IX. PROCESSING AND TRANSMISSION OF INFORMATION*

Prof. E. Arthurs
Prof. J. B. Dennis
Prof. M. Eden
Prof. P. Elias
Prof. R. M. Fano
Prof. R. G. Gallager
Prof. E. M. Hofstetter
Prof. D. A. Huffman
Prof. W. F. Schreiber
Prof. C. E. Shannon
Prof. J. M. Wozencraft
S. Asano
G. Cheadle
J. E. Cunningham
P. M. Ebert
H. A. Ernst
E. F. Ferretti
T. J. Goblick, Jr.
U. F. Gronemann
F. C. Hennie III
T. S. Huang
F. Jelinek
T. Kailath
L. Kleinrock
H. H. Loomis, Jr.
J. L. Massey
J. W. Pan
D. L. Reich
L. G. Roberts
O. J. Tretiak
J. Ziv

A. PICTURE-PROCESSING RESEARCH

To facilitate the study of images and image-coding methods, we are assembling a research device that will enable us to make computer tapes from original photographs and to reverse this process. The equipment will be of such quality and flexibility that

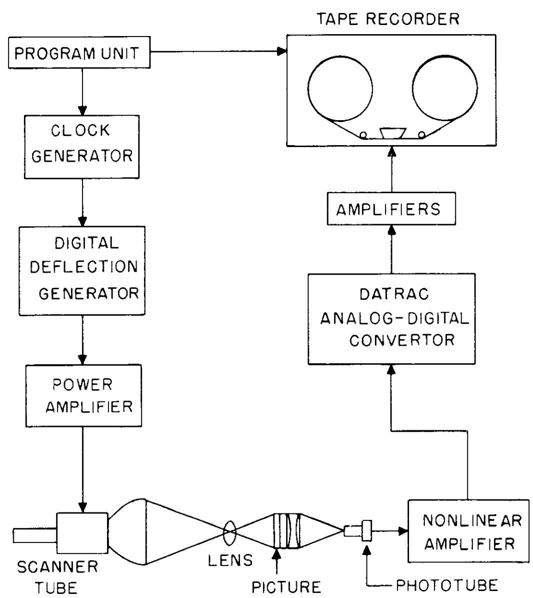


Fig. IX-1. Making tapes from pictures.

it can be operated easily and reproducibly so that the appearance of the pictures will be affected only by the television system parameters and not by the idiosyncrasies of the

*This research was supported in part by Purchase Order DDL B-00306 with Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the joint support of the U.S. Army, Navy, and Air Force under Air Force Contract AF19(604)-5200.

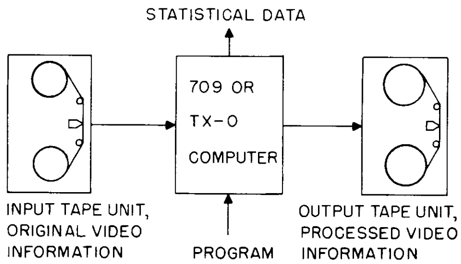


Fig. IX-2. Use of the computer for measurements and video processing.

processing circuitry and tubes. With this device we shall be able to make a tape from any picture, including color originals. We shall also be able to so program the computer that the action of any conceivable circuit or system can be simulated. The operation of this equipment is illustrated in Figs. IX-1, IX-2, and IX-3.

As a part of the precision picture recording and reproducing system, a digitally controlled deflection circuit has been designed. It is completely transistorized and uses magnetic deflection. It can be programmed to scan parts of a picture in any desired sequence. A tunnel diode oscillator generates clock pulses; the clock rate, the rate of moving from one picture element to the next, will be 6 kc (as limited by the computer tape requirements), but could be as high as 30 kc.

Two nonlinear amplifiers are being constructed: one for the recording mode, and one for the playback mode. They are required to have variable transfer curves. We employ piecewise linear approximation techniques. Thus, by using quite simple diode circuits, we achieve a variety of different transfer curves that should be adequate for our purpose.

