X. PROCESSING AND TRANSMISSION OF INFORMATION[*]

Prof. W. B. Davenport, Jr.
Prof. P. Elias
Prof. R. M. Fano
Prof. R. G. Gallager
Prof. F. C. Hennie III
Prof. E. M. Hofstetter
Prof. D. A. Huffman
Prof. I. M. Jacobs
Prof. A. M. Manders
Prof. B. Reiffen
Prof. W. F. Schreiber
Prof. C. E. Shannon
Prof. J. M. Wozencraft
Dr. C. L. Liu
Dr. A. Wojnar

T. G. Adcock
T. M. Anderson
P. G. Arnold
M. H. Bender
E. R. Berlekamp
D. G. Botha
J. E. Cunningham
H. Dym
P. M. Ebert
D. Ecklein
D. D. Falconer
E. F. Ferretti
G. D. Forney, Jr.
R. L. Greenspan

U. F. Gronemann
P. W. Hartman
A. R. Hassan
J. L. Holsinger
T. S. Huang
N. Imai
R. J. Purdy
J. F. Queenan
J. E. Savage
J. R. Sklar
I. G. Stiglitz
W. R. Sutherland
O. J. Tretiak
W. J. Wilson
H. L. Yudkin

A. TWO-DIMENSIONAL POWER DENSITY SPECTRUM OF TELEVISION RANDOM NOISE

1. Introduction

Several workers have investigated the subjective effect of television random noise.[1-4] They have tried to relate the subjective effect of additive Gaussian noise to the one-dimensional power density spectrum of the noise considered as a function of time. Although the noise is one-dimensionally generated as a function of time, it is nevertheless displayed in two-dimensional fashion. Therefore, it has been our opinion that it might be more fruitful to try to relate the subjective effect of additive Gaussian noise to the two-dimensional power density spectrum of the noise considered as a function of two space variables. In this connection the following question naturally arises. A one-dimensional Gaussian noise with known power density spectrum $\Phi_o(\omega)$ is given. From this noise a two-dimensional noise is obtained by scanning. What is the power density spectrum $\Phi^*(u,v)$ of the two-dimensional noise in terms of $\Phi_o(\omega)$? In this report we shall attempt to answer this question.

2. Two-Dimensional Power Density Spectrum

We restrict our attention to black-and-white still pictures. The two-dimensional noise is completely specified by its magnitude $n^*(x,y)$ as a function of the space variables x and y. The geometry of the scanning lines is shown in Fig. X-1. We assume that the scanning is from left to right and from top to bottom, and that there is no interlacing.
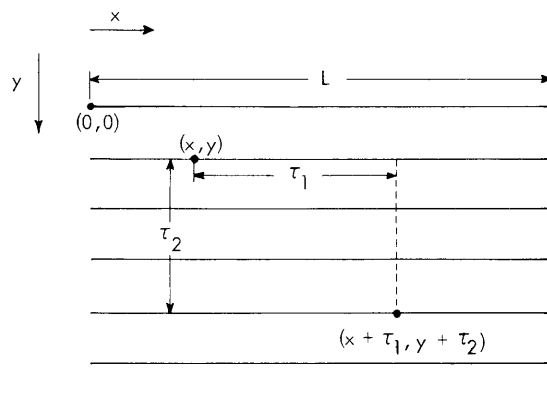

Fig. X-1. Geometry of scanning lines.

The length of each scan line is L, and the distance between two successive scan lines is taken as the unit length. Let the one-dimensional noise be $n_o(t)$, and let $(x,y) = (0,0)$ correspond to $t = 0$. Then we have

$$n^*(x,y) = n(x,y) \sum_k \delta(y-k),$$ (1)

where

$$n(x,y) = n_o(x+yL)$$ (2)

and

$$k = 0, \pm 1, \pm 2, \ldots; \quad -\infty < x < \infty, \quad -\infty < y < \infty.$$

Any particular noise picture can be considered as a finite piece of the sample function (1) which is infinite in extent. The impulses $\delta(y-k)$ are used so that we can deal with the Fourier transform of n^* instead of the Z-transform.

