#4

# ON BINARY CHANNELS AND THEIR CASCADES

RICHARD A. SILVERMAN

LOAN COPY only

Technical Report 297

May 25, 1955

RESEARCH LABORATORY OF ELECTRONICS
Massachusetts Institute of Technology
Cambridge, Massachusetts

*Reprinted from the* IRE TRANSACTIONS OF THE PROFESSIONAL GROUP ON *INFORMATION THEORY*
Volume IT-1, Number 3, December, 1955
PRINTED IN THE U.S.A.

The Research Laboratory of Electronics is an interdepartmental laboratory of the Department of Electrical Engineering and the Department of Physics.

The research reported in this document was made possible in part by support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the Army (Signal Corps), the Navy (Office of Naval Research), and the Air Force (Office of Scientific Research, Air Research and Development Command), under Signal Corps Contract DA36-039 sc-42607, Project 102B; Department of the Army Project 3-99-10-022.

# On Binary Channels and Their Cascades

RICHARD A. SILVERMAN†

**Summary—A detailed analysis of the general binary channel is given, with special reference to capacity (both separately and in cascade), input and output symbol distributions, and probability of error. The infinite number of binary channels with the same capacity lie on double-branched equicapacity lines. Of the channels on the lower branch of a given equicapacity line, the symmetric channel has the smallest probability of error and the largest capacity in cascade, unless the capacity is small, in which case the asymmetric channel (with one noiseless symbol) has the smallest probability of error and the largest capacity in cascade. By simply reversing the designation of the output (or input) symbols, we can decrease the probability of error of any channel on the upper branch of the equicapacity line and increase the capacity in cascade of any asymmetric channel on the upper branch.**

**In a binary channel neither symbol should be transmitted with a probability lying outside the interval [1/e, 1 − (1/e)] if capacity is to be achieved. The maximally asymmetric input symbol distributions are approached by certain low-capacity channels. For these channels, redundancy coding permits an appreciable fraction of capacity in cascade if sufficient delay can be tolerated.**

## Capacity and Symbol Distributions

Discussion of the binary channel is usually confined to the symmetric case, where each of the transmitted digits is similarly perturbed by the noise. However, many interesting features of binary channels are concealed if only *symmetric* channels are considered. Accordingly, this paper will be devoted to a detailed study of the arbitrary binary channel.

Let the channel be characterized by the transition-probability matrix

$$C = \left\| \begin{matrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{matrix} \right\|, \qquad 0 \leq \alpha, \beta \leq 1, \tag{1}$$

where $\alpha$ is the probability that a zero be received as a zero, $\beta$ the probability that a one be received as a zero, etc. We shall use the symbol $C$ for both the channel and its matrix, but no confusion will arise. Computations are simplified by defining (after Muroga[1]) an auxiliary vector

$$\vec{X} = \left\| \begin{matrix} X_0 \\ X_1 \end{matrix} \right\|.$$

which solves the equation

$$C\vec{X} = -\vec{H}.$$

Here $\vec{H}$ is the row-entropy vector of the channel $C$; *i.e.*,

$$\vec{H} = \left\| \begin{matrix} H(\alpha) \\ H(\beta) \end{matrix} \right\|.$$

where $H(x)$ is the entropy function

$$H(x) = -x \log x - (1-x) \log (1-x).$$

(All logarithms are to the base 2 unless otherwise indicated; as another notation for $2^x$, we shall write $\exp_2 x$.) Muroga[1] has shown that the capacity $c(C)$ of the channel $C$ can be written in terms of the components of the vector $\vec{X}$ as

$$c(C) = \log (2^{X_0} + 2^{X_1}). \tag{2}$$

The transmitted and received symbol distributions are also simply related to $\vec{X}$. Let $\vec{P}$ be a vector representing the transmitted symbol distribution which achieves capacity, *i.e.*,

$$\vec{P} = \left\| \begin{matrix} P_0 \\ 1-P_0 \end{matrix} \right\|,$$

where $P_0$ is the probability with which zeros should be chosen if capacity (maximum rate) is to be achieved. Let $\vec{P}'$ be a vector representing the corresponding received symbol distribution, *i.e.*,

$$\vec{P}' = \left\| \begin{matrix} P_0' \\ 1-P_0' \end{matrix} \right\|,$$

where $P_0'$ is the probability that a zero will be received if the transmitted symbol distribution achieves capacity. The vectors $\vec{P}$ and $\vec{P}'$ are related by the equation

$$\vec{P}' = \tilde{C}\vec{P},$$

where $\tilde{C}$ denotes the transpose of the matrix $C$. In terms of the auxiliary vector $\vec{X}$, Muroga finds that

$$P_0 = \frac{2^{-c}}{\det (C)} \det \left\| \begin{matrix} 2^{X_0} & 2^{X_1} \\ \beta & 1-\beta \end{matrix} \right\|, \tag{3}$$

and that

$$P_0' = 2^{X_0 - c}. \tag{4}$$

Our task is to express the quantities (2), (3), and (4) in terms of the parameters $\alpha$ and $\beta$ of the binary channel (1). After some algebraic manipulation we find that

$$c(\alpha, \beta) = \frac{-\beta H(\alpha) + \alpha H(\beta)}{\beta - \alpha} + \log \left[ 1 + \exp_2 \left( \frac{H(\alpha) - H(\beta)}{\beta - \alpha} \right) \right], \tag{5}$$

$$P_0(\alpha, \beta) = \beta(\beta - \alpha)^{-1} - (\beta - \alpha)^{-1} \left[ 1 + \exp_2 \left( \frac{H(\beta) - H(\alpha)}{\beta - \alpha} \right) \right]^{-1}, \tag{6}$$

$$P_0'(\alpha, \beta) = \left[ 1 + \exp_2 \left( \frac{H(\beta) - H(\alpha)}{\beta - \alpha} \right) \right]^{-1}. \tag{7}$$

† Formerly Dept. of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Mass.; now Institute of Mathematical Sciences, Div. of Electromagnetic Res., New York Univ., N. Y.

[1] S. Muroga, "On the capacity of a discrete channel. I. *J. Phys. Soc. Japan*, vol. 8, pp. 484–494; July–August, 1953.

Each of the quantities (5), (6), and (7) depends on the channel parameters $\alpha$ and $\beta$: (5) gives the capacity of $C$, (6) the probability with which zeros should be chosen at the transmitter if capacity is to be achieved, and (7) the probability of a zero appearing at the receiver if zeros are chosen at the transmitter in accordance with (6).

