *error matrix* of code c in channel s. Its (d, d') element is the probability that, when message d is presented to the sender, the receiver will decide that message d' is intended, when code c is used on channel s. We shall be especially interested in the *average error probability* over all messages in the set D. This is the number

$$\pi(s, c) = 1 - |D|^{-1} \operatorname{trace} \epsilon(s, c),$$

where $|D|$ denotes the number of elements in the set D.

A code $c = (s_1, s_2)$ will be called *pure* if only 0's and 1's occur as elements of s_1, s_2. The (finite) set of all pure (D, U, V) codes will be denoted by $C(D, U, V)$, and a probability distribution k over $C(D, U, V)$ will be called a *random*

Received October 21, 1959.

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-53). This paper in whole or in part may be reproduced for any purpose of the United States Government.

(D, U, V) *code*. We define the error matrix $\epsilon(s, k)$ and average error probability $\pi(s, k)$ for a random code k by

$$\epsilon(s, k) = \sum_{c \varepsilon C(D,U,V)} k(c)\, \epsilon(s, c), \qquad \pi(s, k) = 1 - |D|^{-1} \text{trace } \epsilon(s, k).$$

It was observed by Shannon [4] that every (D, U, V) code c is equivalent to some random (D, U, V) code k, in the sense that

$$\epsilon(s, k) = \epsilon(s, c) \qquad \text{for all } s \varepsilon M(U, V).$$

The converse is not true. The greater generality of random codes lies in the possibility, with random codes, of correlated randomization in the encoding and decoding processes. This is a special case of the fact in game theory, noted by Kuhn [3], that every behavior strategy (code) is equivalent to some mixed strategy (random code), but the converse holds only in games of perfect recall (which the communication game is not).

Shannon's basic work in information theory [5], and most later work, has been concerned with the question: for a given U, V channel s and message set D, is there a pure code c which makes the average error probability $\pi(s, c)$ (or the maximum error probability) small? For this question, the distinction between pure codes and random codes is irrelevant (though even here random codes are useful as tools [5]), since

$$\pi(s, k) = \sum k(c)\pi(s, c),$$

so that there is a pure code whose average error probability is at least as small as that of any random code. We shall be concerned with some cases in which D, U, V are given, but the transmission channel is known only to be some U, V channel in a given closed set $S \subset M(U, V)$. We ask: is there a random code k for which $\pi(s, k)$ is small for every $s \varepsilon S$? For this question, as we shall see, the distinction between random codes and pure codes is essential, for some sets S.

Specifically, we shall be interested in D, U, V, S defined as follows. We are given a message set D (only $|D|$, the number of elements in D, will be relevant), an input set A, an output set B, a closed set S_0 of A, B channels, and a positive integer N. The sender will be given some message d from D, and will then choose a sequence $u = (a_1, \cdots, a_N)$ of N elements of A. These inputs will be placed successively into channels $s_1, \cdots, s_N$, $s_n \varepsilon S_0$, and the receiver will observe the resulting output sequence $v = (b_1, \cdots, b_N)$. The receiver must then estimate which message d was presented to the sender. Thus U is the set of all sequences $u = (a_1, \cdots, a_N)$ of length N of elements of A and V is the set of all sequences $v = (b_1, \cdots, b_N)$ of length N of elements of B. The set S of possible U, V channels will depend on what restrictions we place on the sequences $s_1, \cdots, s_N$. We consider three cases.

CASE 1. *Fixed unknown channel.* Here we are given that the same element of S_0 is the transmission channel for each period. There is then one U, V channel s for each A, B channel $s_0 \varepsilon S_0$. The s corresponding to s_0 is defined by

$$s(u, v) = \prod_{n=1}^{N} s_0(a_n, b_n).$$

We shall denote the set of all such U, V channels by S_1.

CASE 2. *Arbitrarily varying channel.* Here there is one U, V channel for each sequence $(s_1, \cdots, s_N)$ of elements of S_0, defined by

$$s(u, v) = \prod_{n=1}^{N} s_n(a_n, b_n).$$

We shall denote the set of all such U, V channels by S_2.

