

## XII. PROCESSING AND TRANSMISSION OF INFORMATION\*

Prof. E. Arthurs	Prof. W. A. Youngblood	M. Horstein
Prof. P. Elias	D. C. Coll	T. S. Huang
Prof. R. M. Fano	J. E. Cunningham	F. Jelinek
Prof. J. Granlund	J. B. Dennis	T. Kailath
Prof. D. A. Huffman	H. A. Ernst	R. A. Kirsch
Prof. R. C. Jeffrey	E. F. Ferretti	G. E. Mongold, Jr.
Prof. H. Rodgers, Jr.	R. G. Gallager	B. Reiffen
Prof. C. E. Shannon	T. L. Grettenberg	L. G. Roberts
Prof. W. M. Siebert	F. C. Hennie III	W. L. Wells
Prof. J. M. Wozencraft	E. M. Hofstetter	H. L. Yudkin

### A. REPRODUCTION OF PICTURES BY COMPUTER DISPLAY

Digital, sampled picture data have been used in digital computers for studying pictures. The computer operations consist of various recoding processes that are used to determine upper bounds on the amount of information necessary for specifying pictures. The recoding and subsequent decoding must be done in such a manner that the picture quality is not degraded excessively. To determine whether or not a particular process is successful in this respect, a picture must be resynthesized from the decoded data.

Hitherto, this resynthesis was done by reading paper tape (upon which a decoded version of the recorded data had been punched) into a modified facsimile reproducer. This process took 4.5 hours; during this time the modulated light source was subject to drift, and photographic distortion followed in a subsequent copying process.

A program has been written for the TX-0 computer which will read in paper tapes of this kind and display the sample values on an addressable-coordinate cathode-ray oscilloscope. Each point of the oscilloscope display corresponds to a sample point of the picture, and each point is intensified a number of times equal to the value of the reflectance at that sample point. This display is photographed with a Polaroid camera and yields photographs such as those shown in Fig. XII-1. The pictures labeled (a) and (b) are reproductions of the original data tape; these pictures are quantized to 64 levels; a 6-digit binary number is required for the specification of each sample point. The pictures labeled (c) and (d) are reproduced to represent a 16-level quantization; a 4-digit binary number is required for each sample point. The successful reproduction of the pictures is due, in part, to the excellent cooperation of the TX-0 computer staff in overcoming display problems.

The time required for reproducing one of these pictures is 10 minutes. Provision is made in the program for masking out any or all of the 6 binary numbers used to

---

\*This research was supported in part by Purchase Order DDL-B222 with Lincoln Laboratory, which is supported by the Department of the Army, the Department of the Navy, and the Department of the Air Force under Contract AF19(604)-5200 with M.I.T.

(XII. PROCESSING AND TRANSMISSION OF INFORMATION)



(a)



(b)



(c)



(d)

Fig. XII-1. Pictures reproduced by computer display.

specify the 64 intensity levels of the data, and to add noise or bias as desired.

By means of a program of this sort we can get a picture with resynthesized data by performing any encoding and decoding operation before displaying. The program thus provides a high-speed reproduction of the picture which we can study to determine the effects of the encoding processes.

J. E. Cunningham

#### B. THE CAPACITY OF A MULTIPLICATIVE CHANNEL

Consider the memoryless channel whose input is the set of real numbers  $-\infty < x < \infty$ , whose output is the set of real numbers  $-\infty < y < \infty$ , and whose transition probability

(XII. PROCESSING AND TRANSMISSION OF INFORMATION)

$p(y|x)$  is a function of the ratio  $y/x$ . This defines a multiplicative channel

$$y = nx \tag{1}$$

where the probability distribution of the multiplicative noise  $n$  is given. We consider here only symmetric probability densities of  $n$ , that is,

$$p_n(n) = p_n(-n)$$

We ask, "Does this channel have a capacity when the second moment of  $x$  is constrained?"