Each of the functions $c(\alpha, \beta)$, $P_0(\alpha, \beta)$, and $P_0'(\alpha, \beta)$ defines a surface over the unit square $0 \leq \alpha, \beta \leq 1$. A study of the expressions (5), (6), and (7) reveals the following symmetries:

$$c(\alpha, \beta) = c(\beta, \alpha) = c(1 - \alpha, 1 - \beta) = c(1 - \beta, 1 - \alpha), \quad (8)$$

$$P_0(\alpha, \beta) = P_0(1 - \alpha, 1 - \beta) = 1 - P_0(\beta, \alpha) = 1 - P_0(1 - \beta, 1 - \alpha), \quad (9)$$

$$P_0'(\alpha, \beta) = P_0'(\beta, \alpha) = 1 - P_0'(1 - \alpha, 1 - \beta) = 1 - P_0'(1 - \beta, 1 - \alpha). \quad (10)$$

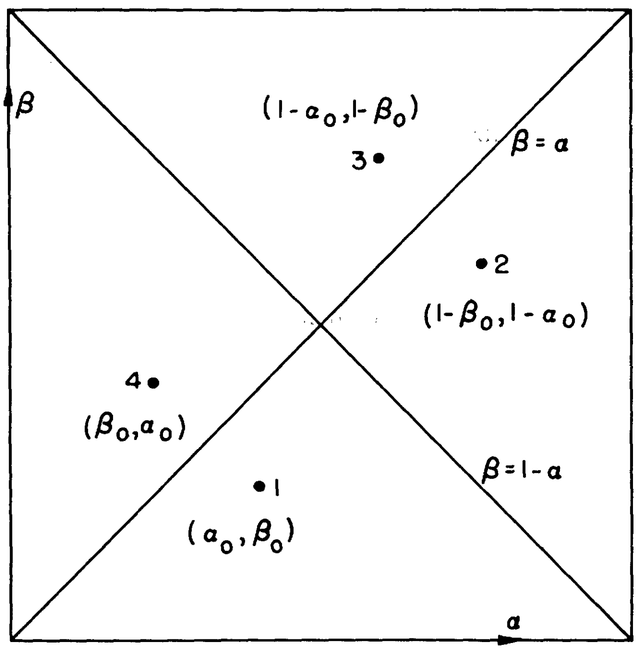


Fig. 1—Illustrating the symmetries of $c(\alpha, \beta)$, $P_o(\alpha, \beta)$, etc.

These symmetries are illustrated in Fig. 1, where a point $(\alpha_0, \beta_0)$ and its reflections in the $\beta = \alpha$ and $\beta = 1 - \alpha$ lines are shown. Eq. (8) shows that the capacity has the same value at any four such symmetrically placed points (for a reason to be discussed shortly), (9) that $P_0$ has the same value at points 1 and 3 and one minus that value at points 2 and 4, and (10) that $P_0'$ has the same value at points 1 and 4 and one minus that value at points 2 and 3.

Fig. 2 shows lines of constant capacity (equicapacity lines). Along the line $\beta = \alpha$ the capacity vanishes, corresponding to the vanishing of det $(C)$. The line $\beta = 1 - \alpha$ is the locus of symmetric channels, and along this line (5) reduces to the familiar expression

$$c(\alpha, \alpha) = 1 - H(\alpha).$$

Along the lines $\alpha = 0$, $\alpha = 1$, $\beta = 0$, and $\beta = 1$, (5) reduces to especially simple expressions. For example,

$$c(\alpha, 0) = \log\,[1 + \exp_2(-H(\alpha)/\alpha)], \qquad 0 \leq \beta \leq 1.$$

The slope of the curves $c(\alpha, 0)$ and $c(0, \beta)$ at the point $\alpha = \beta = 0$ is $\log e/e$. It is clear from Fig. 2 that there are an infinite number of binary channels with the same capacity. This is to be expected since *two* parameters are required to uniquely specify a binary channel.

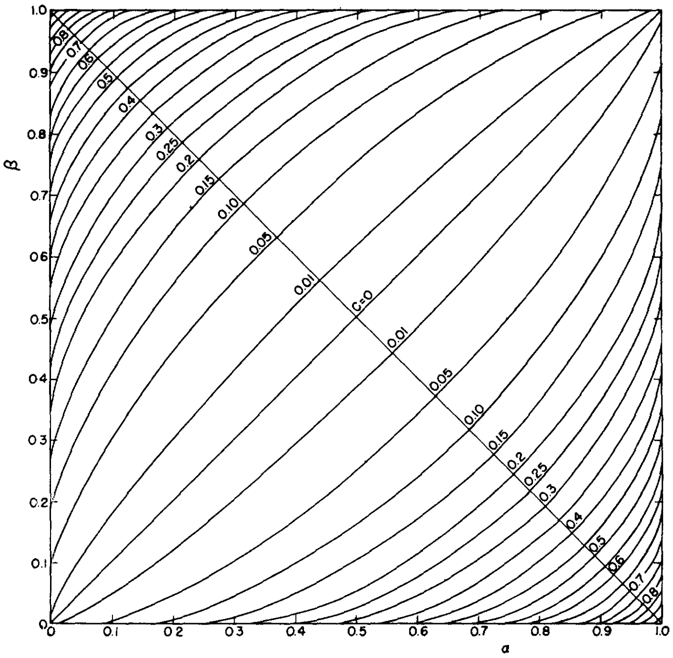


Fig. 2—Lines of constant $c(\alpha, \beta)$.

The fact that the four channels $C(\alpha, \beta)$, $C(\beta, \alpha)$, $C(1 - \alpha, 1 - \beta)$, and $C(1 - \beta, 1 - \alpha)$ have the same capacity, which produces two symmetrically placed branches of each equicapacity curve is easily explained. Clearly it is a matter of indifference which input (or output) symbol we choose to call a zero and which we choose to call a one. Reversing the designation of the input symbols corresponds to premultiplication by the noiseless matrix

$$I = \begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix},$$

and maps the channel $C(\alpha, \beta)$ into the channel $C(\beta, \alpha)$. Reversing the designation of the output symbols corresponds to postmultiplication by the matrix $I$ and maps the channel $C(\alpha, \beta)$ into the channel $C(1 - \alpha, 1 - \beta)$. Reversing the designation of *both* the input and output symbols corresponds to premultiplication *and* postmultiplication by the matrix $I$, and maps the matrix $C(\alpha, \beta)$ into the matrix $C(1 - \beta, 1 - \alpha)$. As (9) and (10) show, there are properties that, unlike capacity, are not invariant under all these mappings.

We have just seen that from a given point on an equicapacity line at least three other points can be reached by multiplying the channel matrix by another channel matrix. That there are no more such points is an immediate consequence of a theorem proved by DeSoer[2] to the effect that the capacity of two channels in cascade is less than the capacity of either unless one is a noiseless channel (*i.e.*, the unit matrix or one of its permutations) or unless one is a completely noisy (zero-capacity) channel.[3,4] (The reader is reminded that connecting two (or more) channels in cascade corresponds to multiplying the corresponding matrices.) Our statement follows from the fact that there are only two noiseless binary channels, namely $I$ and the unit matrix.