We assume that the one-dimensional noise is ergodic, then $n(x,y)$ is also ergodic. The autocorrelation function of $n(x,y)$ is

$$\phi(\tau_1,\tau_2) = \overline{n(x,y) \, n(x+\tau_1,y+\tau_2)}$$

$$= \overline{n(t) \, n(t+\tau_1+\tau_2 L)}$$

$$= \phi_o(\tau_1+\tau_2 L),$$ (3)

where $\phi_o(\tau)$ is the autocorrelation function of $n_o(t)$. Let $\Phi(u,v)$ and $\Phi_o(\omega)$ be the power density spectra of $n(x,y)$ and $n_o(t)$, respectively. We have

$$\Phi(u,v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(\tau_1, \tau_2) \, e^{j(\tau_1 u + \tau_2 v)} \, d\tau_1 d\tau_2$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_0(\tau_1 + \tau_2 L) \, e^{j(\tau_1 u + \tau_2 v)} \, d\tau_1 d\tau_2$$

$$= \frac{2\pi}{L} \, \delta\left(u - \frac{v}{L}\right) \, \Phi_0\left(\frac{v}{L}\right). \tag{4}$$

Figure X-2 shows a possible $\Phi_0(\omega)$ and its corresponding $\Phi(u,v)$. It is to be noted that $\Phi(u,v)$ is zero everywhere except on the line $u - \frac{v}{L} = 0$ where it is an impulse sheet.
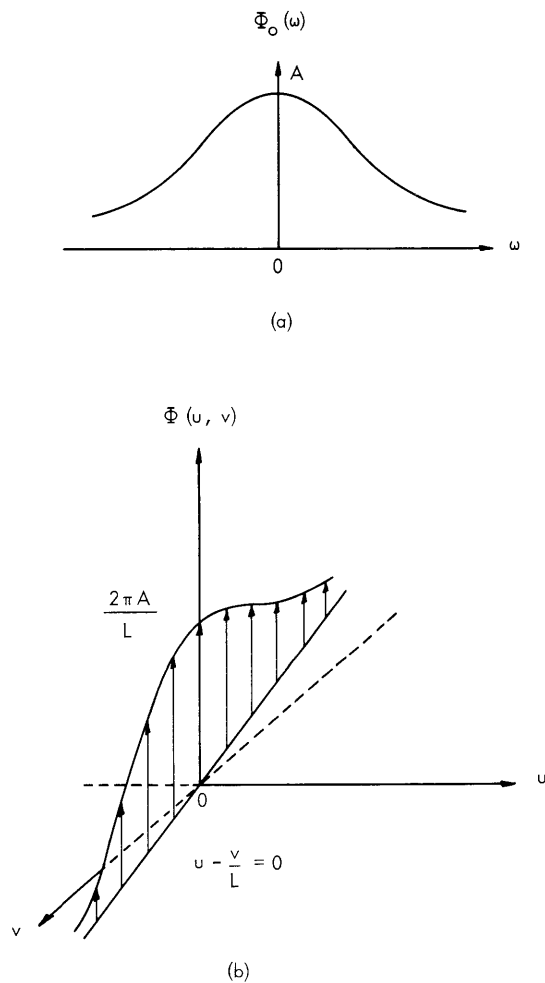

(a)

(b)

Fig. X-2. Relation between $\Phi(u, v)$ and $\Phi_0(\omega)$.

For ordinary commercial television, we have $L \approx 500$. Hence the line $u - \frac{v}{L} = 0$ is very close to the v-axis.

Letting $\Phi^*(u,v)$ be the power density spectrum of $n^*(x,y)$, we have

$$\Phi^*(u,v) = \sum_k \Phi(u,v+2\pi k), \tag{5}$$

where the summation is over all integers. $\Phi^*(u,v)$ consists of identical impulse sheets on the lines $u - \dfrac{v}{L} = \dfrac{2\pi k}{L}$, $k = 0, \pm 1, \pm 2, \ldots$.

3. One-Dimensional Power Density Spectrum along a Particular Direction

We have derived the power density spectrum $\Phi^*(u,v)$ of a two-dimensional noise in terms of the one-dimensional power density spectrum along the x-direction. It is reasonable to believe that the subjective effect of noise depends also on the one-dimensional power density spectra along directions other than the x-axis.
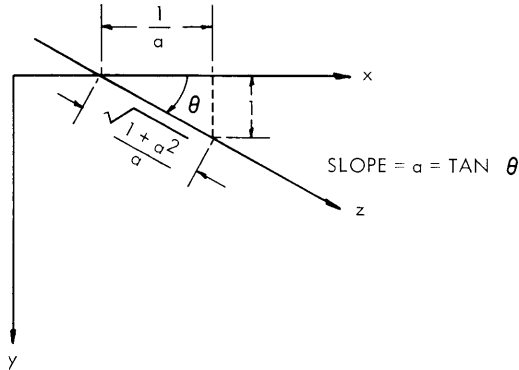
Fig. X-3. Calculation of the one-dimensional power density spectrum along the a direction.