We show in this report that the capacity of this channel is indeed infinite, and that this capacity may be obtained with a suitably shifted version of any nontrivial probability density function  $p_x(x)$ .

Note that because  $p_n(n)$  is symmetric,  $p_y(y)$  will also be symmetric, irrespective of  $p_x(x)$ . Thus without loss of generality, we can confine our attention to the three variables  $x'$ ,  $y'$ ,  $n'$  which may not assume negative values:

$$\Pr(x') = \begin{cases} p_x(x) + p_x(-x) & x > 0 \\ p_x(0) & x = 0 \\ 0 & x < 0 \end{cases} \tag{2}$$

$$\Pr(y') = \begin{cases} p_y(y) + p_y(-y) & y > 0 \\ p_y(0) & y = 0 \\ 0 & y < 0 \end{cases} \tag{3}$$

$$\Pr(n') = \begin{cases} p_n(n) + p_n(-n) & n > 0 \\ p_n(0) & n = 0 \\ 0 & n < 0 \end{cases} \tag{4}$$

Thus

$$y' = n'x' \tag{5}$$

Define

$$w = \log y', u = \log n', v = \log x' \tag{6}$$

Therefore

$$w = u + v \tag{7}$$

## (XII. PROCESSING AND TRANSMISSION OF INFORMATION)

Thus, by a change of variables, we have constructed a channel with additive noise:  $u$ .

Since  $\Pr(n')$  is known, the density function  $p_u(u)$  can be obtained by standard methods. The second-moment constraint on  $x$  and  $x'$ ,

$$\int_{-\infty}^{\infty} x^2 p_x(x) dx = \int_0^{\infty} (x')^2 \Pr(x') dx = \sigma^2 \quad (8)$$

in terms of the transformed variable  $v = \log x'$ , becomes

$$\int_{-\infty}^{\infty} e^{2v} p_v(v) dv = \sigma^2 \quad (9)$$

We desire a choice of  $p_x(x)$  or  $\Pr(x')$  that is consistent with the constraint of Eq. 8, which maximizes the mutual information between input and output,  $I(X;Y)$  or equivalently,  $I(X';Y')$ . However,  $v$  and  $w$  are related to  $x'$  and  $y'$  in one-to-one correspondence. Thus

$$I(X';Y') = I(V;W) = H(W) - H(W|V) \quad (10)$$

where the  $H$ 's are the respective entropies.

Since the noise  $u$  is additive in its channel,  $H(W|V)$  is the noise entropy and is a constant, being a function only of the density function  $p_u(u)$ .

Now, since  $w = u + v$ , and  $u$  and  $v$  are independent random variables,  $H(W) \geq H(V)$ . Therefore,

$$I(V;W) \geq H(V) - H(W|V) \quad (11)$$

Consider any probability density  $p_v(v)$  that makes the right-hand side of Eq. 11 equal to an arbitrarily large rate. In general, this will not satisfy the constraint of Eq. 9.

Define the new density

$$p'_v(v) = p_v(v+\lambda) \quad (12)$$

Applying the constraint of Eq. 9 to  $p'_v$ , we have

$$\int_{-\infty}^{\infty} e^{2v} p'_v(v) dv \quad (13)$$

Making the substitution  $q = v + \lambda$ , we obtain

$$\int_{-\infty}^{\infty} e^{2(q-\lambda)} p_v(q) dq = e^{-2\lambda} \int_{-\infty}^{\infty} e^{2v} p_v(v) dv \quad (14)$$

Thus by suitable choice of parameter  $\lambda$ , we can be assured that the density  $p'_v$  satisfies the constraint.

## (XII. PROCESSING AND TRANSMISSION OF INFORMATION)

However, the entropy of density  $p'_v$  is clearly identical to the entropy of density  $p_v$ . Thus we have succeeded in transmitting an arbitrarily high rate with arbitrarily low second moment on the input variable  $x$ .