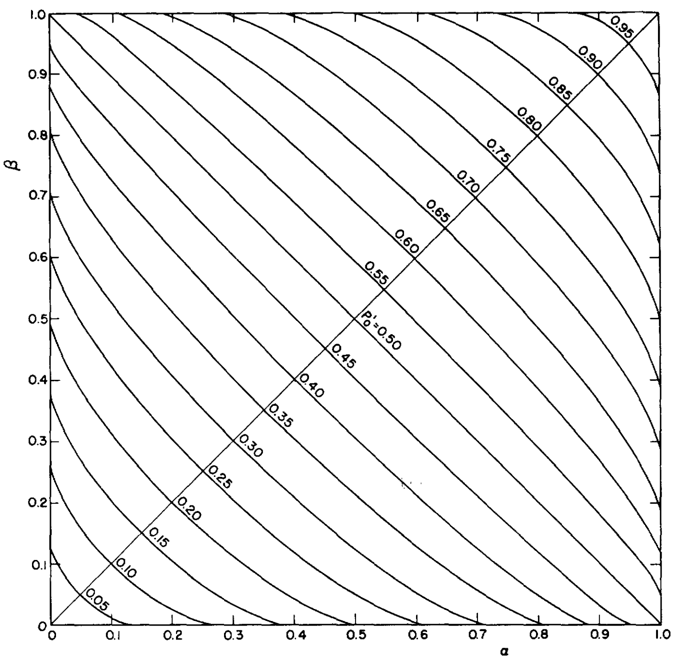


Fig. 3—Lines of constant $P'_o(\alpha, \beta)$.

Fig. 3 shows lines of constant $P_0'(\alpha, \beta)$, the probability of receiving a zero if the input symbol distribution achieves capacity. Along the line $\beta = 1 - \alpha$, the locus of symmetric channels, $P_0'(\alpha, \beta)$ has the familiar value $\frac{1}{2}$. Along the zero-capacity line $\beta = \alpha$, $P_0'(\alpha, \beta)$ has the limiting value $\alpha$, although (7) is indeterminate for $\beta = \alpha$. That is

$$\lim_{\epsilon \to 0} P_0'(\alpha + \epsilon, \alpha) = \lim_{\epsilon \to 0} P_0'(\alpha, \alpha + \epsilon) = \alpha.$$

Fig. 4 shows lines of constant $P_0(\alpha, \beta)$, the probability with which zeros should be transmitted to achieve capacity. The surface is saddle-shaped with the saddle point at $\alpha = \beta = \frac{1}{2}$. Along the symmetric channel line $\beta = 1 - \alpha$, $P_0(\alpha, \beta)$ has the familiar value $\frac{1}{2}$. Along the zero-capacity line $\beta = \alpha$, $P_0(\alpha, \beta)$ has the limiting value $\frac{1}{2}$, although (6) is indeterminate for $\beta = \alpha$. That is,

$$\lim_{\epsilon \to 0} P_0(\alpha + \epsilon, \alpha) = \lim_{\epsilon \to 0} P_0(\alpha, \alpha + \epsilon) = \tfrac{1}{2}.$$

The behavior of $P_0(\alpha, \beta)$ at the corners $\alpha = \beta = 0$ and $\alpha = \beta = 1$ is sufficiently remarkable to warrant special discussion.

Suppose we approach the point $\alpha = \beta = 0$ along the line $\alpha = \epsilon$, $\beta = r\epsilon$, where $0 \leq r < \infty$; *i.e.*, along any line between the positive $\alpha$-axis ($r = 0$) and the positive $\beta$-axis ($r = \infty$). Then $\lim_{\epsilon \to 0} P(\epsilon, r\epsilon)$ takes on all values between $1/e$ and $1 - (1/e)$, depending on the value of $r$, provided that for the value $r = 1$ (for which the single limit is indeterminate) we take the double limit $\lim_{\epsilon \to 0} \lim_{\epsilon' \to 0} P_0(\epsilon, \epsilon + \epsilon')$. For example,

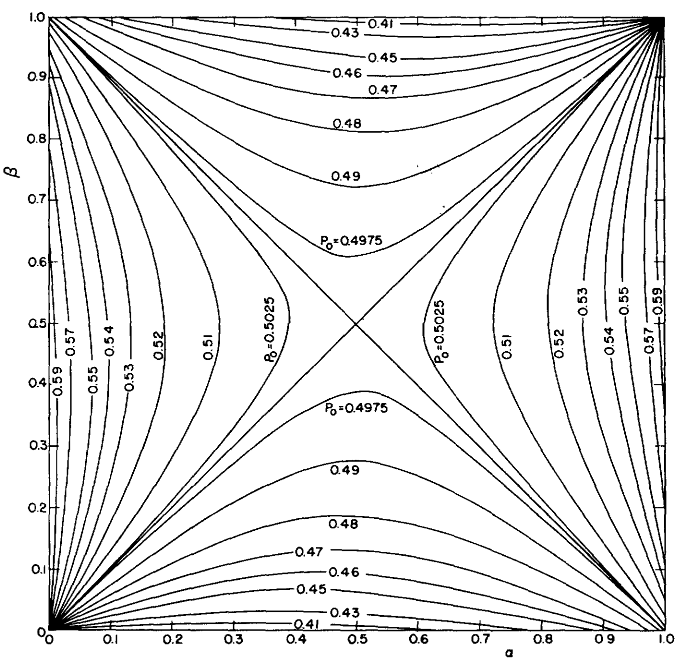


Fig. 4—Lines of constant $P_o(\alpha, \beta)$.

$$\lim_{\epsilon \to 0} P_0(\epsilon, 0) = \lim_{\epsilon \to 0} [\epsilon(1 + \exp_2 (H(\epsilon)/\epsilon)]^{-1}$$

$$= \lim_{\epsilon \to 0} \left[ e + \left(1 - \frac{e}{2}\right)\epsilon \right]^{-1} = 1/e,$$

$$\lim_{\epsilon \to 0} P_0(0, \epsilon) = 1 - \lim_{\epsilon \to 0} P_0(\epsilon, 0) = 1 - (1/e),$$

$$\lim_{\epsilon \to 0} P_0(\epsilon, \tfrac{1}{2}\epsilon) = \lim_{\epsilon \to 0} - 1 + [(e/4) + (\epsilon/2) - (3e\epsilon/16)]^{-1}$$

$$= (4/e) - 1,$$

$$\lim_{\epsilon \to 0} \lim_{\epsilon' \to 0} P_0(\epsilon, \epsilon + \epsilon') = \lim_{\epsilon \to 0} \lim_{\epsilon' \to 0}$$

$$1 + \frac{\epsilon(1 - \epsilon)}{2 \log e} H''(\epsilon) + \epsilon 0(\epsilon') = \frac{1}{2}.$$

[2] C. A. DeSoer, "Communication through channels in cascade," Sc.D. Thesis, January, 1953, Dept. of Elect. Eng., M.I.T.