The link between the scanning device and the computer that is used to simulate coding systems, consists of various gating, recording, and playback functions.

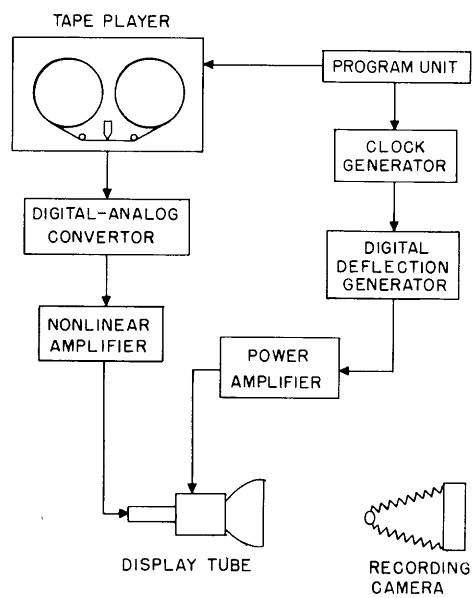


Fig. IX-3. Displaying and recording the processed picture.

An analog signal is converted, first, to digital form by the DATRAC converter, then the 10-bit output of the converter is recorded on magnetic tape in IBM 709 computer format by gating the 10 bits to the recording amplifier in 5-bit groups.

Playing back a digital record is accomplished by reversing the operation and combining the two 5-bit groups into a 10-bit group which is then decoded by a transistorized digital-to-analog converter. The DAC converter furnishes the analog voltage for playback into the nonlinear amplifier.

A tunnel diode oscillator furnishes the 6-kc clock source for the entire system.

Most of the circuitry has been constructed, and only a small part of it is still in the breadboard stage.

J. W. Pan, U. F. Gronemann, T. S. Huang,
J. E. Cunningham, W. F. Schreiber

B. ENCODING FOR TIME-DISCRETE MEMORYLESS CHANNELS

We report here upon some aspects of the problem of communication by means of memoryless channels. A block diagram of a general communication system for such a channel is shown in Fig. IX-4.

Fig. IX-4. Communication system for memoryless channel.

In 1959, Shannon (1) studied coding and decoding systems for a time-discrete but amplitude-continuous channel with additive Gaussian noise, subject to the constraint that all code words were required to have exactly the same power. Upper and lower bounds were found for the probability of error when optimal codes and optimal decoding systems are used. The lower bound followed from sphere-packing arguments, and the upper bound was derived by using random-coding arguments.

In random coding for such a Gaussian channel, one considers the ensemble of codes obtained by placing M points randomly on the surface of a sphere of radius $(nP)^{1/2}$, where nP is the power of each code word, and n = 2WT, with W the bandwidth of the channel, and T the time length of each code word. More precisely, each point is placed independently of all other points with a probability measure proportional to surface area or, equivalently, to solid angle. Shannon's upper and lower bounds for the probability of error are very close together for signaling rates from some R_{crit} to channel capacity, C.

Fano (2) has recently studied the general discrete memoryless channel. For this case the signals are not constrained to have exactly the same power. If random coding is used, the upper and lower bounds for the probability of error are, again, close together for all rates R above some R_{crit}.

The detection scheme that was used in both of these studies is an optimal one, that is, one that minimizes the probability of error for the given code. Such a scheme requires that the decoder compute an a posteriori probability measure, or a quantity equivalent to it, for each of, say, the M allowable code words. Since R is defined for a given code (block) length n by the equation

$$R = \frac{1}{n} \ln M$$

we have $M = \exp(nR)$.

In all cases in which optimum decoding is used, the lower bound on the probability of error, P_e, has the form $P_e \geq K \exp(-E^*(R) \cdot n)$, where K is a constant that is independent of n. The behavior of $E^*(R)$ as a function of R is described in Fig. IX-5. Similarly, when optimum random coding is used, the probability of error is upper-bounded by $P_e \leq 2 \exp(-E(R) \cdot n)$. (In general, construction of a random code involves the selection of messages with some probability density P(x) from the set of all possible messages. When P(x) is such that E(R) and $E^*(R)$ coincide for $R > R_{crit}$, the random code is called "optimum.") It has been shown that $E(R) \approx E^*(R)$ for all $R \geq R_{crit}$, as illustrated in Fig. IX-5. We see that specification of an extremely small probability of error implies (in general) a significantly large value for the code length n, and hence an extremely large value for the number of code words M.