CASE 3. *Channel selected by jammer with knowledge of past inputs and outputs.* Here we suppose that the element s_n of S_0 which will be the transmission channel during the nth period if selected by a jammer after he has observed the inputs $a_1, \cdots, a_{n-1}$ and outputs $b_1, \cdots, b_{n-1}$ during the first $n - 1$ periods. A pure strategy f for the jammer is a sequence $(f_1, \cdots, f_N)$ of functions, where f_n maps every sequence $x_n = (a_1, \cdots, a_{n-1}, b_1, \cdots, b_{n-1})$ into a corresponding element $f_n(x_n)$ of S_0. There is then one U, V channel s_f for each pure strategy f, defined by

$$s_f(u, v) = \prod_{n=1}^{N} s_n(a_n, b_n), \qquad \text{where } s_n = f_n(x_n).$$

We shall denote the set of all such U, V channels by S_3.

Let us define, for $i = 1, 2, 3$,

$$\pi_i(|D|, N, S_0) = \min_k \max_{s \varepsilon S_i} \pi(s, k).$$

The number $\pi_i(|D|, N, S_0)$ is the minimum average error probability which can be guaranteed, by using a suitable random code, when there are $|D|$ possible messages, N transmission periods, the channel at each period is some element of S_0, and the channel variation from period to period is as described in Case i above. It is also the value of the following two-person zero sum game: Player I (the jammer) chooses any U, V channel s in S_i, and Player II independently chooses a pure (D, U, V) code c. A message is then selected at random from D, so that each d has probability $|D|^{-1}$ of being selected, and transmitted over channel s using code c. If an error is made, Player I wins one unit; otherwise he wins zero.

Since π is linear in s, we have

$$\pi_i(|D|, N, S_0) = \min_k \max_{s \varepsilon S_i^*} \pi(s, k),$$

where S_i^* is the convex hull of S_i, i.e. the smallest convex set containing S_i. Let us for the moment denote by T the convex hull of S_0, by T_i the set of U, V channels defined by T in the same way that S_i is defined by S_0, and by T_i^* the convex hull of T_i. It is not hard to verify that

$$S_2^* \supset T_2, \; S_3^* \supset T_3, \text{ so that } S_2^* = T_2^*, \; S_3^* = T_3^*.$$

We conclude that

$$(1) \qquad \pi_i(|D|, N, S_0^*) = \pi_i(|D|, N, S_0) \qquad \text{for } i = 2, 3,$$

a fact which will be used later.

We shall call a number $R \geqq 0$ an *attainable rate of type* i for S_0 if

$$\pi_i([2^{RN}], N, S_0) \to 0 \text{ as } N \to \infty.$$

The upper bound of the set of attainable rates of type i for S_0 will be called the *type* i *capacity* of the set S_0 and denoted by $R_i(S_0)$. Thus if R is an attainable rate of type i we can, by random encoding in large blocks, transmit R binary symbols per transmission period, with small error probability.

If, in the definition of π_i above, we had minimized over pure codes instead of random codes, we would have obtained numbers $r_i(S_0)$, which we shall call the *type* i *pure capacity* of S_0. The present authors in an earlier paper obtained a simple formula for $r_1(S_0)$. The principal result of the present paper is that

$$R_3(S_0) = R_2(S_0) = R_1(S_0^*) = r_1(S_0^*),$$

where S_0^* is the convex hull of S_0. In addition we show that always $R_1(S_0) = r_1(S_0)$ and give an example in which $R_3(S_0) > 0$, $r_2(S_0) = r_3(S_0) = 0$. The evaluation of $r_2(S_0)$ and $r_3(S_0)$ for general S_0 remains unsettled.