[3] There is also the intermediate case where the channel matrix is reducible and one of the submatrices is completely noisy, *e.g.*, the channel with matrix

$$\begin{Vmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{Vmatrix}.$$

This channel (cited by Shannon in reference 4) is effectively the unit matrix, since the symbols corresponding to the first and second rows produce indistinguishable effects at the receiver.

[4] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, 1949, pp. 44–45.

In evaluating the limits we have made free use of the expressions

$$H(x) = x \log (e/x) - (x^2/2) \log e + 0(x^3),$$

$$H'(x) = (d/dx)H(x) = \log (1 - x)/x,$$

$$-H''(x) = -(d^2/dx^2)H(x) = x^{-1}(1 - x)^{-1} \log e.$$

We see that the point $\alpha = \beta = 0$ (and its image $\alpha = \beta = 1$ in the line $\beta = 1 - \alpha$) is a point of discontinuity of the function $P_0(\alpha, \beta)$, whose limiting behavior there depends on the direction of approach in a sort of "spiral staircase" fashion.

The maximum value of $P_0(\alpha, \beta)$ is the limiting value $1 - (1/e)$ obtained when we approach the point $\alpha = \beta = 0$ along the positive $\beta$-axis, and the minimum value of $P_0(\alpha, \beta)$ is the limiting value $1/e$ obtained when we approach the point $\alpha = \beta = 0$ along the positive $\alpha$-axis. (There are two corresponding limits at the point $\alpha = \beta = 1$.) Of course, the channel capacity is zero in both limits, so that there is no channel with positive capacity whose input symbol distribution is as asymmetric as $\vec{P} = [1/e, 1 - (1/e)]$ or $\vec{P} = [1 - (1/e), 1/e]$. However, there are low-capacity channels whose input symbol distributions are arbitrarily close to these maximally asymmetric ones. These low-capacity channels will be discussed further below.

We see that in a binary channel neither transmitted symbol can be selected with a probability lying outside the interval $[1/e, 1 - (1/e)]$ if capacity is to be achieved. If we are compelled to send digits from a more asymmetric distribution (as we may well be), the possibility of signaling at capacity is precluded from the start. Intuitively, this means that in a binary channel no decrease in equivocation obtained by skewing the input symbol distribution can justify making the source entropy less than $H(1/e)$, at least if obtaining maximum rate (capacity) is the objective.

For channels with larger alphabets it may be quite proper to choose one or more transmitted symbols with probability less than $1/e$, or indeed to suppress one or more transmitted symbols. Thus, for example, in the ternary channel

$$\begin{Vmatrix} \alpha & 1-\alpha & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1-\alpha & \alpha \end{Vmatrix},$$

capacity cannot be achieved if the symbol corresponding to the second row is transmitted, unless $\alpha < \alpha_0$, where $\alpha_0 \sim 0.64$ is the solution of the equation

$$\log \alpha = -\alpha.$$

Muroga gives many other examples of the need for suppressing possible transmitted symbols in his basic paper[1]. He was the first to point out the need of taking special care that $P_0$ does not become negative in capacity calculations.

The four channels which in the zero-capacity limit achieve one of the maximally asymmetric input distributions $\vec{P} = [1/e, 1 - (1/e)]$ or $\vec{P} = [1 - (1/e), 1/e]$ have matrices

$$\begin{Vmatrix} \epsilon & 1-\epsilon \\ 0 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 1-\epsilon & \epsilon \\ 1 & 0 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 1 \\ \epsilon & 1-\epsilon \end{Vmatrix}, \quad \begin{Vmatrix} 1 & 0 \\ 1-\epsilon & \epsilon \end{Vmatrix}. \tag{11}$$

We shall refer to these channels as "$\epsilon$-channels". To the first order in $\epsilon$, they all have capacity

$$c = \frac{\log e}{e} \epsilon \sim 0.53\epsilon \text{ bits.} \tag{12}$$

Introducing the abbreviation $k = (2 - e)/e \sim -0.27$, we find that the input symbol distribution which achieves capacity for the *first* and *second* of these channels is

$$\vec{P} = \begin{Vmatrix} (e + k\epsilon)^{-1} \\ 1 - (e + k\epsilon)^{-1} \end{Vmatrix}, \tag{13}$$

whereas for the *third* and *fourth* it is

$$\vec{P} = \begin{Vmatrix} 1 - (e + k\epsilon)^{-1} \\ (e + k\epsilon)^{-1} \end{Vmatrix}. \tag{14}$$

The corresponding output symbol distributions are

$$\vec{P}' = \begin{Vmatrix} \epsilon(e + k\epsilon)^{-1} \\ 1 - \epsilon(e + k\epsilon)^{-1} \end{Vmatrix} \tag{15}$$

for the *first* and *third* channels, and

$$\vec{P}' = \begin{Vmatrix} 1 - \epsilon(e + k\epsilon)^{-1} \\ \epsilon(e + k\epsilon)^{-1} \end{Vmatrix} \tag{16}$$

for the *second* and *fourth* channels. Eqs. (13) through (16) are accurate only to the first order in $\epsilon$.

## Probability of Error

We have seen that there are infinitely many binary channels with the same capacity. It is natural to ask whether there are contexts in which any of these channels with the same capacity is to be preferred to the others. Two questions that we might ask are: 1) Which of the channels with the same capacity has the smallest probability of error (in a received digit), and 2) Which has the largest end-to-end capacity if its output terminals are connected to the input terminals of an identical channel?[5] We answer the first question in this section and defer discussion of the second question until the next section.

The probability of error (at capacity) is given by the expression

$$P_\epsilon(\alpha, \beta) = \beta + (1 - \alpha - \beta)P_0(\alpha, \beta), \tag{17}$$

[5] This question arises naturally if we consider building up a cascade of repeaters, using a given binary channel as a unit.

where $P_0(\alpha, \beta)$ is the probability of a transmitted zero as given by (6). It is easily verified that $P_e(\alpha, \beta)$ has the following symmetries

$$P_e(\alpha, \beta) = P_e(1 - \beta, 1 - \alpha) = 1 - P_e(\beta, \alpha)$$
$$= 1 - P_e(1 - \alpha, 1 - \beta). \quad (18)$$

In deriving (18) free use has been made of the symmetries of $P_0(\alpha, \beta)$ as given by (9). Referring to Fig. 1, we see that $P_e(\alpha, \beta)$ has the same value at the points 1 and 2 and one minus that value at the points 3 and 4. (Note that none of the functions $c(\alpha, \beta)$, $P_0(\alpha, \beta)$, $P_0'(\alpha, \beta)$, and $P_e(\alpha, \beta)$ has the same symmetries.)

Fig. 5 shows lines of constant $P_e(\alpha, \beta)$. Along the symmetric channel line $\beta = 1 - \alpha$, $P_e(\alpha, 1 - \alpha)$ has the familiar value $1 - \alpha$. Along the zero-capacity line $\beta = \alpha$, $P_e(\alpha, \alpha)$ has the limiting value $\frac{1}{2}$. At the point $\alpha = 0$, $\beta = 1$ corresponding to the channel matrix $I$, $P_e = 1$, whereas at the point $\alpha = 1$, $\beta = 0$ corresponding to the unit matrix, $P_e = 0$.