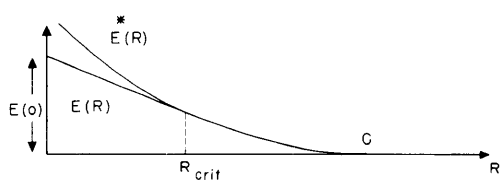


Fig. IX-5. The behavior of $E^*(R)$ and E(R) as a function of R.

Kelly (3) has derived a class of codes for continuous channels. These are block codes in which the (exponentially large) set of code words can be computed from a much smaller set of generators by a procedure analogous to group coding for discrete channels. Unfortunately, there seems to be no simple detection procedure for these codes. The receiver must generate each of the possible transmitted combinations and must then compare them with the received signal.

The sequential encoding-decoding scheme of Wozencraft (4), extended by Reiffen (5), is a code that is well suited to the purpose of reducing the average number of computations. They have shown (4, 5) that, for the general discrete and symmetric channel, the average number of computations can be bounded by $k \cdot n^2$ for all rates below some R_{comp}.

In this report, we consider the effect of constructing a discrete signal space in such a way as to make the application of sequential encoding-decoding possible for the general (continuous) memoryless channel. The effect of this constraint on the probability of error will be considered. In particular, Shannon's work (1) considered code selection from an infinite ensemble, but in this investigation the ensemble is a finite ensemble. The engineering advantage is that each code word can be sequentially generated from a small set of basic waveforms. Ultimately, we propose to consider the problem of sequentially decoding transmissions of this form.

1. Signal-Space Structure

We proceed to introduce a structured signal space, and to investigate the effect of a particular structure on the exponent of the probability of error. Two cases will be considered.

Case I: The power of each sample is less than or equal to P.

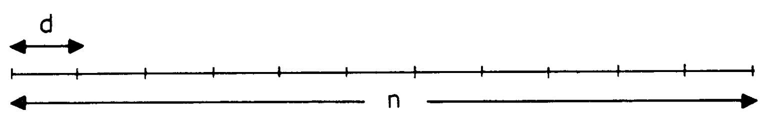


Fig. IX-6. Construction of a code word as a series of elements.

Let each code word of length n be constructed as a series of m elements, each of which has the same length d, as shown in Fig. IX-6. Thus we have

$$m = \frac{n}{d}; \quad d = \frac{n}{m} \tag{1}$$

Each of the m elements is to be picked at random, with probability $1/\ell$, from a set X_ℓ of ℓ waveforms (vectors)

$$X_\ell = \{x_k; k=1\ldots\ell\} \tag{2}$$

The length, or "dimensionality," of each x_k is d.

Let the set X_ℓ be generated in the following manner: Each vector x_k is picked at random, with probability density $P(x_k)$, from the ensemble of all d-dimensional vectors meeting the power constraint P. The probability density $P(x_k)$ is the same as that used for the generation of the optimum unrestricted random code that yields the optimum exponent E(R).

We have established the following theorem.

THEOREM: Let a general memoryless channel be represented by the set of transition probability densities $\{P(y|x)\}$. Let $E_{\ell,d}(R)$ be the exponent of the probability of error for the random code constructed above. Let $\overline{E_{\ell,d}(R)}$ be the expected value of $E_{\ell,d}(R)$ averaged over all possible sets X_ℓ. Then

$$\overline{E_{\ell,d}(R)} \geq E(R) - \frac{1}{d} \ln \left(\frac{e^{dE(0)} + \ell - 1}{\ell} \right) \tag{3}$$

for $R \leq R_{crit}$, and

$$\overline{E_{\ell,d}(R)} \geq E(R) - \frac{1}{d} \ln \left(\frac{e^{2dC} + \ell - 1}{\ell} \right) \tag{4}$$

for any $R \leq C$, where C, the channel capacity, is the rate for which $E(R) = 0$. Thus, if we define C_s, the source capacity, as $C_s = \frac{1}{d} \ln \ell$, we get

$$\overline{E_{\ell,d}(R)} > E(R) - \frac{1}{d} \ln \left(e^{d(E(0)-C_s)} + 1 \right) \tag{5}$$