We shall find the one-dimensional power density spectrum $\Phi_a^*(\omega)$ along a line of slope a (Fig. X-3). Let $n_a^*(z)$ be the noise along such a line:

$$n_a^*(z) = n_a(z) \sum_k \delta\left(z - \frac{\sqrt{1+a^2}}{a}k\right); \qquad k = 0, \pm 1, \pm 2, \ldots , \tag{6}$$

where

$$n_a(z) = n(z \cos\theta, z\cos\theta+b) \quad \text{for some fixed } b. \tag{7}$$

The autocorrelation function of $n_a(z)$ is

$$\phi_a(\tau) = \overline{n_a(z)\, n_a(z+\tau)}$$

$$= \overline{n(x,y)\, n(x+\tau\cos\theta,\, y+\tau\sin\theta)}$$

$$= \phi(\tau\cos\theta,\, \tau\sin\theta) \tag{8}$$

or

$$\phi_a'(\tau_1) \equiv \phi_a\left(\frac{\tau_1}{\cos\theta}\right) = \phi(\tau_1,\, a\tau_1). \tag{9}$$

The Fourier transform of $\phi_a'(\tau_1)$ is

$$\Phi_a'(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi(\omega-av,v)\, dv. \tag{10}$$

Hence the Fourier transform of $\phi_a(\tau_1)$ is

$$\Phi_a(\omega) = \frac{1}{|\cos\theta|}\, \Phi_a'\left(\frac{\omega}{\cos\theta}\right)$$

$$= \frac{1}{2\pi|\cos\theta|} \int_{-\infty}^{\infty} \Phi\left(\frac{\omega}{\cos\theta} - av, v\right) dv. \tag{11}$$

Putting Eq. 4 into Eq. 11, we have

$$\Phi_a(\omega) = \frac{1}{|(1+La)\cos\theta|}\, \Phi_0\left[\frac{\omega}{(1+La)\cos\theta}\right]$$

$$= \frac{\sqrt{1+a^2}}{|1+La|}\, \Phi_0\left(\frac{\sqrt{1+a^2}}{1+La}\,\omega\right). \tag{12}$$

In particular, for $a = 0$, we have

$$\Phi_0(\omega) = \Phi_0(\omega)$$

which checks with our assumption. For $a = \infty$, we have

$$\Phi_\infty(\omega) = \frac{1}{L}\, \Phi_0\left(\frac{\omega}{L}\right). \tag{13}$$

We note that for $L \approx 500$, the bandwidth of $\Phi_\infty(\omega)$ is 500 times that of $\Phi_0(\omega)$. Figure X-4 shows how the factor $\dfrac{\sqrt{1+a^2}}{|1+La|}$ varies with the slope a. We note that the area under $\Phi_a(\omega)$ is independent of a.
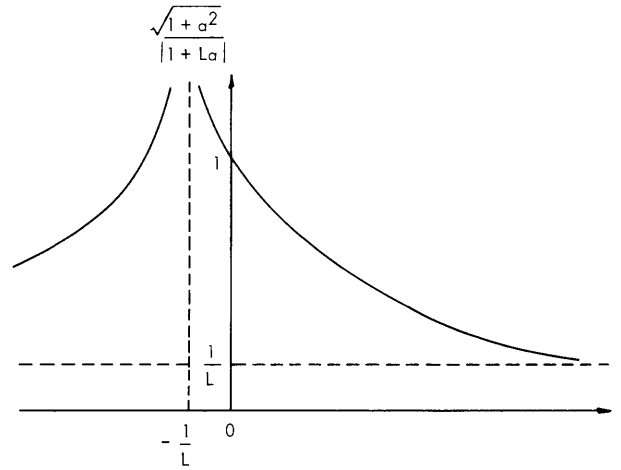
Fig. X-4. $\dfrac{\sqrt{1 + a^2}}{|1+La|}$ vs a.

Finally, from Eq. 6, we find that the one-dimensional power density spectrum along the direction a is

$$\Phi_a^*(\omega) = \sum_{k} \Phi_a\left(\omega + \frac{2\pi a}{\sqrt{1 + a^2}}k\right), \tag{14}$$

where $k = 0, \pm 1, \pm 2, \ldots$.

4. Discussion

From subjective tests, it has been found[1-4] that for pictures that are more or less isotropic, low-frequency noise (low-frequency when considered as a time function) is in general more annoying. From Eq. 12 we know, however, that a two-dimensional noise obtained from a low-frequency one-dimensional noise by scanning may contain high frequencies along directions other than the x-axis. In particular, the bandwidth along the y-axis is approximately 500 times as wide as that along the x-axis. Since the spatial frequency response of the human eye has a high-frequency cutoff, we suspect that the following hypothesis might be true for isotropic pictures contaminated by additive Gaussian noise. The more anisotropic a noise is, the more annoying it will be. Work is being carried out to test, among other things, this hypothesis, but the results are still not conclusive.

It is important to note that the mathematical model from which we obtained the power density spectra is quite crude. In order to relate the spectra to the subjective effect of noise, modifications may have to be made to take care

of the finiteness of scanning aperture.

We should like to thank W. L. Black who offered many very helpful comments on this work.

