XVIII. PROCESSING AND TRANSMISSION OF INFORMATION

Academic and Research Staff

Prof. P. Elias
Prof. R. M. Gallager
Prof. M. M. Goutmann
Prof. E. V. Hoversten
Prof. D. A. Huffman
Prof. R. S. Kennedy
Prof. J. L. Massey
Prof. E. Mortenson
Prof. C. E. Shannon
Prof. R. N. Spann
Prof. J. T. Wagner

Graduate Students

D. S. Arnstein
E. A. Bucher
D. Chase
R. L. Greenspan
H. M. Heggestad
J. A. Heller
M. Khanna
Jane W-S. Liu
J. Max
J. C. Moldon
J. T. Pinkston III
E. M. Portner, Jr.
J. S. Richters
A. H. M. Ross
S. Thongthammachat
D. A. Wright

A. SHIFT-REGISTER SYNTHESIS AND APPLICATIONS

The general form of a linear feedback shift-register (FSR) is shown in Fig. XVIII-1. The register is completely described by its length and its connection polynomial

$$C(D) = 1 + c_1 D + \ldots + c_t D^t.$$

Given a finite sequence $s_1, \ldots, s_N$ of digits from some number field (for example, the field of binary numbers), the problem is posed of finding (one of) the shortest

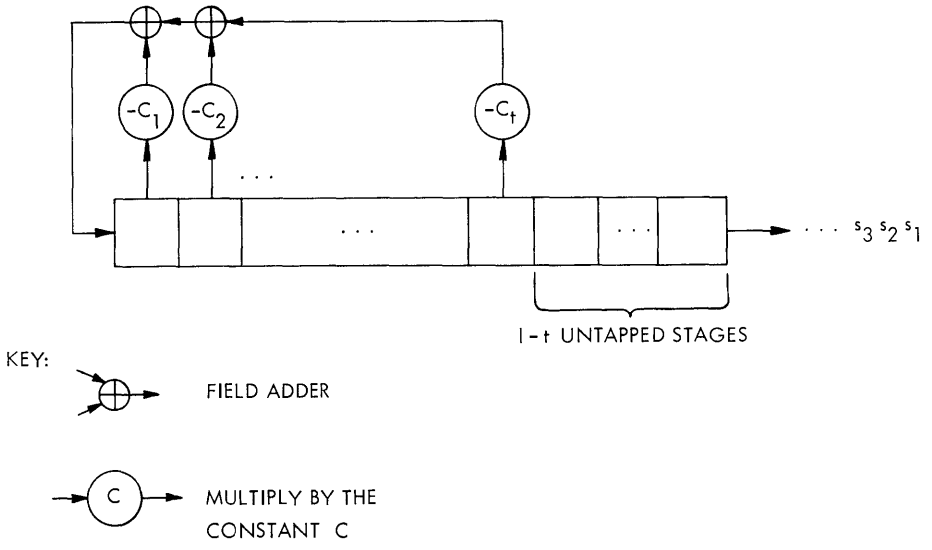

Fig. XVIII-1. General linear feedback shift register.

*This work was supported principally by the National Aeronautics and Space Administration (Grant NsG-334); and in part by the Joint Services Electronics Programs (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract DA 28-043-AMC-02536(E).

linear FSR that could generate this sequence when loaded initially with $s_1, s_2, \ldots s_\ell$. The following algorithm, which solves this problem by a recursive technique, has been obtained. Defining

$$d_n = s_{n+1} + \sum_{i=1}^{\ell_n} c_i^{(n)} s_{n+1-i},$$

where

$$C^{(n)}(D) = 1 + c_1^{(n)}(D) + \ldots + c_{\ell_n}^{(n)} D^{\ell_n},$$

gives a linear FSR of length ℓ_n, the shortest possible one for a linear FSR that generates $s_1, s_2, \ldots s_n$. Initializing with $n' = -1$, $n = 0$, $\ell_{n'} = \ell_n = 0$, $d_{n'} = 1$, $C^{(n')}(D) = 1$, $C^{(n)}(D) = 1$, we compute the registers for $n = 1, 2, \ldots, N$ by the following recursion:

1. If $d_n = 0$, set $C^{(n+1)}(D) = C^{(n)}(D)$, $\ell_{n+1} = \ell_n$, and leave all other quantities unchanged.

2. If $d_n \neq 0$, set

$$C^{(n+1)}(D) = C^{(n)}(D) - d_n d_{n'}^{-1} D^{n-n'} C^{(n')}(D)$$

and

$$\ell_{n+1} = \max [\ell_n, n-n'+\ell_{n'}].$$

If $n - \ell_n \leq n' - \ell_{n'}$, leave all other quantities unchanged. But if $n - \ell_n > n' - \ell_{n'}$ replace n', $\ell_{n'}$, $d_{n'}$ and $C^{(n')}(D)$ with n, ℓ_n, d_n, and $C^{(n)}(D)$, respectively.