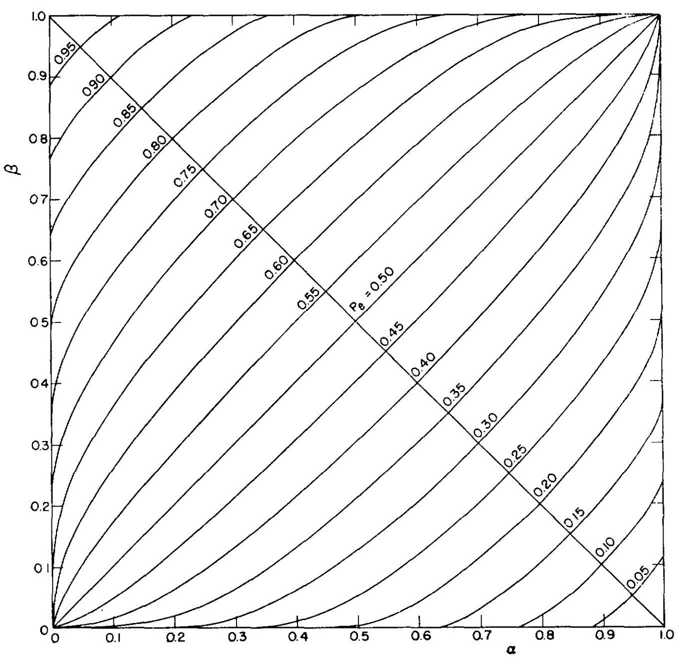


Fig. 5—Lines of constant $P_e(\alpha, \beta)$.

We have already noted that $P_0(\alpha, \beta)$ has discontinuities at the corners $\alpha = \beta = 0$ and $\alpha = \beta = 1$, and indeed that any value between $1/e$ and $1 - (1/e)$ can be obtained by approaching these discontinuities along the proper directions. Eq. (17) shows that $P_e(\alpha, \beta)$ shares these discontinuities. For since

$$\lim_{\alpha, \beta \to 0} P_e(\alpha, \beta) = \lim_{\alpha, \beta \to 0} P_0(\alpha, \beta),$$

the limiting behavior of $P_e(\alpha, \beta)$ at the two discontinuities is identical with that of $P_0(\alpha, \beta)$, however different the over-all appearance of the two sets of curves. This fact is apparent from Fig. 5. In particular, it follows that the curve $P_e(\alpha, \beta) = 1/e$ (not shown) must come into the points $\alpha = \beta = 0$ and $\alpha = \beta = 1$ with zero slope.

As we have seen, the portion of the equicapacity curve lying in the triangular region $\beta \leq \alpha$, $\beta \leq 1 - \alpha$ generates the rest of the equicapacity curve under the mappings corresponding to reversing the designation of the input symbols or output symbols or both. Eq. (18) shows that the probability of error for channels on the upper branch of the equicapacity curve (above the line $\beta = \alpha$) is greater than for channels on the lower branch (below the line $\beta = \alpha$), and indeed is greater than $\frac{1}{2}$. However, a value of $P_e$ greater than $\frac{1}{2}$ is artificial, for if communication is through such a channel, the receiver can obtain information at the same rate and with probability of error one minus that value merely by reversing the designation of the received symbols. (The transmitter need not be informed of this reversal, for (9) shows that the input symbol distribution remains the same in the reversed channel.) Thus our problem reduces to finding which of the channels on the portion of the equicapacity curve lying in the triangular region $\beta \leq \alpha$, $\beta \leq 1 - \alpha$ has the smallest probability of error.

The question is immediately answered if we superimpose the curves of Figs. 2 and 5. We find that a symmetric channel with a given capacity has a smaller probability of error than an asymmetric channel with the same capacity, *unless the channels have very low capacity.* In the latter case it is easily verified that the symmetric channel

$$\begin{Vmatrix} 1/2 + (\epsilon/2e)^{1/2} & 1/2 - (\epsilon/2e)^{1/2} \\ 1/2 - (\epsilon/2e)^{1/2} & 1/2 + (\epsilon/2e)^{1/2} \end{Vmatrix}$$

has the same capacity as the asymmetric $\epsilon$ channel

$$\begin{Vmatrix} \epsilon & 1 - \epsilon \\ 0 & 1 \end{Vmatrix},$$

namely

$$\frac{\log e}{e} \epsilon \text{ bits.}$$

Using (13) and (17), we find that the probability of error for the asymmetric channel is

$$(P_e)_{\text{asym.}} = (1 - \epsilon)/(e + k\epsilon) \qquad [k = (2 - e)/e],$$

whereas that of the symmetric channel is obviously

$$(P_e)_{\text{sym.}} = 1/2 - (\epsilon/2e)^{1/2}.$$

For small $\epsilon$, it is apparent that

$$(P_e)_{\text{asym.}} < (P_e)_{\text{sym.}},$$

as asserted. Indeed

$$\lim_{\epsilon \to 0} (P_e)_{\text{asym.}} = 1/e, \quad (19)$$

whereas

$$\lim_{\epsilon \to 0} (P_e)_{\text{sym.}} = \tfrac{1}{2}. \quad (20)$$

### Cascaded Channels

We turn to a discussion of the second of the questions raised at the beginning of the preceding section: Which of the channels with the same capacity has the largest end-to-end capacity if its output terminals are connected to the input terminals of an identical channel?

First we remind the reader (see the first section) that

$$c(C^2) < c(C),$$

*i.e.*, capacity is a decreasing function of the number of cascaded stages, unless $C$ is one of the noiseless channels

$$\begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix},$$

or one of the zero-capacity channels

$$\begin{Vmatrix} a & 1-a \\ a & 1-a \end{Vmatrix}. \tag{21}$$

The reader should further note that, unless $C$ is one of the two noiseless channels,

$$\lim_{n\to\infty} c(C^n) = 0.$$

For then $C$ is irreducible, and by a well-known theorem on Markov chains[6] $\lim_{n\to\infty} C^n$ has the form (21), and consequently

$$\lim_{n\to\infty} c(C^n) = c(\lim_{n\to\infty} C^n) = 0.$$

We begin by squaring the matrix $C(\alpha, \beta)$, obtaining

$$C^2 = \begin{Vmatrix} A(\alpha,\beta) & 1 - A(\alpha,\beta) \\ B(\alpha,\beta) & 1 - B(\alpha,\beta) \end{Vmatrix},$$

where

$$A(\alpha, \beta) = \alpha^2 + (1-\alpha)\beta,$$
$$B(\alpha, \beta) = \alpha\beta + (1-\beta)\beta.$$