T. S. Huang

References

1. P. Mertz, Perception of TV random noise, SMPTE J. 54, 8-34 (January 1950).

2. P. Mertz, Data on random-noise requirements for theater TV, SMPTE J. 57, 89-107 (August 1951).

3. J. M. Barstow and H. N. Christopher, The measurement of random monochrome video interference, Trans. AIEE, Vol. 73, Part 1, pp. 735-741, 1954.

4. R. C. Brainard, F. W. Kammerer, and E. G. Kimme, Estimation of the subjective effects of noise in low-resolution TV systems, IRE Trans. on Information Theory, Vol. IT-8, pp. 99-106, February 1962.

B. SEQUENTIAL DECODING FOR AN ERASURE CHANNEL WITH MEMORY

1. Introduction

It has been shown that sequential decoding is a computationally efficient technique for decoding with high reliability on memoryless (or constant) channels.[1] It is of interest to determine whether or not similar statements are true for channels with memory.

In order to gain insight into the problem of sequential decoding with memory, a simple channel model, a two-state Markov erasure channel, has been analyzed. The presence of memory is shown to increase the average number of computations above that required for a memoryless erasure channel with the same capacity. For the erasure channel, the effects of memory can be reduced by scrambling.

2. Channel Model

Assume as a channel model an erasure channel with an erasure pattern that is generated by a Markov process. Assume the process shown in Fig. X-5. Assume for
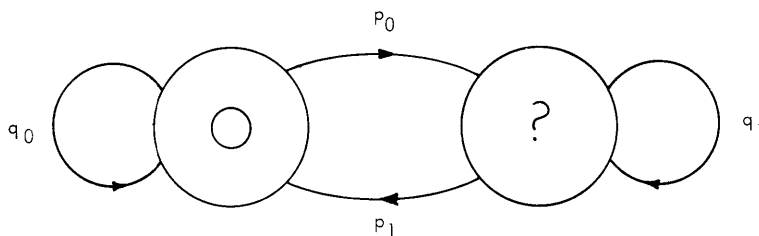

Fig. X-5. Noise model.

simplicity that the process begins in the 0 state. This will not affect the character of the results.

It can be shown that the probability of being in the 0 state after many transitions becomes independent of the starting probabilites and approaches $\frac{p_1}{p_0 + p_1}.^2$ Channel capacity is defined as

$$C = \lim_{n \to \infty} \frac{1}{n} \max_{p(x^n)} I(x^n; y^n),$$

where x^n and y^n are input and output sequences of n digits, respectively. Then, intuitively, $C = \frac{p_1}{p_0 + p_1}$, since information is transmitted only when the noise process is in

the 0 state and the frequency of this event approaches $n \frac{p_1}{p_0 + p_1}$ with large n.

3. Decoding Algorithim

We encode for this channel in the following fashion: A stream of binary digits is supplied by the source. Each digit selects ℓ digits from a preassigned tree (see Fig. X-6). If a digit has the value 1, the upper link at a node is chosen, the lower link being chosen when it is 0. (In our example (1, 0, 0) produces (011, 101, 000).)

Our object in decoding will be to determine the first information digit of a sequence of n digits. Having done this, we move to the next node and repeat the decoding process, again using n digits. The following decoding algorithim is used: Build two identical decoders that operate separately on each subset of the tree in time synchronism. Each
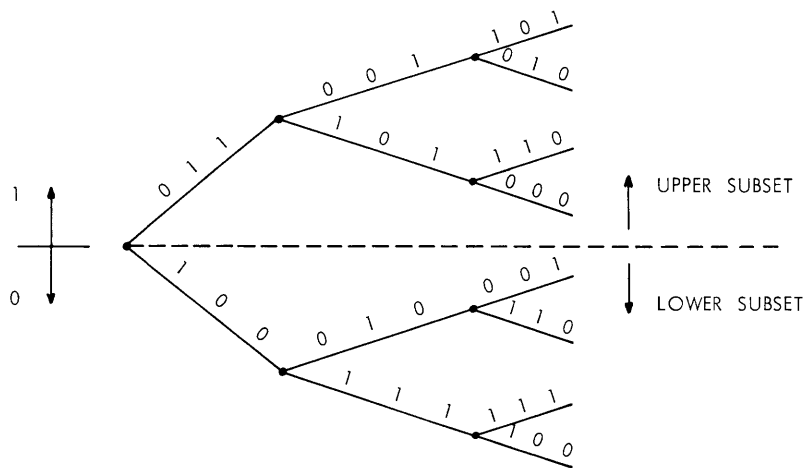

Fig. X-6. Encoding tree with $\ell = 3$.

decoder compares the received sequence with paths in its subset, discarding a path (and all longer paths with this as a prefix) as soon as a digit disagrees with an unerased received digit. When either a subset is identified (one subset is eliminated) or an ambiguity occurs (that is, when an acceptable path is found in each subject), we stop. If an ambiguity occurs, ask for a repeat. The computation, then, is within a factor of 2 of that necessary to decode the incorrect subset. We average the computation over the ensemble of all erasure patterns and all incorrect subsets.