We may already conclude from (1) that

$$(2) \qquad R_i(S_0) = R_i(S_0^*) \qquad \text{for } i = 2, 3.$$

2. Direct half of principal result. For any random variable X with a finite set of values x, we denote by $I(X)$ the random variable whose value when $X = x$ is $-\log_2 P\{X = x\}$. For any two random variables X, Y, each with a finite set of values, we define

$$\begin{aligned} I(X \mid Y) &= I(X, Y) - I(Y) \\ J(X, Y) &= I(X) + I(Y) - I(X, Y) \\ &= I(X) - I(X \mid Y) \\ &= I(Y) - I(Y \mid X). \end{aligned}$$

$I(X)$ is usually called the information, entropy, or uncertainty in X, $I(X \mid Y)$ the information in X given Y, and $J(X, Y)$ the mutual information in X, Y. These concepts, introduced by Shannon [5], are basic in information theory.

Associated with each probability distribution s on A and A, B channel s is a probability distribution $P_{\alpha s}$ on the set $A \times B$ of pairs (a, b), defined by

$$P_{\alpha s}(a, b) = \alpha(a)s(a, b).$$

Let X, Y be the input, output variables on $A \times B$: $X(a, b) = a$, $Y(a, b) = b$ and define, for any closed subset $S \subset M(A, B)$,

$$H_\alpha(S) = \min_{s \varepsilon S} E_{\alpha s} J_{\alpha s}(X, Y),$$

$$H(S) = \max_{\alpha} H_\alpha(S),$$

where the subscripts αs indicate that expectation and mutual information are with respect to $P_{\alpha s}$.

THEOREM 1. *$R_3(S_0) \geqq H(S_0^*)$, where S_0^* is the convex hull of S_0.*

PROOF. We shall first suppose S_0 finite. It suffices to show that, for any α and any number σ with $0 < \sigma < H_\alpha(S_0^*)$, the number $H_1 = H_\alpha(S_0^*) - \sigma$ is an attainable type 3 rate for S_0. Let δ be any number for which $0 < | B |\delta \leqq 1$, where $| B |$ is the number of elements in B, and let s_δ be the B, B channel whose nondiagonal elements are all equal to δ, so that its diagonal elements are all equal to $1 - (| B | - 1)\delta$. Finally, let q be any probability distribution on the finite set F of jamming strategies f.

Let us choose a sequence $X_N^* = (X_1, \cdots, X_N)$ of N independent input variables, each with distribution α, and let L be a jamming strategy, selected independently of X_N^* with distribution q. The input sequence X_N^* and jamming strategy L determine a sequence of output variables $Y_N^* = (Y_1, \cdots, Y_N)$. We use Y_n as an input variable on the B, B channel s_δ, and let Z_n be the resulting output variable. Write $Z_n^* = (Z_1, \cdots, Z_n)$, $n = 1, \cdots, N$. Then

$$\begin{aligned} &P\{Y_N^* = v \mid X_N^* = u\} = s(u, v), \qquad s = \sum q(f)s_f, \\ (3) \qquad &P\{Z_n = b \mid X_N^*, Y_N^*, L\} = s_\delta(Y_n, b), \\ &P\{(X_n, Z_n) = (a, b) \mid X_{n-1}^*, Y_{n-1}^*, Z_{n-1}^*, L\} = P_{\alpha s^* s_\delta}(a, b), \end{aligned}$$

where s^* is the element of S_0 selected by L for the nth transmission period when the previous input-output history is X_{n-1}^*, Y_{n-1}^*. From (3) we obtain

$$\begin{aligned} &P\{(X_n, Z_n) = (a, b) | X_{n-1}^*, Z_{n-1}^*\} = \sum_{y,f} P\{Y_{n-1}^* = y, L = f \mid X_{n-1}^*, Z_{n-1}^*\} \\ (4) \qquad &\cdot P\{(X_n, Z_n) = (a, b) | X_{n-1}^*, Z_{n,1}^*, Y_{n,1}^* = y, L = f\} = P_{\alpha t s_\delta}(a, b), \end{aligned}$$

where $t = t(X_{n-1}^*, Z_{n-1}^*) \, \varepsilon \, S_0^*$, the convex hull of S_0.