Since  $x' = e^v$ , we can find the density of  $x'$  associated with densities  $p_v$  and  $p'_v$ :

$$P'r(x') = \frac{1}{x'} p'_v(\log x'), \quad Pr(x') = \frac{1}{x'} p_v(\log x') \quad (15)$$

Since  $p'_v(v) = p_v(v+\lambda)$ , we have

$$P'r(x) = \frac{1}{x'} p_v(\log [e^\lambda x']) \quad (16)$$

Thus the  $P'r(a)$  and  $Pr(b)$  densities are related by the transformation  $a = e^{-\lambda} b$ . If we note this relationship between  $a$  and  $b$ , it is clear that the effect of a linear shift  $\lambda > 0$  in the  $v$ -domain is equivalent in the  $x'$ -domain to a compression toward zero and renormalization.

I wish to acknowledge helpful discussions with Professor C. E. Shannon on this problem.

B. Reiffen

### C. ASYMPTOTIC BEHAVIOR OF OPTIMUM FIXED-LENGTH AND SEQUENTIAL DICHOTOMIES

One of the classical problems of statistics is the determination of the probability distribution of a random variable from measurements made on the variable itself. The simplest, nontrivial version of this problem is the dichotomy – a situation in which the unknown distribution is known to be one of two possible distributions. In this report two of the most important solutions of the dichotomy are described and their performances compared.

#### 1. The Optimum Fixed-Length Dichotomy

Let  $x$  be a random variable admitting one of two possible (discrete) probability distributions,  $p_0(x)$  or  $p_1(x)$ , and let  $\xi$  denote the a priori probability that distribution "0" pertains. It is our object to decide, on the basis of  $N$  independent measurements on  $x$ , which distribution is present. To do this with a minimum probability of error (1), we calculate the a posteriori probability,  $\xi_N$ , that "0" is true,

$$\xi_N = \frac{\xi p_0(x_1) \dots p_0(x_N)}{\xi p_0(x_1) \dots p_0(x_N) + (1-\xi) p_1(x_1) \dots p_1(x_N)} \quad (1)$$

and decide in favor of "0" if  $\xi_N \geq 1/2$ , and in favor of "1" if  $\xi_N < 1/2$ . Since Eq. 1

## (XII. PROCESSING AND TRANSMISSION OF INFORMATION)

may be written in the equivalent form,

$$\xi_N = \frac{1}{1 + \frac{1-\xi}{\xi} \ell_N}$$

where

$$\ell_N = \frac{p_1(x_1)}{p_0(x_1)} \cdots \frac{p_1(x_N)}{p_0(x_N)}$$

the decision procedure amounts to selecting "0" if  $\ell_N \leq (1-\xi)/\xi$ , and "1" otherwise.

This decision rule may be put into a still more useful form by defining

$$L_N = \log \ell_N = \sum_{i=1}^N \log \left[ \frac{p_1(x_i)}{p_0(x_i)} \right]$$

In terms of this quantity, our decision rule reads: select "0" if  $L_N \leq \log \left( \frac{1-\xi}{\xi} \right)$ , and "1" otherwise. The probability of error,  $p_\epsilon$ , incurred by this process is

$$p_\epsilon = \xi P_0 \left\{ L_N > \log \left( \frac{1-\xi}{\xi} \right) \right\} + (1-\xi) P_1 \left\{ L_N \leq \log \left( \frac{1-\xi}{\xi} \right) \right\} \quad (2)$$

where the subscripts denote which probability distribution is to be used in the calculation of the pertinent probability.

Since the  $x_i$  are independent, we see that  $L_N$  is the sum of  $N$  independently and identically distributed random variables. The problem of computing the asymptotic behavior of  $p_\epsilon$  for large  $N$  thus reduces to the classical problem of finding the asymptotic behavior of the probability of the "tails" of the distribution of the sum of a large number of independent random variables. This problem has been solved by Shannon (2), whose results we shall now state in a form suited to our needs.

### 2. Asymptotic Behavior of the Tails of the Distribution of the Sum of $N$ Independently and Identically Distributed Random Variables

Let  $\{y_i\}$  denote a sequence of independent random variables with the common cumulative distribution function,  $F(x) = P(y_i \leq x)$ .