4. Computation

Let x_i equal the number of nodes that are acceptable at the ith state of the incorrect subset. We trace two links for each node or perform two computations. Then, the average computation to decode the incorrect subset is I(n).

$$I(n) = 2 \sum_{i=0}^{nR-1} \overline{x_i},$$

where $R = \frac{1}{\ell}$, the rate of the source. We have

$$\overline{x_i} = 2^{i-1} P_R \qquad \text{(a path of } i\ell \text{ digits is acceptable)}$$

$$= 2^{i-1} P_{i\ell}.$$

Now

$$P_m = \sum_{r=0}^{m} \left(\frac{1}{2}\right)^{m-r} P_m(r),$$

where $P_m(r)$ is the probability of r erasures in m transitions. Then

$$I(n) = \sum_{i=0}^{nR-1} \left(\frac{1}{2}\right)^{i(\ell-1)} \sum_{r=0}^{i\ell} 2^r P_{i\ell}(r).$$

We recognize that the sum on r is the moment-generating function of erasures, $g_m(s)$, evaluated at s = ln 2.

$$g_m(s) = \sum_{e_m} e^{s\phi(e_m)} P(e_m),$$

where e_m is a sequence of m noise states and

$$\phi(e_m) = \sum_{i=1}^{m} \phi(e_i); \qquad \phi(x) = \begin{cases} 1 & x = ? \\ 0 & x \neq ? \end{cases}$$

It can be shown that

$$g_m(s) = \lfloor 1, \ 0 \rfloor \Pi^{m-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

where $\Pi = \begin{bmatrix} q_0 & p_0 e^s \\ p_1 & q_1 e^s \end{bmatrix}$, and that Π^m is asymptotically equal to the mth power of the

largest eigenvalue of Π. Then, we find the following asymptotically correct expression for $I(n)$

$$I(n) \simeq \begin{cases} \omega_0 & R < R_{comp} \\ \omega_1 2^{n(R-R_{comp})} & R > R_{comp} \end{cases}$$

where ω_0, ω_1 are constants and

$$R_{comp} = 1 - \log_2 \frac{1}{2} \left[3 - p_0\left(\frac{1+c}{1-c}\right) + \sqrt{\left(3-p_0\left(\frac{1+c}{1-c}\right)\right)^2 - 8\left(1 - \frac{p_0}{1-c}\right)^2} \right].$$

Using similar techniques, we can find an upper bound to the probability of ambiguity and show that it is equal to the random block-coding bound and that the zero rate exponent of this bound is R_{comp}.

The computational cutoff rate, R_{comp}, is shown as a function of p_0 for fixed capacity, c, in Fig. X-7. R_{comp} is an increasing function of p_0 for constant c. It is equal to R^o_{comp}, the memoryless rate, when $p_0 = q_1 = 1 - c$, and it exceeds R^o_{comp} when $p_0 > q_1$. However, $p_0 > q_1$ is not, in general, physically meaningful, since this situation corresponds to anticlustering.

5. Conclusions

We conclude that sequential decoding is inefficient when the channel becomes "sticky" (small p_0). It is possible, however, to reduce the effective memory of the channel and increase R_{comp} if we scramble before transmission and unscramble after reception.

Since the ergodic probability, $\frac{p_1}{p_0 + p_1}$, of erasures is not changed by this maneuver, channel capacity remains constant.