We shall find an upper bound for $P\{J(X_N^*, Z_N^*) \leqq N(H_1 + \gamma)\}$, where γ is a positive number less than σ. We write

$$J(X_N^*, Z_N^*) = \sum_{n=1}^{N} [J(X_n^*, Z_n^*) - J(X_{n-1}^*, Z_{n-1}^*)] = \sum_{n=1}^{N} J_n,$$

where

$$J_n = I(X_n) + I(Z_n \mid Z_{n-1}^*) - I((X_n, Z_n) \mid (X_{n-1}^*, Z_{n-1}^*)).$$

Let us fix x^*, z^* and denote by μ the conditional joint distribution of (X_n, Z_n) given $X_{n-1}^* = x^*$, $Z_{n-1}^* = z^*$ and by β the conditional distribution of Z_n given $Z_{n-1}^* = z^*$. The conditional distribution of J_n, given $X_{n-1}^* = x^*$, $Z_{n-1}^* = z^*$ is then that of $T = I(X) - \log_2\beta(Z) - I(X, Z)$, where X, Z are the input-output variables on $A \times B$ and μ is the distribution on $A \times B$. Now

$$T = J(X, Z) + \log_2\beta'(Z) - \log_2\beta(Z),$$

where β' is the distribution of Z. Since

$$E(\log_2\beta'(Z) - \log_2\beta(Z)) = -\sum\beta'(z) \log_2(\beta(z)/\beta'(z)) \geqq 0$$

(using convexity of $-\log_2$), we obtain $ET \geqq EJ(X, Z)$. From (4), μ is a distribution $P_{\alpha t s_\delta}$ for some $t \varepsilon S_0^*$, so that, denoting by S_δ^* the set of all A, B channels of the form ts_δ, $t \varepsilon S_0^*$, we have

$$ET \geqq H_\alpha(S_\delta^*) = h(\delta). \tag{5}$$

We next find an upper bound for $| T |$. We have $T = -\log_2\beta(Z) - I(Z \mid X)$. Now $\beta(b) \geqq \delta$ and $ts_\delta(a, b) \geqq \delta$ for all a, b. Thus, since $0 \leqq -\log_2\beta(Z) \leqq -\log_2\delta$ and $0 \leqq I(Z \mid X) \leqq -\log_2\delta$, we have

$$| T | \leqq -\log_2\delta. \tag{6}$$

Using (5) and (6), we find a bound for the moment generating function of the variable $T_1 = T - h(\delta) + \lambda$, where λ is a positive number. From (5), (6) we obtain $E(T_1) \geqq \lambda$, $| T_1 | \leqq \lambda - \log_2\delta = Q = Q(\lambda, \delta)$. For $t \leqq 0$, we have $e^{tT_1} \leqq 1 + tT_1 + [(tQ)^2/2]e^{|t|Q}$, so that $\phi(t) = Ee^{tT_1} \leqq 1 + \lambda t + [(tQ)^2/2]e^{|t|Q}$. From now on, we restrict λ, δ to the set

$$\lambda/Q \leqq \log(4/3). \tag{9}$$

With this restriction, and $t_0 = -\lambda/Q^2$, we obtain

$$\phi(t_0) \leqq 1 - (\lambda^2/3Q^2) = \rho_1 = \rho_1(\lambda, \delta). \tag{10}$$

Now ϕ is the conditional moment generating function of $J_n - h(\delta) + \lambda$, given $X_{n-1}^* = x^*$, $Z_{n-1}^* = z^*$. It follows that $E(\exp t_0(\sum_{n=1}^N (J_n - h(\delta) + \lambda))) \leqq \rho_1^N$, so that

$$P\{J(X_N^*, Z_N^*) \leqq N(h(\delta) - \lambda)\} \leqq \rho_1^N(\delta, \gamma). \tag{11}$$