We now define

$$\mu(s) = \log \phi(s)$$

where

$$\phi(s) = \int_{-\infty}^{\infty} e^{xs} dF(x)$$

and

(XII. PROCESSING AND TRANSMISSION OF INFORMATION)

$$Y_N = \sum_{i=1}^N y_i$$

In terms of these definitions, Shannon's results may be summarized by the following statement: For any A,

$$\left. \begin{array}{l} P(Y_N \leq A) \\ P(Y_N \geq A) \end{array} \right\} \sim \frac{K}{N^{1/2}} e^{\mu(s)N} \quad (3)$$

as  $N \rightarrow \infty$ . Here  $s$  is the root of the equation  $\mu'(s) = 0$ , and  $K$  is a constant that depends on  $A$  but is independent of  $N$ . The upper formula holds when the expectation of  $y$  is greater than zero ( $\mu'(0) > 0$ ), and the lower formula holds when the expectation of  $y$  is less than zero ( $\mu'(0) < 0$ ).

### 3. Asymptotic Behavior of $p_\epsilon$ for the Optimum Fixed-Length Dichotomy

We now apply relation 3 to our fixed-length decision problem, in order to determine the asymptotic behavior of  $p_\epsilon$ . To this end, we need the functions:

$$\phi_i(s) = \int_{-\infty}^{\infty} \left[ \frac{p_1(x)}{p_0(x)} \right]^s p_i(x) dx$$

$$\mu_i(s) = \log \phi_i(s) \quad i = 0, 1$$

We note that  $\phi_0(s) = \phi_1(s-1)$  and, therefore,  $\mu_0(s) = \mu_1(s-1)$ . It follows that, if  $s_0$  and  $s_1$  are the roots of the equations  $\mu_0'(s_0) = 0$  and  $\mu_1'(s_1) = 0$ , respectively,  $s_1 = s_0 - 1$ . This implies that  $\mu_0(s_0) = \mu_1(s_1)$  which, in turn, implies that both terms in Eq. 2 have the same exponential behavior. In other words, the large  $N$  behavior of  $p_\epsilon$  is given by the formula

$$p_\epsilon \sim \frac{K}{N^{1/2}} e^{\mu_0(s_0)N} \quad (4)$$

Where  $s_0$  is the root of the equation,  $\mu_0'(s_0) = 0$ , and  $K$  is a constant that depends on  $\xi$  but is independent of  $N$ . We shall now derive the asymptotic behavior of the optimum sequential dichotomy and compare it with expression 4.

### 4. The Optimum Sequential Dichotomy

It can be shown (1) that the sequential decision scheme that minimizes the probability of error for a given expected length of test is of the following form:

At the  $N^{\text{th}}$  step of the test compute  $L_N$  and  
if  $L_N \leq \log B$ , accept "0"

(XII. PROCESSING AND TRANSMISSION OF INFORMATION)

if  $L_N \geq \log A$ , accept "1"

if  $\log B < L_N < \log A$ , take another measurement and go through the same procedure, using  $L_{N+1}$ . The constants  $A$  and  $B$  are to be chosen so as to minimize  $p_\epsilon$  for a given expected length,  $\bar{N}$ .

Wald (3) has shown that the asymptotic behavior of  $p_\epsilon$  and  $\bar{N}$  as  $A \rightarrow \infty$  and  $B \rightarrow 0$  is given by

$$\left. \begin{aligned} p_\epsilon &\sim \frac{\xi}{A} + (1-\xi) B \\ \bar{N} &\sim \frac{\xi}{I_0} \log B + \frac{1-\xi}{I_1} \log A \end{aligned} \right\} \quad (5)$$

where

$$I_i = \mu'_i(0) \quad (6)$$

It is now easy to minimize  $p_\epsilon$  for a fixed  $\bar{N}$ ; the result is

$$p_\epsilon \sim K' e^{-I \bar{N}} \quad (7)$$

where

$$\frac{1}{I} = \frac{-\xi}{I_0} + \frac{1-\xi}{I_1}$$

and  $K'$  is a constant dependent upon  $\xi$  but independent of  $N$ . Equation 7 is the desired asymptotic relationship for the sequential test.