Most of these results appear to be extendable to binary-input, binary-output channels. It is not clear, however, that scrambling will improve performance on binary channels,
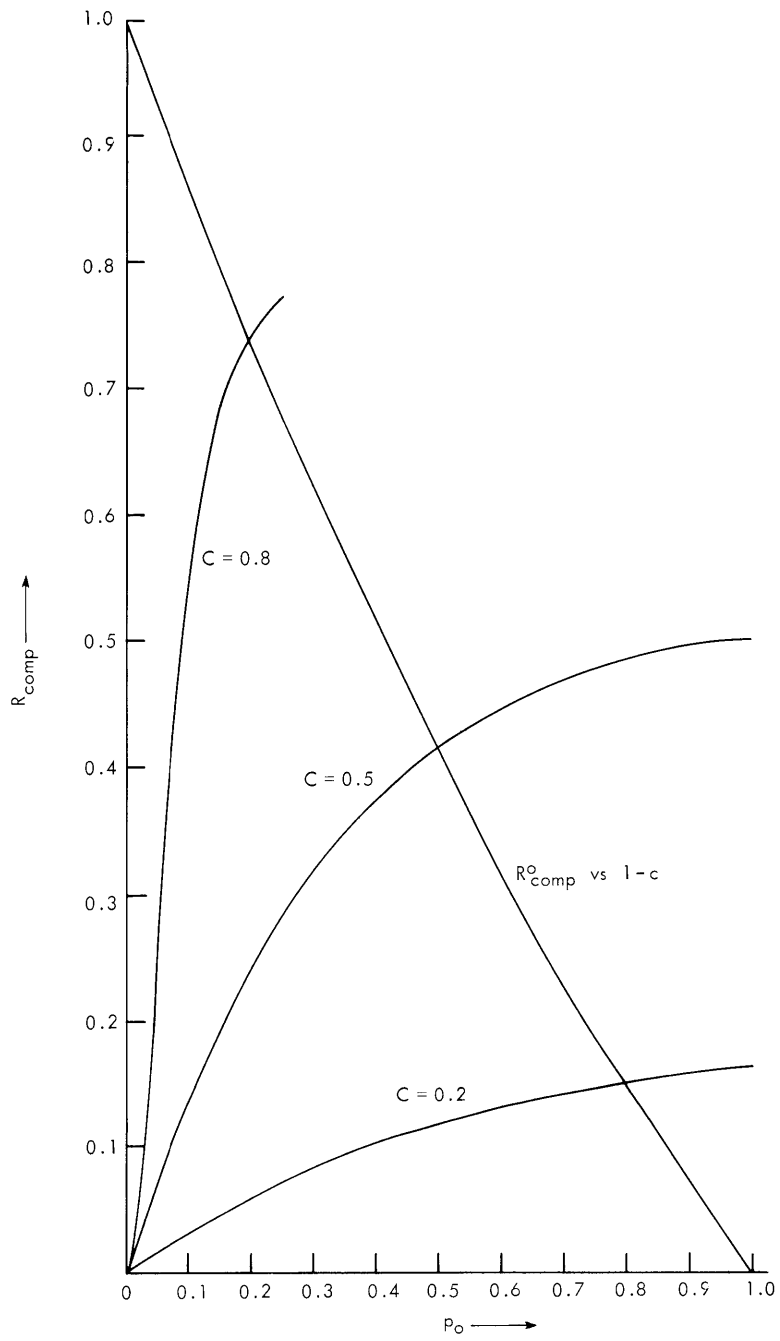
Fig. X-7. R_{comp} versus p_o.

since capacity is reduced in the process. Work is continuing on this problem.

J. E. Savage

References

1. J. M. Wozencraft and B. Reiffen, Sequential Decoding (The Technology Press of the Massachusetts Institute of Technology, Cambridge, Mass., 1961).

2. R. A. Howard, Dynamic Programming and Markov Processes (The Technology Press of the Massachusetts Institute of Technology, Cambridge, Mass., 1960), p. 7.

C. A SIMPLE DERIVATION OF THE CODING THEOREM

This report will summarize briefly a new derivation of the Coding Theorem. Let $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_M$ be a set of M code words for use on a noisy channel, and let $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_L$ be the set of possible received sequences at the channel output. The channel is defined by a set of transition probabilities on these sequences, $\Pr(\underline{y}|\underline{x}_m)$. Define

$$\Pr(\underline{y}) = \sum_m \Pr(\underline{x}_m)\,\Pr(\underline{y}|\underline{x}_m) \tag{1}$$

$$\Pr(\underline{x}_m|\underline{y}) = \frac{\Pr(\underline{y}|\underline{x}_m)\,\Pr(\underline{x}_m)}{\Pr(\underline{y})}. \tag{2}$$

For a given received $\underline{y}$, the decoder will select the number m for which $\Pr(\underline{x}_m|\underline{y})$ is a maximum. We call this number $m_{\underline{y}}$. The probability that the decoder selects correctly in this case is, then, $\Pr(\underline{x}_{m_{\underline{y}}}|\underline{y})$. Therefore the probability of decoding error is

$$P_e = \sum_{\underline{y}} \Pr(\underline{y})\left[1 - \Pr\!\left(\underline{x}_{m_{\underline{y}}}\big|\underline{y}\right)\right]. \tag{3}$$

We now upper-bound the expression in brackets in Eq. 3.

$$1 - \Pr\!\left(\underline{x}_{m_{\underline{y}}}\big|\underline{y}\right) = \sum_{m \neq m_{\underline{y}}} \Pr(\underline{x}_m|\underline{y}) \tag{4}$$

$$\leq \left[\sum_{m \neq m_{\underline{y}}} \Pr(\underline{x}_m|\underline{y})^{1/(1+\rho)}\right]^{1+\rho} \qquad \text{for any } \rho > 0. \tag{5}$$