Now $h(\delta) \rightarrow H_\alpha(S_0^*)$ as $\delta \rightarrow 0$. Choose δ_0 sufficiently small so that

$$h(\delta_0) > H_\alpha(S_0^*) - \sigma + \gamma = H_1 + \lambda$$

and $h(\delta_0) - H_1 - \gamma \leqq -\log_2 \delta_0 \log(4/3)$, and set $\lambda_0 = h(\delta_0) - H_1 - \gamma$. From (11) we obtain

$$P\{J(X_N^*, Z_N^*) \leqq N(H_1 + \gamma)\} \leqq \rho^N = \rho^N(\sigma - \gamma) \tag{12}$$

where $\rho = \rho_1(\lambda_0, \delta_0) < 1$ and depends only on $\sigma - \gamma$ and the modulus of continuity of the function h. Inequality (12) is the first, and most difficult, step in our proof.

Now

$$P\{Z_N^* = v \mid X_N^* = u\} = \sum_{v'} P\{Y_N^* = v' \mid X_N^* = u\}P\{Z_N^* = v \mid Y_N^* = v'\} = ss_3(u, v),$$

where s is the U, V channel defined in (3) and s_3 is the V, V channel which sends inputs Y_N^* into outputs Z_N^*, with $\delta = \delta_0$. We now apply a fundamental inequality of Shannon [6], which asserts the existence, for any message set D

with $|D| \leqq 2^{NH_1}$, of a pure U, V code $c = (s_1, s_2)$, whose average error probability, on channel ss_3, is at most $P\{J(X_N^*, Z_N^*) \leqq N(H_1 + \gamma)\} + 2^{-N\gamma}$. Thus, using (12), we obtain $\pi(ss_3, c) \leqq 2\rho_2^N$, where $\rho_2 = \min_{0<\gamma<\delta} \max(2^{-\gamma}, \rho(\sigma - \gamma)) < 1$. Now $\pi(ss_3, c) = \pi(s, c^*)$, where $c^* = (s_1, s_3s_2)$.

We have now proved the

Lemma. *There is a constant $\rho_2 < 1$ such that, for $|D| = [2^{NH_1}]$ and any probability distribution q on the set F of U, V jamming strategies, there is a D, U, V code c^* for which*

$$\sum_f q(f)\pi(s_f, c^*) \leqq 2\rho_2^N .$$

We now consider the two-person zero sum game in which the pure strategies for Player I are the U, V jamming strategies f, the pure strategies for Player II are the pure D, U, V codes c, and the payoff to Player I for f, c is $\tau(s_f, c)$, the average error probability for code c on the channel sf determined by the jamming strategy f. The lemma asserts that, for any given mixed strategy q of Player I, there is a corresponding strategy for Player II which makes the payoff to I at most $2\rho_2^N$. The minimax theorem then asserts the existence of a mixed strategy for Player II, i.e., a probability distribution k over the set C of pure D, U, V codes, for which $\Sigma k(c)\pi(s_f, c) = \pi(s_f, k) \leqq 2\rho_2^N$ for all jamming strategies f, i.e., $\pi(s, k) \leqq 2\rho_2^N$ for all $s \varepsilon S_3$. Thus

$$\pi_3([2^{H_1N}], N, S_0) \leqq 2\rho_2^N \rightarrow 0 \text{ as } N \rightarrow \infty,$$

H_1 is an admissible rate of type 3, and the proof of Theorem 1 is complete for the case of finite S_0.

The restriction to finite S_0 was made only to avoid irrelevant details, e.g., measurability of jamming strategies. This restriction can now easily be removed by approximation. For an arbitrary S_0, let T be any set which contains S_0 and which is the convex hull of a finite set. Clearly $R_3(S_0) \geqq R_3(T)$, and we have shown that $R_3(T) \geqq H(T)$. Thus $R_3(S_0) \geqq \sup_T H(T)$. It is not difficult to show that $\sup_T H(T) = H(S_0^*)$, completing the proof.

3. Converse half of principal result.

Theorem 2. *For any closed S_0, $R_1(S_0) \leqq H(S_0)$.*

Proof. It was proved in [1] that $r_1(S_0) \leqq H(S_0)$. The present proof is a minor modification of the earlier one. Again, we shall use

Fano's inequality [2], [7]. *For any two random variables W, W',*

$$EI(W \mid W') \leqq -[g \log_2 g + (1 - g) \log_2(1 - g)] + g \log_2(G - 1),$$

where $g = \Pr\{W \neq W'\}$ and G is the number of values of W.