Among the applications for this algorithm are (a) solving Newton's identities, which is the fundamental problem in decoding the Bose-Chaudhuri-Hocquenghem codes, (b) finding simple digital devices to produce a specified binary sequence, and (c) compressing the output of certain data sources with memory.

J. L. Massey

References

1. J. L. Massey, "Shift-Register Synthesis and BCH Decoding" (submitted to IEEE Transactions on Information Theory).

B. CODING THEOREMS FOR SOURCE-CHANNEL PAIRS

In a recently completed thesis,[1] we have studied the communication system shown in Fig. XVIII-2 when the capacity of the communication channel is not sufficiently high to allow perfect transmission of the source. The resulting (nonzero) distortion is measured by a non-negative distortion function, d(w, z), which gives the distortion in the event that the source letter w has occurred at the source output but been reproduced at the decoder output as the letter z. It is assumed that both the source and channel are discrete, constant, and memoryless, and that the channel is available for use at a rate of once per source output. It is also assumed that the encoder and decoder are allowed to operate on blocks of letters; the encoder maps n-letter source output words into n-letter channel

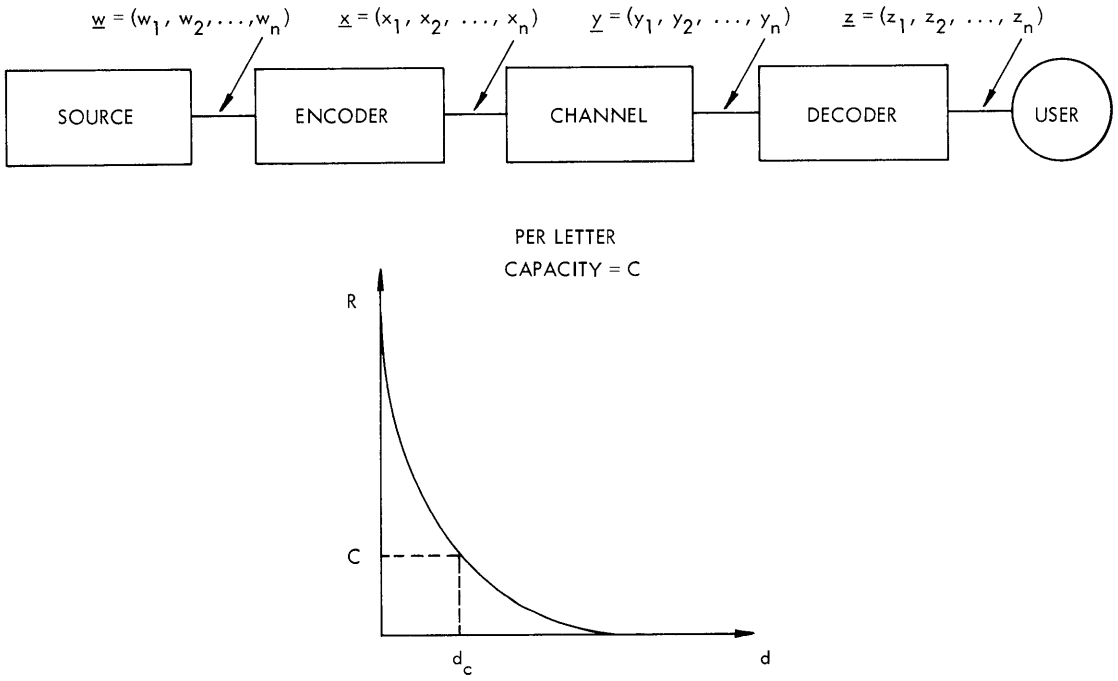

Fig. XVIII-2. Rate-distortion curve for the source.

input words, and the decoder maps n-letter channel output words into n-letter decoder output words. When block operators of this type are used, and one "transmission" contains n information letters from the source, the system performance is measured by the normalized sum of the n letter distortions, or

$$d(\underline{w}, \underline{z}) = \frac{1}{n} \sum_{i=1}^{n} d(w_i, z_i).$$

For such transmission systems, Shannon has introduced a rate-distortion function[2] that specifies the minimum attainable transmission distortion, d_c, in terms of the channel capacity, C. In general, though, the distortion level d_c is attainable only in the limit as the encoder and decoder are allowed to be arbitrarily complex, that is, the block lengths on which they operate are arbitrarily long. In this work, the block length was included as a variable, and upper and lower bounds were found to the minimum attainable transmission distortion as a function of this block length. Particular emphasis was placed on finding the asymptotic form of these bounds.