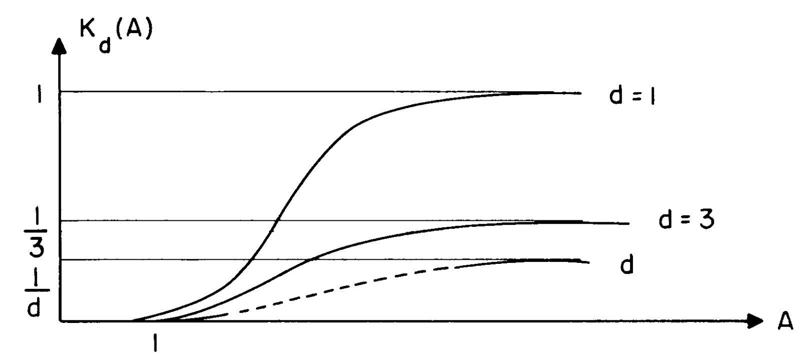
for $R \leq R_{crit}$, and

$$E_{\ell,d}(R) > E(R) - \frac{1}{d} \ln \left(e^{d(2C-C_s)} + 1 \right) \tag{6}$$

for any $R \leq C$. In Appendix B, we include a proof of Eq. 3. The proof of Eq. 4 is essentially the same as that for Eq. 3, but is much more complicated.

From Eqs. 3-6, we see that $\overline{E_{\ell,d}(R)}$ can be made arbitrarily close to E(R) by proper selection of ℓ and d. Usually, the required number (ℓ) of vectors in the set X_ℓ is quite small.

All of the M code words may be sequentially generated (5) from the set X_k of ℓ basic waveforms. Thus, only ℓ waveforms have to be stored, rather than the $M = \exp(nR)$ waveforms which have to be stored with unrestricted optimal coding. Furthermore, ℓ is not a function of n. In our derivation we have used the probability density $P(x_k)$ that



$$\left.\begin{array}{l} \lim_{A\to\infty} K_d(A) = 1/d \\ \lim_{A\to 0} K_d(A) = 0 \end{array}\right\} \text{ for any } d$$

Fig. IX-7. The behavior of $K_d(A)$ as a function of A and d.

makes the upper and lower bounds of the probability of error for unrestricted coding close together for rates above R_{crit}. Thus, with a peak-power constraint we find that $E_{\ell,d}(R)$ can be made close to the optimum exponent of the probability of error.

Case II: All code words have exactly the same power nP.

The requirement that all code words shall have exactly the same power can be met by making each of the m elements in our signal space have exactly the same power. This additional constraint produces an additional reduction in the value of $E_{\ell,d}(R)$ that we have computed for the Gaussian channel. In this case, we have shown that

$$\overline{E_{\ell,d}(R)} \geq E(R) - K_d(A)\, E(0) - \frac{1}{d} \ln \left(\frac{e^{E(0)d} + \ell - 1}{\ell} \right)$$

for $R \leq R_{crit}$. Here, $K_d(A)$ is a function both of the signal-to-noise ratio A and of the length d of each element. This function is plotted in Fig. IX-7.

(a) Deterministic signal spaces. In the particular case in which the channel is Gaussian the probability of error is simply related to a set of energy distances between the code words, and it is better not to generate the set X_ℓ of ℓ vectors by choosing them at random. In order to make all code words have exactly the same power, each of the basic waveforms must be picked as points on a d-dimensional sphere of radius $(dP)^{1/2}$. The set X_ℓ is then preferably generated in such a way as to maximize the minimum distance between any two points (whenever possible). In Fig. IX-8 such a set X_ℓ is shown for the case d = 2. As an example, the binary case (d=1, ℓ=2) will be discussed.