Thus to every channel $(\alpha, \beta)$ on a given equicapacity curve corresponds a squared channel (cascaded with itself with matrix elements $A(\alpha, \beta)$ and $B(\alpha, \beta)$. The functions $A(\alpha, \beta)$ and $B(\alpha, \beta)$ exhibit the following symmetries:

$$\begin{aligned} A(\alpha, \beta) &= 1 - B(1-\beta, 1-\alpha), \\ B(\alpha, \beta) &= 1 - A(1-\beta, 1-\alpha), \end{aligned} \tag{22}$$

and

$$\begin{aligned} A(\beta, \alpha) &= 1 - B(1-\alpha, 1-\beta), \\ B(\beta, \alpha) &= 1 - A(1-\alpha, 1-\beta). \end{aligned} \tag{23}$$

However, there is no simple relation between $A(\alpha, \beta)$ and $A(\beta, \alpha)$, or between $B(\alpha, \beta)$ and $B(\beta, \alpha)$ [unless $\beta = 1 - \alpha$], so that two quite different curves are produced by squaring the matrices corresponding to a given equicapacity line, one originating from the upper branch of the equicapacity curve (above the line $\beta = \alpha$), and the other originating from the lower branch (below the line $\beta = \alpha$). The portion of the $(A, B)$ curve which originates from the upper branch of the equicapacity curve has no intercepts on the $\alpha$- and $\beta$-axes; we shall call it the *upper branch* of the $(A, B)$ curve. The portion of the $(A, B)$ curve which originates from the lower branch of the equicapacity curve has intercepts on the $\alpha$- and $\beta$-axes; we shall call it the *lower branch* of the $(A, B)$ curve. Eqs. (22) and (23) show that both branches of the $(A, B)$ curve are symmetric in the line $\beta = 1 - \alpha$. Moreover, since

$$A(\alpha, 1-\alpha) = A(1-\alpha, \alpha),$$
$$B(\alpha, 1-\alpha) = B(1-\alpha, \alpha),$$

the two branches contact on the line $\beta = 1 - \alpha$.

These facts are illustrated by Fig. 6, which shows three $(A, B)$ curves, those generated by squaring the channels with capacities 0.1, 0.4, and 0.7. The $(\alpha, \beta)$ values corresponding to these channels were read off the corresponding equicapacity lines of Fig. 2. Note that all the $(A, B)$ curves lie below the line $\beta = \alpha$. This is because $\beta > \alpha$ implies $B(\alpha, \beta) < A(\alpha, \beta)$, whereas $\beta < \alpha$ implies $B(\beta, \alpha) < A(\beta, \alpha)$.

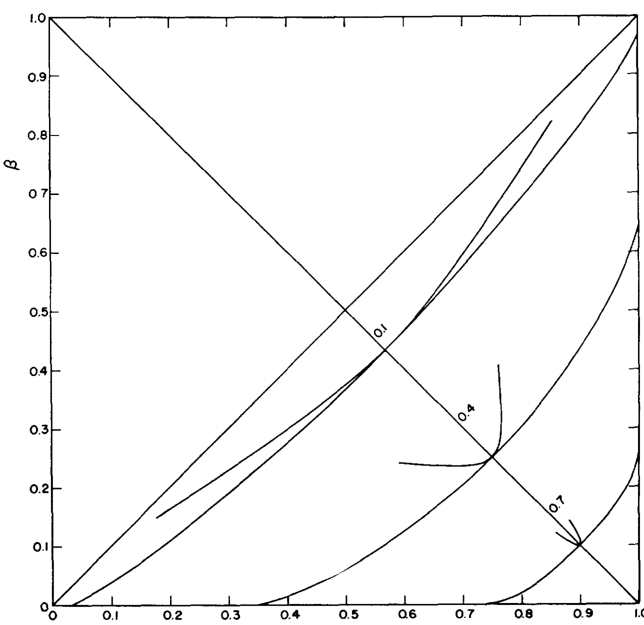


Fig. 6—$(A, B)$ curves corresponding to the channels of capacities 0.1, 0.4, and 0.7.

Of all the channels with the same capacity, the two corresponding to the end-points of the upper branch of the $(A, B)$ curve have the smallest capacity under cascade. Moreover, any channel on the upper branch of the $(A, B)$ curve has lower capacity than its images (under multiplication by $I$) on the lower branch. However, we can avoid the low capacity in cascade exhibited by these channels by an extremely simple intermediate station behavior, namely by crossing the connections between the outputs

[6] W. Feller, "Probability Theory and Its Applications," vol. 1 New York, John Wiley and Sons, 1950. Reference is made to Theorem 2, p. 325.

of one channel and the inputs of the next. In this way we arrive on the lower branch of the $(A, B)$ curve, which has higher capacity in cascade. (It will be recalled that in the preceding section we avoided probabilities of error greater than $\frac{1}{2}$ by exactly the same expedient.) We shall comment on the significance of this intermediate station behavior below.

The problem thus reduces to finding the channel on the lower branch of the $(A, B)$ curve with the highest capacity. As in the preceding section, we resort to a superposition of curves, this time superimposing the curves of Figs. 2 and 6. We find that a symmetric channel with a given capacity has a higher capacity under cascade than an asymmetric channel with the same capacity, *unless the channels have very low capacity.* As an example of this exceptional behavior at low capacity, we cite again the two channels

$$\begin{Vmatrix} \epsilon & 1-\epsilon \\ 0 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 1/2 + (\epsilon/2e)^{1/2} & 1/2 - (\epsilon/2e)^{1/2} \\ 1/2 - (\epsilon/2e)^{1/2} & 1/2 + (\epsilon/2e)^{1/2} \end{Vmatrix}, \tag{24}$$

which both have capacity of $(\epsilon \log e)/e$ bits. The squares of the matrices (24) are

$$\begin{Vmatrix} \epsilon^2 & 1-\epsilon^2 \\ 0 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 1/2 + \epsilon/e & 1/2 - \epsilon/e \\ 1/2 - \epsilon/e & 1/2 + \epsilon/e \end{Vmatrix},$$

respectively. The corresponding capacities are

$$c[C^2_{\text{asym.}}] = \frac{\log e}{e} \epsilon^2 \text{ bits}, \tag{25}$$

and

$$c[C^2_{\text{sym.}}] = \frac{2 \log e}{e^2} \epsilon^2 \text{ bits} \tag{26}$$

Thus

$$c[C^2_{\text{asym.}}] > c[C^2_{\text{sym.}}],$$

as asserted.[7]

We can now answer both questions posed at the beginning of the preceding section as follows: Of all the binary channels on the lower branch of a given equicapacity line, the symmetric channel has the smallest probability of error and the largest capacity in cascade, *unless the capacity is small,* in which case the asymmetric channel (with one noiseless symbol, *i.e.*, $\beta = 0$) has the smallest probability of error and the largest capacity in cascade. Continuity requires that there be a small range of values of the capacity for which asymmetric channels with $\beta \neq 0$ are superior in these two respects, but we have made no detailed study of this intermediate case. There is no point in trying to establish a preference among the channels on the upper branch of the equicapacity line, since by reversing the designation of the output (or input) symbols we arrive at a channel with a smaller probability of error and a larger capacity in cascade. (An exception to this statement in the symmetric case is noted below.)