We consider a random (D, U, V) code k, take any U, V channel $s \varepsilon S_1$, and suppose that a message is selected from D with a uniform distribution and transmitted over s using k. We denote by W, X, Y, W' the resulting message, U, V input, U, V output, and estimated message respectively. Let $g = \pi(s, k) =$

$\Pr\{W' \neq W\}$. Let us denote by Z the pure code selected, so that Z is independent of W and has distribution k. Then

$$
\begin{aligned}
EJ(X, Y \mid Z) &\geqq EJ(W, W' \mid Z) \\
&= EI(W) - EI(W \mid W', Z) \\
&\geqq EI(W) - EI(W \mid W') \\
&\geqq (1 - g)\log_2 | D | - 1,
\end{aligned} \tag{13}
$$

where the last inequality is obtained from Fano's inequality. Also

$$
\begin{aligned}
EJ(X, Y \mid Z) = EI(Y \mid Z) - EI(Y \mid X, Z) &= EI(Y \mid Z) - EI(Y \mid X) \\
&\leqq EI(Y) - EI(Y \mid X) = EJ(X, Y).
\end{aligned}
$$

Combining this inequality with (13) yields

$$EJ(X, Y) \geqq (1 - g)\log_2 | D | - 1, \tag{14}$$

i.e.,

$$g = \pi(s, k) \geqq 1 - [EJ(X, Y) + 1/\log_2 | D |]. \tag{15}$$

Since the distribution of X is independent of s, we maximize (15) over $s \varepsilon S_1$, then minimize over k, to obtain

$$\pi_1(| D |, N, S_0) \geqq 1 - [H(S_1) + 1]/[\log_2 | D |]. \tag{16}$$

But, as shown in [1], $H(S_1) = NH(S_0)$, so that

$$\pi_1([2^{RN}], N, S_0) \geqq 1 - [NH(S_0) + 1]/[\log_2(2^{RN})]. \tag{17}$$

Thus if R is an admissible rate of type 1, $\lim_{N\to\infty} [NH(S_0) + 1]/[\log_2(2^{RN})] \geqq 1$, i.e., $R \leqq H(S_0)$. This completes the proof.

We summarize our results in

THEOREM 3. *For any* S_0,

$$R_3(S_0) = R_2(S_0) = R_1(S_0^*) = r_1(S_0^*),$$

where S_0^* *is the convex hull of* S_0. *Also* $R_1(S_0) = r_1(S_0) = H(S_0)$.

PROOF. That $r_1(S_0) = H(S_0)$ was shown in [1]. Since $r_1(S_0) \leqq R_1(S_0)$ and, from Theorem 2, $R_1(S_0) \leqq H(S_0)$, we have $R_1(S_0) = r_1(S_0) = H(S_0)$. The chain of inequalities

$$H(S_0^*) \leqq R_3(S_0) \leqq R_2(S_0) = R_2(S_0^*) \leqq R_1(S_0^*) \leqq H(S_0^*)$$

completes the proof of Theorem 3.

An example and an open question. We have associated with a set S_0 of A, B channels six capacities, according as (a) we face (1) the same unknown channel in S_0 each period, (2) an unknown channel varying arbitrarily in S_0 from period to period, or (3) an unknown channel in S_0, selected each period by a jammer with knowledge of previous inputs and outputs, and (b) we are restricted to pure codes or are allowed to use random codes. Of these six numbers, we have evaluated four: $r_1(S_0)$ and $R_i(S_0)$, $i = 1, 2, 3$.

The evaluation of $r_2(S_0)$, $r_3(S_0)$ remains unsolved. We conclude with an example in which $r_2(S_0) = r_3(S_0) = 0$, while $R_2(S_0) = R_3(S_0) = \frac{1}{2}$. This example illustrates that, against an unknown arbitrarily varying channel, or against a jammer, random codes are a real improvement over pure codes.