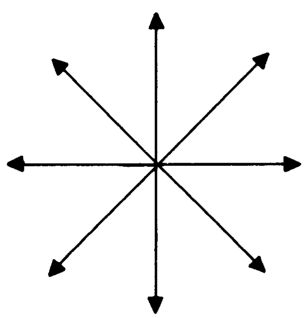


Fig. IX-8. Geometry of the set X_ℓ, for d = 2, ℓ = 8, for a Gaussian channel.

(b) Example: The binary case (d=1, ℓ=2) for Gaussian channels. In this case the set X_ℓ of basic waveforms consists of two oppositely directed vectors. For voltage signal-to-noise ratios $A \ll 1$, $E_{2,1}(R)$ is found to be exactly the same as the optimum exponent E(R) for all $R \leq C$, in the limit as $A \to 0$. In general,

$$E_{2,1}(R) = E(R) - K_1(A)\, E(0)$$

for $R \leq R_{crit}$. Values of $K_1(A)$ for different A's are given in Table IX-1. These results hold true when optimum decoding is performed. The case in which the channel output is quantized into two levels before data processing is considered in Appendix A.

We have established a discrete signal space in which we may hope to encode and decode in a sequential way. This restricted space is optimum in the sense that it yields

Table IX-1.

A	$K_1(A)$
1	0.01
2	0.1
3	0.28
4	0.43

a probability of error that can be made close to optimum through proper selection of ℓ and d for all rates above some R_{crit}.

The author wishes to acknowledge very helpful discussions with Professer J. M. Wozencraft and Professor E. M. Hofstetter. The original work of Professor R. M. Fano provided the foundation for this work, and the author's debt to him is obvious. This research was supported in part by a fellowship from the Government of Israel.

APPENDIX A

A Gaussian Channel Converted into a Binary Symmetric Channel

A Gaussian channel with a binary input ($d=1, \ell=2$) can be converted into a binary symmetric channel by quantizing the output of the channel. Let $E^q_{2,1}(R)$ be the exponent of the probability of error when the output of the channel is quantized as follows:

$$\left.\begin{array}{l}\text{For all } y \geq 0, \; y_q = 1 \\ \text{For all } y < 0, \; y_q = -1\end{array}\right\} \quad \text{for antipodal input elements}$$

where y is the output of the quantizer.

The channel is thus converted into a binary symmetric channel. Then it can be shown that

$$E^q_{2,1}(R) = E_{2,1}(R) - K_q(A)\, E(0)$$

for $R \leq R_{crit}$. For $K_q(A)$ we have

$$\lim_{A\to 0} K_q(A) = \frac{2}{\pi} \qquad \text{(A-1)}$$

$$\lim_{A\to \infty} K_q(A) = 0 \qquad \text{(A-2)}$$

On the other hand, it was established (see Table IX-1) that

$$E_{2,1}(R) \approx E(R) \qquad \text{for } A \leq 2, \text{ approximately}$$

Thus, combining these two results, we see that the loss incurred for small R/C in converting the general Gaussian channel into a binary symmetric channel is approximately equivalent to a signal power loss of 2 db, for all signal-to-noise ratios $A \leq 2$, approximately.

APPENDIX B

Unrestricted Random Coding

Let us consider a constant channel with input events represented by the points x of a d-dimensional space X and output events represented by the points y of a d-dimensional space Y. The conditional probability distribution $P(y|x)$ is defined by the channel and is independent of the past of the transmission. Each particular message of length n is constructed by selecting $m = n/d$ input events independently at random with probability P(x) from the input ensemble X.