Equation 5 follows from the fact that $t^{1/(1+\rho)}$ is a convex upward function of t for $\rho \geqslant 0$. Rewriting Eq. 5, we obtain

$$1 - \Pr\left(\underline{x}_{m_{\underline{y}}} \mid \underline{y}\right) \leqslant \sum_{m \neq m_{\underline{y}}} \Pr\left(\underline{x} \mid \underline{y}\right)^{1/(1+\rho)} \left[\sum_{m' \neq m_{\underline{y}}} \Pr\left(\underline{x}_{m'} \mid \underline{y}\right)^{1/(1+\rho)} \right]^{\rho} \qquad (6)$$

$$\leqslant \sum_{m} \Pr\left(\underline{x}_{m} \mid \underline{y}\right)^{1/(1+\rho)} \left[\sum_{m' \neq m} \Pr\left(\underline{x}_{m'} \mid \underline{y}\right)^{1/(1+\rho)} \right]^{\rho}. \qquad (7)$$

Equation 7 follows from overbounding the first term of Eq. 6 by summing over all m, and overbounding the second term by replacing the missing term, $\Pr\left(\underline{x}_{m_{\underline{y}}} \mid \underline{y}\right)$, with a smaller missing term, $\Pr\left(\underline{x}_{m} \mid \underline{y}\right)$. Now assume that the code words are equally likely, so that $\Pr\left(\underline{x}_{m}\right) = \frac{1}{M}$ for all m, and substitute Eq. 2 in Eq. 7.

$$1 - \Pr\left(\underline{x}_{m_{\underline{y}}} \mid \underline{y}\right) \leqslant \frac{1}{M \Pr(\underline{y})} \sum_{m} \Pr\left(\underline{y} \mid \underline{x}_{m}\right)^{1/(1+\rho)} \left[\sum_{m' \neq m} \Pr\left(\underline{y} \mid \underline{x}_{m'}\right)^{1/(1+\rho)} \right]^{\rho} \qquad (8)$$

$$P_e \leqslant \frac{1}{M} \sum_{\underline{y}} \sum_{m} \Pr\left(\underline{y} \mid \underline{x}_{m}\right)^{1/(1+\rho)} \left[\sum_{m' \neq m} \Pr\left(\underline{y} \mid \underline{x}_{m'}\right)^{1/(1+\rho)} \right]^{\rho}. \qquad (9)$$

Equation 9 bounds P_e for a particular code in terms of an arbitrary parameter $\rho > 0$. We shall now average this result over an ensemble of codes. For each m, let $\underline{x}_m$ be chosen independently according to some probability measure $P(\underline{x})$.

$$\overline{P}_e \leqslant \frac{1}{M} \sum_{\underline{y}} \sum_{m} \sum_{\underline{x}_m} P(\underline{x}_m) \Pr\left(\underline{y} \mid \underline{x}_m\right)^{1/(1+\rho)} \left[\sum_{m' \neq m} \Pr\left(\underline{y} \mid \underline{x}_{m'}\right)^{1/(1+\rho)} \right]^{\rho}. \qquad (10)$$

The bar over the last term in Eq. 10 refers to the average over the ensemble of all code words other than m. Now let $\rho \leqslant 1$. Then we are averaging over a convex upward function, and we can upper-bound Eq. 10 by averaging before raising to the power ρ. This is, then, the average of a sum of $M-1$ identically distributed random variables. Thus, for $0 \leqslant \rho \leqslant 1$,

$$\overline{P}_e \leqslant \frac{1}{M} \sum_{\underline{y}} \sum_{m} \sum_{\underline{x}_m} P(\underline{x}_m) \Pr\left(\underline{y} \mid \underline{x}_m\right)^{1/(1+\rho)} \left\{ (M-1) \sum_{\underline{x}} P(\underline{x}) \Pr\left(\underline{y} \mid \underline{x}\right)^{1+\rho} \right\}^{\rho}. \qquad (11)$$

Removing the index m in the sum over $\underline{x}_m$, and summing over m, we have

$$\overline{P}_e \leq \sum_{\underline{y}} \left[\sum_{\underline{x}} P(\underline{x}) \, Pr \, (\underline{y}|\underline{x})^{1/(1+\rho)} \right]^{1+\rho} (M-1)^\rho . \tag{12}$$