[7] Eqs. (19), (20), (25), and (26) suggest the conjecture that the capacity in cascade of channels with the same *low* capacity is inversely proportional to their probabilities of error (see also (32) in the Appendix).

A particularly striking example of the difference between the capacity in cascade of a channel on the upper branch of the equicapacity line and its images on the lower branch is afforded by comparing the two $\epsilon$-channels

$$C_\epsilon = \begin{Vmatrix} \epsilon & 1-\epsilon \\ 0 & 1 \end{Vmatrix}, \quad C'_\epsilon = C_\epsilon I = \begin{Vmatrix} 1-\epsilon & \epsilon \\ 1 & 0 \end{Vmatrix}.$$

$C_\epsilon$ is on the lower branch of the equicapacity line (with capacity $(\epsilon \log e)/e$ bits), and $C'_\epsilon$, obtained by reversing the designation of the output terminals of $C_\epsilon$, is on the upper branch. The squares of $C_\epsilon$ and $C'_\epsilon$ are

$$C^2_\epsilon = \begin{Vmatrix} \epsilon^2 & 1-\epsilon^2 \\ 0 & 1 \end{Vmatrix}$$

and

$$C'^2_\epsilon = \begin{Vmatrix} 1-\epsilon+\epsilon^2 & \epsilon-\epsilon^2 \\ 1-\epsilon & \epsilon \end{Vmatrix}.$$

The corresponding capacities are

$$c[C^2_\epsilon] = \frac{\log e}{e} \epsilon^2 \text{ bits}$$

and

$$c[C'^2_\epsilon] = \frac{\log e}{8} \epsilon^3 \text{ bits},$$

so that, in this extreme case, the capacity of one channel is an order of magnitude less than that of the other.

DeSoer[2] has emphasized the importance of proper intermediate station behavior in maximizing the end-to-end capacity of a cascade of channels. In particular, he compares the capacity of a cascade of continuous channels perturbed by white Gaussian noise with that of a cascade of PCM channels with the same signal-to-noise ratio. It is assumed that in the cascade of continuous channels the intermediate stations retransmit the received waveform without change, whereas in the cascade of PCM channels requantization occurs at each intermediate station. Although the continuous channel has a higher capacity than the PCM channel, the PCM channels deteriorate less in cascade. Thus, for some cascade length depending on the signal-to-noise ratio, the cascade of PCM channels has a larger end-to-end capacity than the cascade of continuous channels.

We have an even simpler example of how proper intermediate station behavior can preserve the end-to-end capacity of cascaded channels. For, in the example just given, if the output symbols of the $C'_\epsilon$ channel are reversed at the intermediate station, we have

$$c[C'_\epsilon \; I \; C'_\epsilon] = c[C^2_\epsilon] \gg c[C'^2_\epsilon].$$

On the other hand, the capacity of a cascade of *symmetric* channels is completely insensitive to whether the identity of the symbols is preserved or reversed at the intermediate stations. Graphically, this is a consequence of the point of contact of the two branches of the $(A, B)$ curve on the symmetric channel line $\beta = 1 - \alpha$ (See Fig. 6).

Suppose we regard the behavior of the intermediate station as a detection scheme. Then preserving the designation of the output symbols of the preceding channel at the intermediate station is minimum probability of error detection[8] if the channel lies in the square region $\alpha \geq \frac{1}{2}$, $\beta \leq \frac{1}{2}$, and reversing the designation of the output symbols is minimum probability of error detection if the channel lies in the square region $\alpha \leq \frac{1}{2}$, $\beta \geq \frac{1}{2}$. If the channel lies in the square region $\alpha, \beta \leq \frac{1}{2}$, then minimum probability of error detection requires that both zeros and ones be changed to ones, and information-destroying mapping. Similarly, if the channel lies in the square region $\alpha, \beta \geq \frac{1}{2}$, then minimum probability of error detection requires that both zeros and ones be changed to zeros. Thus it is apparent that maximum rate in cascade and minimum probability of error detection at intermediate stations are not always compatible. (DeSoer[2] gives a complicated example that illustrates this fact.) If information-destroying mappings are precluded, as they must be if maximum rate is the objective, we conclude that the larger capacity in cascade is given by minimum probability of error detection, which requires that the identity of the symbols be preserved at the intermediate station if the channel lies below the line $\beta = \alpha$, but reversed if the channel lies above the line $\beta = \alpha$.

The proofs of the statements of the preceding paragraph are left to the reader. It is merely necessary to examine the expressions for the probabilities of error of each of the four delayless detection schemes and ascertain which is smallest in each square region.

## Appendix

### *Redundancy Coding in the $\epsilon$-Channel*

In our discussion of cascaded channels in the last section we considered only delayless operation of the intermediate station. If sufficient intermediate station delay is allowed, it follows from Shannon's second coding theorem that the end-to-end capacity of a cascade of identical channels can be made arbitrarily close to the common capacity of the separate channels. Studies of probability of error and rate as a function of delay are still in progress,[9] and it is perhaps too early to apply the theory to cascaded channels. However, the $\epsilon$-channel is susceptible to a simple type of *redundancy* coding, which is effective just because of its low capacity. This redundancy coding, although not ideal in the sense of achieving capacity with a vanishingly small probability of error, nonetheless achieves a rate which is an appreciable fraction of capacity with a small probability of error. Moreover, it serves to illustrate how delay can be exchanged for enhanced rate in a cascade of channels.[10]

In the redundancy coding to which we refer, each transmitted digit is repeated $r$ times, and the receiver decides whether a zero or a one was sent by examining sequences of $r$ digits. More specifically, let the channel have matrix

$$C_\epsilon = \begin{Vmatrix} \epsilon & 1-\epsilon \\ 0 & 1 \end{Vmatrix},$$

with capacity $(\log e)/e\ \epsilon$ bits. Have the receiver examine the output in blocks of $r$ symbols (properly synchronized with the transmitter) and decode a block of $r$ ones as a one and a block of $r$ digits with a zero at *any* position as a zero. In other words, the symbols 0 and 1 are mapped into the sequences $00 \cdots 0$ ($r$ times) and $11 \cdots 1$ ($r$ times) at the transmitter, and the events $S$ and $F$ are mapped into 0 and 1 at the receiver, where $S$ designates the appearance of a zero in a block of $r$ digits, and $F$ the nonappearance of a zero in a block of $r$ digits.[11] Transmission can then be regarded as taking place in an equivalent channel $C(r)$ with matrix

$$C(r) = \begin{Vmatrix} 1-(1-\epsilon)^r & (1-\epsilon)^r \\ 0 & 1 \end{Vmatrix}.$$

The capacity of $C(r)$ is

$$c[C(r)] = \log\left[1 + \exp_2\left(\frac{-H((1-\epsilon)^r)}{1-(1-\epsilon)^r}\right)\right],$$

and its probability of error is

$$P_e(r) = (1-\epsilon)^r P_0(\mathrm{r}),$$

where by $P_0(r)$ we mean the probability that the sequence $00 \cdots 0$ ($r$ times) should be transmitted if the capacity $c[C(r)]$ is to be achieved. $P_0(r)$ is not the same as $P_0$ for the $\epsilon$-channel, as given by (13) of the first section.