For this situation, Fano (2) has shown that the average probability of error for all rates below R_{crit} is given by

$$P_e \leq MP_1 \qquad \text{for } R \leq R_{crit} \tag{B-1}$$

where $M = \exp(nR)$, $P_1 = \exp[m\gamma(1/2)]$, and

$$\gamma\left(\frac{1}{2}\right) = \ln \sum_y \left[\sum_x P(x)(P(y|x))^{1/2}\right]^2 \tag{B-1a}$$

Thus $P_e \leq \exp[-nE(R)]$, where $nE(R) = -m\gamma(1/2) - R$, and

$$E(R) = -\frac{m}{n}\gamma\left(\frac{1}{2}\right) - R = -\frac{1}{d}\gamma\left(\frac{1}{2}\right) - R \tag{B-2}$$

where $d = n/m$.

Restricted Random Coding

Next, let us consider the restricted random coding of the theorem, and evaluate the corresponding exponent of the probability of error. In this case, each code word of length n consists of a series of m elements with each element picked at random with probability $1/\ell$ from a subset X_ℓ of the X. The subset X_ℓ contains ℓ elements, and is generated in the following way: Each member x_k of the set X_ℓ is picked at random with probability $P(x_k)$ from the set X. Thus

$$P(x_k=x) = P(x) \qquad \text{for all } k \tag{B-3}$$

Given the subset X_ℓ, each of the m successive elements of a code word is generated by first picking the index k at random with probability $1/\ell$ and then taking x_k to be the element. Under these circumstances, by direct analogy with the unrestricted case, the probability of error can be bounded by

$$P(e|X_\ell) \leq MP_1' \qquad \text{for } R \leq R_{crit}$$

$$M = e^{nR}$$

$$P_1' = e^{m \cdot \gamma_{\ell, d}(1/2)}$$

where

$$\gamma_{\ell, d}\left(\frac{1}{2}\right) = \ln \sum_y \left[\sum_k P(k)(P(y|k))^{1/2}\right]^2$$

$$= \ln \sum_y \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \frac{1}{\ell^2} P(y|x_i)^{1/2} P(y|x_k)^{1/2} \tag{B-4}$$

Thus $P(e|X_\ell) \leq \exp[-nE_{\ell, d}(R)]$, where

$$E_{\ell, d}(R) = -\frac{m}{n} \gamma_{\ell, d}\left(\frac{1}{2}\right) - R$$

$$= -\frac{1}{d} \gamma_{\ell, d}\left(\frac{1}{2}\right) - R \tag{B-5}$$

The average value of $E_{\ell, d}(R)$ over all choices of the subset X_ℓ is given by

$$\overline{E_{\ell, d}(R)} = -\frac{1}{d} \overline{\gamma_{\ell, d}\left(\frac{1}{2}\right)} - R \tag{B-6}$$

Now

$$\gamma_{\ell, d}\left(\frac{1}{2}\right) = \ln g_{\ell, d}\left(\frac{1}{2}\right)$$

where

$$g_{\ell, d}\left(\frac{1}{2}\right) = \sum_y \sum_{j=1}^{\ell} \sum_{i=1}^{\ell} \frac{1}{\ell^2} P(y|x_i)^{1/2} P(y|x_j)^{1/2}$$

Thus

$$-\overline{\gamma_{\ell, d}\left(\frac{1}{2}\right)} = -\overline{\ln g_{\ell, d}\left(\frac{1}{2}\right)} \geq -\ln \overline{g_{\ell, d}\left(\frac{1}{2}\right)}$$

where

$$\overline{g_{\ell, d}\left(\frac{1}{2}\right)} = \sum_y \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \frac{1}{\ell^2} \overline{P(y|x_i)^{1/2} P(y|x_j)^{1/2}}$$

Thus

$$E_{\ell,d}(R) \geq -\frac{1}{d} \ln \overline{g_{\ell,d}\left(\frac{1}{2}\right)} - R \tag{B-7}$$