5. Comparison of the Asymptotic Behavior of the Two Dichotomies

We shall now compare the asymptotic behavior of the fixed-length dichotomy given by expression 4 with that of the sequential dichotomy given by expression 7. Our main result in this direction is that the exponent in the sequential case is larger than the exponent in the fixed-length case; that is,

$$I > -\mu_0'(s_0) \quad (8)$$

To derive this, we note that Eq. 6 implies that  $\mu'_0(0) = I_0$  and  $\mu'_0(1) = I_1$ . Recalling

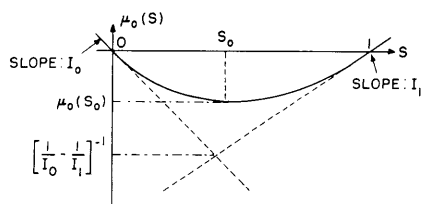


Fig. XII-2. Lower bound for  $\mu_0'(s_0)$ .



(XII. PROCESSING AND TRANSMISSION OF INFORMATION)

that  $s_0$  is the root of the equation  $\mu'_0(s_0) = 0$ , we can make a sketch of  $\mu_0(s)$ , as shown in Fig. XII-2. From the sketch we see that the convexity ( $\mu''_0(s) > 0$ ) of the function  $\mu_0(s)$  enables us to use the intersection of the tangents to  $\mu_0(s)$  at  $s = 0$  and  $s = 1$  as a lower bound for  $\mu_0(s_0)$ . Symbolically,

$$\mu_0(s_0) \geq - \left[ \frac{1}{I_1} - \frac{1}{I_0} \right]^{-1}$$

However,

$$\begin{aligned} \frac{-\mu_0(s_0)}{I} &= -\mu_0(s_0) \left[ \frac{1}{I_1} - \frac{1}{I_0} \right] \left[ I \left( \frac{1}{I_1} - \frac{1}{I_0} \right) \right]^{-1} \\ &\leq 1 \cdot \left[ I \left( \frac{1}{I_1} - \frac{1}{I_0} \right) \right]^{-1} = \frac{1-\xi}{I_1} - \frac{\xi}{I_0} \bigg/ \frac{1}{I_1} - \frac{1}{I_0} \end{aligned} \quad (9)$$

Expression 9 can be written in two equivalent forms as follows,

$$\begin{aligned} \frac{-\mu_0(s_0)}{I} &\leq \xi \frac{\frac{1-\xi}{\xi} - \frac{I_1}{I_0}}{1 - \frac{I_1}{I_0}} = 1 - \xi \frac{1 - \frac{\xi}{1-\xi} \frac{I_1}{I_0}}{1 - \frac{I_1}{I_0}} \\ &\leq \begin{cases} \xi & \text{if } \frac{1-\xi}{\xi} \leq 1 \\ 1-\xi & \text{if } \frac{1-\xi}{\xi} \geq 1 \end{cases} < 1 \end{aligned}$$

Inequality 8 follows at once. It shows that for long tests (large  $N$  or  $\bar{N}$ ) the sequential test has a probability of error that is almost exponentially smaller than the fixed-length test. In other words, not only is the probability of error for the sequential test less than the corresponding probability of error for the fixed-length test, but the ratio of the two probabilities goes to zero almost (except for a factor of  $N^{1/2}$ ) exponentially.

E. M. Hofstetter

References

1. D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions* (John Wiley and Sons, Inc., New York, 1954).
2. C. E. Shannon, *Information Theory Seminar Notes*, M.I.T., 1956 (unpublished).
3. A. Wald, *Statistical Decision Functions* (John Wiley and Sons, Inc., New York, 1950).