The bound in Eq. 12 applies to any channel for which $Pr \, (\underline{y}|\underline{x})$ can be defined, and is valid for all choices of $P(\underline{x})$ and all ρ, $0 \leq \rho \leq 1$. If the channel is memoryless (that is, if $Pr \, (\underline{y}|\underline{x}) = \prod_{n=1}^{N} Pr \, (\underline{y}_n|\underline{x}_n)$, where $\underline{y} = (\underline{y}_1, \cdots \underline{y}_n, \cdots \underline{y}_N)$, $\underline{x} = (\underline{x}_1, \cdots \underline{x}_n, \ldots, \underline{x}_N)$), then the bound can be further simplified. Let $P(\underline{x})$ be a probability measure that chooses each letter independently with the same probability (that is, $P(\underline{x}) = \prod_{n=1}^{N} P(\underline{x}_n)$). Then

$$\overline{P}_e \leq \sum_{\underline{y}} \left[\sum_{\underline{x}} P(\underline{x}) \prod_{n=1}^{N} Pr(\underline{y}_n|\underline{x}_n)^{1/(1+\rho)} \right]^{1+\rho} (M-1)^\rho . \tag{13}$$

The term in brackets in Eq. 13 is the average of a product of independent random variables, and is equal to the product of the averages. Thus

$$\overline{P}_e \leq \sum_{\underline{y}} \prod_{n=1}^{N} \left[\sum_{k=1}^{K} P(x_k) \, Pr \, (\underline{y}_n|x_k)^{1/(1+\rho)} \right]^{1+\rho} (M-1)^\rho , \tag{14}$$

where $x_1, \ldots, x_K$ are the letters in the channel-input alphabet. Applying an almost identical argument to the sum on $\underline{y}$, we obtain

$$\overline{P}_e \leq \left[\sum_{j=1}^{J} \left[\sum_{k=1}^{K} P(x_k) \, Pr \, (y_j|x_k)^{1/(1+\rho)} \right]^{1+\rho} \right]^{N} (M-1)^\rho , \tag{15}$$

where $y_1, \cdots y_J$ are the letters in the channel-output alphabet.

If the rate is defined in natural units as $R = \dfrac{\ln M}{N}$, Eq. 15 can be rewritten

$$\overline{P}_e \leq e^{-NE(\rho)} \qquad \text{for any } \rho, \; 0 \leq \rho \leq 1 \tag{16}$$

$$E(\rho) = E_o(\rho) - \rho R \tag{17}$$

$$E_o(\rho) = \ln \sum_{j} \left[\sum_{k} P(x_k) \, Pr \, (y_j|x_k)^{1/(1+\rho)} \right]^{1+\rho} . \tag{18}$$

It can be shown by straightforward but tedious differentiation that $E_o(\rho)$ has a positive first derivative and negative second derivative with respect to ρ for a fixed $P(x_k)$. Optimizing over ρ, we get the following parametric form for $E(\rho)$ as a function of R:

$$E(\rho) = E_o(\rho) - \rho E_o'(\rho) \tag{19}$$

$$R(\rho) = E_o'(\rho) \qquad \text{for } E_o'(1) \leqslant R \leqslant E_o'(0) \tag{20}$$

$$E = E_o(1) - R \qquad \text{for } R \leqslant E_o'(1). \tag{21}$$

From the properties of $E_o(\rho)$, it follows immediately that the E,R curve for a given choice of $P(x_k)$ appears as shown in Fig. X-8. $E_o'(0)$ turns out to be the mutual information on the channel for the input distribution $P(x_k)$. Choosing $P(x_k)$ to achieve channel
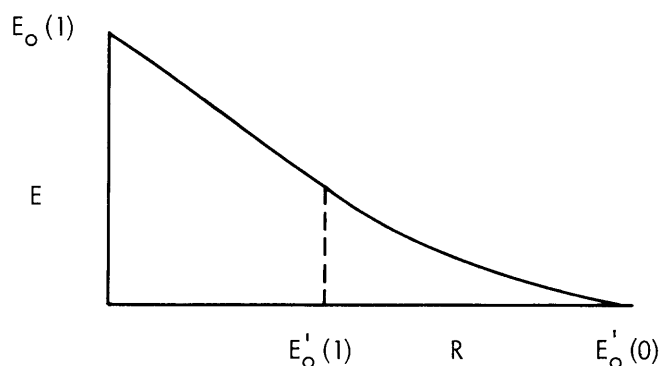

Fig. X-8. Exponent as a function of rate for fixed $P(x_k)$.

capacity, we see that for R < C, the probability of error can be made to decrease exponentially with the block length, N. For rates other than channel capacity, $P(x_k)$ must often be varied away from the capacity input distribution to achieve the best exponent.

Equations 19-21 are equivalent to those derived by Fano,[1] and for $E_o'(1) \leqslant R \leqslant E_o'(0)$, a lower bound to P_e can also be derived[1] of the form $P_e \geqslant K(N) e^{-NE(\rho)}$, where $E(\rho)$ is the same as in Eq. 19, and K(N) varies as a small negative power of N.

R. G. Gallager

References

1. R. M. Fano, The Transmission of Information (The M.I.T. Press, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1961), Chapter 9.