In our example, S_0 consists of two noiseless channels, labeled 0 and 1. Each channel has two inputs, 0 and 1, and three outputs, 0, 1, and 2. Channel i transmits input i perfectly, but changes the other input $1 - i$ into 2:

Input	*Channel* 0 *output*	*Channel* 1 *output*
0	0	2
1	2	1

We shall prove that, for any number N and any pure D, U, V code $c = (s_1, s_2)$, there is a channel $s \varepsilon S_2$ for which

$$\pi(s, c) \geqq (G - 1)/2G, \tag{18}$$

where $G = |D| =$ number of messages in D.

Thus no set with two or more messages can be transmitted by a pure code with average error probability less than $\frac{1}{4}$ over every sequence of channels in S_0, no matter how many transmission periods are allowed. It follows that $r_2(S_0) = 0$, and *a fortiori* $r_3(S_0) = 0$. On the other hand, our formula

$$R_3(S_0) = \max_{\alpha} \min_{s \varepsilon S_0^*} E_{\alpha s} J_{\alpha s}(X, Y)$$

yields $R_3(S_0) = \frac{1}{2}$, with $\alpha = (\frac{1}{2}, \frac{1}{2})$ as the maximizing input distribution and the channel s with matrix

$$\begin{vmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{vmatrix}:$$

the midpoint of the channels in S_0, as the minimizing channel in S_0^*.

To verify (18), let N be any positive integer, let D be any message set with $|D| = G$ elements, and let $c = (s_1, s_2)$ be any pure D, U, V code. Let x_{dn} denote the nth input specified by c for transmitting message d, and let x_d denote the vector whose coordinates are x_{dn}, $n = 1, 2, \cdots, N$: x_d is the vector in U for which $s_1(d, x_d) = 1$. Let us denote by $s(d)$ that U, V channel in S_2 which transmits x_d perfectly: $s(d)$ has channel number x_{dn} as its nth coordinate. We note that the output v corresponding to any input u and any U, V channel $s \varepsilon S_2$ has for its nth coordinate the common vaue of the nth coordinate of u and the number of the nth channel of s, if these numbers agree, and has 2 if they do not. Thus, denoting this output vector by $v(u, s)$, we have

$$v(x_d, s(d')) = v(x_{d'}, s(d)).$$

The probability $p(d, d')$ of an error in transmitting message d over channel

$s(d')$ is 0 if $v(x_d, s(d'))$ is decoded as d, and 1 otherwise. If $d' \neq d$, the vector $v(x_d, s(d')) = v(x_{d'}, s(d))$ cannot be decoded as both d and d', so that $p(d, d') + p(d', d) \geqq 1$ for $d' \neq d$. Summing this inequality over all pairs d, d' with $d' \neq d$ yields

$$2G\sum_d \pi(s(d), c) \geqq G(G - 1),$$

so that, for some d, $\pi(s(d), c) \geqq (G - 1),/2G$, and (18) is verified.

REFERENCES

[1] David Blackwell, Leo Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 1229–41.

[2] R. M. Fano, "Statistical theory of communication," notes on a course given at the Massachusetts Institute of Technology, 1952, 1953.

[3] H. W. Kuhn, "Extensive Games and the Problem of Information, Contributions to the Theory of Games," Vol. II, *Ann. Math. Studies*, No. 28, pp. 103–216, Princeton, 1953.

[4] Claude E. Shannon, "A note on a partial ordering for communications channels," *Information and Control* Vol. 1 (1958), pp. 390–398.

[5] C. E. Shannon, "A mathematical theory of communication," Bell System Tech. J., Vol. 27 (1948), pp. 379–423, 623–656.

[6] Claude E. Shannon, "Certain results in coding theory for noisy channels," *Information and Control*, Vol. 1 (1957), pp. 6–25.

[7] A. Feinstein, *Foundation of Information Theory*, McGraw-Hill, New York, 1958, pp. 35–36.