Now

$$\overline{P(y|x_i)^{1/2} P(y|x_j)^{1/2}} = \sum_{x_j x_i} P(x_i) P(x_j) P(y|x_i)^{1/2} P(y|x_j)^{1/2}$$

for all $i \neq j$, and

$$\overline{P(y|x_i)^{1/2} P(y|x_j)^{1/2}} = \sum_{x_i} P(x_i) P(y|x_i) = \sum_{x_j} P(x_j) P(y|x_j)$$

for $i = j$. There are ℓ terms in expression B-7 for which $i = j$, and there are $\ell^2 - \ell$ terms for which $i \neq j$. Thus

$$\overline{g_{\ell,d}\left(\frac{1}{2}\right)} = \sum_y \frac{\ell}{\ell^2} \sum_{x_i} P(x_i) P(y|x_i) + \sum_y \frac{\ell(\ell-1)}{\ell^2} \sum_{x_i} P(x_i) P(y|x_i)^{1/2} \sum_{x_j} P(x_j) P(y|x_j)^{1/2}$$

Now, since $P(x_i=x) = P(x_j=x) = P(x)$, we obtain

$$\overline{g_{\ell,d}\left(\frac{1}{2}\right)} = \frac{1}{\ell} \sum_y \sum_x P(x) P(y|x) + \frac{\ell-1}{\ell} \sum_y \left[\sum_x P(x)(P(y|x))^{1/2}\right]^2$$

or, from equation B-1a,

$$\overline{g_{\ell,d}\left(\frac{1}{2}\right)} = \frac{1}{\ell} + \frac{\ell-1}{\ell} e^{\gamma(1/2)} \tag{B-8}$$

Thus

$$\overline{g_{\ell,d}\left(\frac{1}{2}\right)} = e^{\gamma(1/2)}\left[\frac{e^{-\gamma(1/2)} + \ell - 1}{\ell}\right]$$

and

$$\ln \overline{g_{\ell,d}\left(\frac{1}{2}\right)} = \gamma\left(\frac{1}{2}\right) + \ln\left[\frac{e^{-\gamma(1/2)} + \ell - 1}{\ell}\right] \tag{B-9}$$

Thus, from inequality B-7 and Eq. B-9, we get

$$\overline{E_{\ell,d}(R)} \geq -\frac{1}{d} \gamma\left(\frac{1}{2}\right) - R - \frac{1}{d} \ln\left[\frac{e^{-\gamma(1/2)} + \ell - 1}{\ell}\right] \tag{B-10}$$

Using Eq. B-2, we have

$$\overline{E_{\ell,d}(R)} \geq E(R) - \frac{1}{d} \ln\left[\frac{e^{-\gamma(1/2)} + \ell - 1}{\ell}\right] \tag{B-11}$$

and also

$$-\gamma\left(\frac{1}{2}\right) = E(o) \cdot d$$

Thus

$$\overline{E_{\ell, d}(R)} \geq E(R) - \frac{1}{d} \ln \left[\frac{e^{dE(o)} + \ell - 1}{\ell}\right]$$

Q.E.D.

The extension of this proof to include continuous channels for which the function

$$\gamma\left(\frac{1}{2}\right) = \ln \int_y \left[\sum_x P(x) P(y|x)^{1/2}\right]^2$$

exists is straightforward.

J. Ziv

References

1. C. E. Shannon, Probability of error for optimal codes in a Gaussian channel, Bell System Tech. J. 38, 611 (1959).

2. R. M. Fano, Transmission of Information (The M.I.T. Press, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1961); see Chapter 9, Eqs. 9.116, 9.247, 9.248, 9.249.

3. J. L. Kelly, A class of codes for signaling on a noisy continuous channel, Trans. IRE, Vol. IT-6, No. 1, pp. 22-24 (March 1960).

4. J. M. Wozencraft, Sequential Decoding for Reliable Communication, Technical Report 325, Research Laboratory of Electronics, M.I.T., Aug. 9, 1957.

5. B. Reiffen, Sequential Encoding and Decoding for the Discrete Memoryless Channel, Technical Report 374, Research Laboratory of Electronics, M.I.T., Aug. 12, 1960; Technical Report 231, Lincoln Laboratory, M.I.T., Aug. 12, 1960.