Suppose that each of the transmitted symbols is repeated $r = n/\epsilon$ times. Then, since $(1-\epsilon)^{n/\epsilon} \sim e^{-n}$ for small $\epsilon$, the equivalent channel becomes

$$C(n) = \begin{Vmatrix} 1-e^{-n} & e^{-n} \\ 0 & 1 \end{Vmatrix},$$

with capacity

$$c[C(n)] = \log\left[1 + \exp_2\left(\frac{-H(1-e^{-n})}{1-e^{-n}}\right)\right], \tag{27}$$

input symbol distribution

$$P_0(n) = (1-e^{-n})^{-1}\left[1 + \exp_2\left(\frac{H(1-e^{-n})}{1-e^{-n}}\right)\right]^{-1}, \tag{28}$$

and probability of error

$$P_e(n) = e^{-n}P_0(n). \tag{29}$$

[8] This is sometimes called maximum *a posteriori* probability detection or the ideal observer.

[9] Reference is made to recent work by C. E. Shannon and by P. Elias, presented at the March, 1955 National Convention of the IRE.

[10] Tentative studies of the effect of delay in cascaded channels have also been made by DeSoer[2].

[11] Of course, redundancy coding is also effective in a low capacity *symmetric* channel, but the analysis is more complicated, since now the events $S$ and $F$ refer to receiving more zeros than ones in a block of $r$ received digits, and vice versa.

Of course, before comparing the capacity of the redundant channel with the capacity of the $\epsilon$-channel, we must normalize (27) by dividing it by $r = n/\epsilon$, since only one information digit is transmitted every $r$ units of time. Moreover, we must now accept a delay of $r$ units. The interesting result (displayed in Table I at the end of this appendix) is that even when properly normalized, the redundancy code (which is, after all, a very simple code) gives rates which are an appreciable fraction of the capacity of the $\epsilon$-channel, with a probability of error which becomes smaller as we tolerate more delay. (Unfortunately, the capacity goes to zero with the probability of error, which is not the case for ideal coding.) Note that as more redundancy is introduced, $C(n)$ becomes a better approximation to the unit matrix, and $P_0(n)$ approaches $\frac{1}{2}$.

TABLE I

| $n$ | $c[C(n)]$ | $c[C(n)](\epsilon/n)$ | $P_0(n)$ | $P_e(n)$ |
|---|---|---|---|---|
| 1 | 0.436 | 0.436$\epsilon$ | 0.413 | 0.152 |
| 2 | 0.707 | 0.353$\epsilon$ | 0.448 | 0.061 |
| 3 | 0.858 | 0.286$\epsilon$ | 0.472 | 0.024 |
| 4 | 0.934 | 0.234$\epsilon$ | 0.487 | 0.0089 |
| 5 | 0.971 | 0.194$\epsilon$ | 0.493 | 0.0033 |
| 6 | 0.987 | 0.165$\epsilon$ | 0.497 | 0.0012 |
| 7 | 0.994 | 0.142$\epsilon$ | 0.499 | 0.00045 |
| 8 | 0.998 | 0.125$\epsilon$ | 0.500 | 0.00017 |
| 9 | 0.999 | 0.111$\epsilon$ | 0.500 | 0.00006 |
| 10 | 1.000 | 0.100$\epsilon$ | 0.500 | 0.00002 |

(For no redundancy: $c[C_\epsilon] = 0.531\epsilon$, $P_0 = 0.368$, $P_e = 0.368$)

Illustrating the redundancy-coded $\epsilon$-channel. A table of the quantities given by (27), (28), and (29), corresponding to a redundancy $r = n/\epsilon$.

If $C(n)$ is cascaded $N$ times, we raise the matrix $C(n)$ to the $N$th power:

$$C^N(n) = \left\| \begin{matrix} (1 - e^{-n})^N & 1 - (1 - e^{-n})^N \\ 0 & 1 \end{matrix} \right\|.$$

The capacity of the product matrix is still normalized by dividing it by the per-stage delay $r = n/\epsilon$, but the over-all delay that must now be tolerated is $Nr$. Assume that $n$ is large enough so that $e^{-n}$ is small and $c[C(n)] \sim 1.0$, and suppose that we agree to let $c[C^N(n)]$ fall off only to the value $k_1 < 1$. This amount of deterioration occurs when

$$(1 - e^{-n})^N = f(k_1), \tag{30}$$

where $f(k_1)$ is the abscissa of the point of intersection of the equicapacity line $k_1$ with the $\alpha$-axis (see Fig. 2). If we assume that $N$, the number of cascaded stages, is large (30) becomes

$$\exp(-Ne^{-n}) = f(k_1)$$

or

$$Ne^{-n} = -\log f(k_1) \equiv k_2 > 0.$$

Thus

$$N = k_2 e^n, \tag{31}$$

*i.e.*, we can tolerate more channels in the cascade if we increase $n$, and consequently the error-proofing and delay per stage. Moreover, since

$$P_e(n) \sim \tfrac{1}{2} e^{-n},$$

(31) can be rewritten as

$$N \sim \frac{k_2}{2P_e(n)}. \tag{32}$$

Eq. (32) says that the amount of cascading permissible to within a given tolerated deterioration of the end-to-end capacity is inversely proportional to the probability of error per stage.[12]

If a noiseless feedback channel is available at the receiver, the rate can be increased by the simple expedient of having the receiver instruct the transmitter to begin a new run of $r$ repetitions whenever a zero is received. For, since received zeros can only originate from transmitted zeros, it is a waste of channel space to continue repeating zeros for the rest of the run of $r$ digits when a zero has already been received. In the limit of high redundancy this feedback procedure increases the rate by a factor approaching 2, because without feedback almost half the channel space is taken up by the needless repetition of zeros.

## Acknowledgment

This work was done while the author held the position of Research Associate in the Department of Electrical Engineering of the Massachusetts Institute of Technology, and was supported by Industrial Fellowship funds of the Research Laboratory of Electronics.

The numerical calculations on which the figures and table are based were done by Mrs. Elgie G. Levin.

[12] If $k_2$ is very small, self-consistency may require $n$ to be quite large, since $N$ is assumed to be large (see (31)).

*Reprinted from the* IRE TRANSACTIONS OF THE PROFESSIONAL GROUP ON *INFORMATION THEORY*
Volume IT-1, Number 3, December, 1955
PRINTED IN THE U.S.A.