Even before these bounds are found, several interesting situations are known to exist. For instance, there are some source-channel pairs for which the minimum attainable transmission distortion is independent of the encoding block length; therefore, it is possible to attain the distortion level d_c even with n = 1. An example of such a pair is the binary symmetric source (equally-likely binary letters with $d(i,j) = 1 - \delta_{ij}$, i, j = 1, 2) used with a binary symmetric channel, where the optimum encoder is a direct connection. Another example is a Gaussian source used with an additive Gaussian noise channel, where the optimum encoder is simply an amplifier. When the source-channel pair is such that the minimum attainable distortion is independent of the coding block length, we shall say that the source and channel are "matched." For the more common situation, wherein the minimum attainable transmission distortion decreases with increasing encoding block length to asymptotically approach the distortion level d_c, we say that there is a "mismatch" between the source and channel, and suggest as a measure of this mismatch the "slowness" of the approach of the distortion to the asymptote d_c. Examples illustrating mismatches between source and channel are given in the author's thesis.[1]

Another interesting situation occurs when there is a choice of using one of several channels of different capacity. Although the channel of highest capacity would be the best choice when one is willing to use infinite block-length coding, it might not be the best choice with finite-length coding. This could easily happen if the high capacity channel were very much more mismatched to the source than some lower-capacity channel.

1. Lower Bound

A generalization of the sphere-packing concept is used to derive the lower bound. The idea involved can be described with the following simple, but poor, bound. It is first assumed that the source word $\underline{w}$ has occurred at the source output and that the channel input word $\underline{x}$ is used for transmission. We list all possible received words, $\underline{y}$, ordered in decreasing conditional probability, $p(\underline{y}|\underline{x})$, and pair with each the decoder output word, $\underline{z}(\underline{y})$, to which it is decoded. The transmission distortion,

$$d(\underline{w}) = \sum_{\underline{y}^n} p(\underline{y}|\underline{x})\, d(\underline{w}, \underline{z}(\underline{y})), \tag{1}$$

can be seen to equal the sum of conditional probability-distortion products on this list. If the set of distortion values that appear on this list are now rearranged (with the list of conditional probabilities fixed) to be ordered in increasing distortion values, the resulting sum of conditional probability-distortion products must be smaller, or at most equal to, the sum in Eq. 1. It therefore provides a lower bound.

An improved lower bound employs the same sort of orderings and rearrangements but includes a probability function, $f(\underline{y})$, in the ordering of the channel output words. This function is defined over the set of all channel output words, denoted by $\underline{y}^n$, and is later chosen to optimize the result. The channel output words are now ordered according to increasing values of the information difference $I(\underline{x}, \underline{y}) = \ln f(\underline{y})/p(\underline{y}|\underline{x})$, and each is again paired with the decoder output word $\underline{z}(\underline{y})$ to which it is decoded. The rearrangement of decoder output words is also slightly different. To describe this rearrangement, we visualize each channel output word, $\underline{y}$, as "occupying" an interval of width $f(\underline{y})$ along the line $[0, 1]$. The decoder output word, $\underline{z}(\underline{y})$, that is paired with a particular channel output word, $\underline{y}$, is also viewed as occupying the same region along $[0, 1]$ as $\underline{y}$, but, since any particular word $\underline{z}_0$ might be the decoding result of several channel output words, the region along $[0, 1]$ occupied by $\underline{z}_0$ could be a set of separated intervals. The arrangement of decoder output words is, this time, a rearrangement of occupancies in $[0, 1]$ toward the desired configuration, wherein the decoder words are ordered in increasing distortion (along this line), and each occupies the same total width in $[0, 1]$ as it did before the ordering. Thus two monotone nondecreasing functions can be defined along the line $[0, 1]$; one, $I(h)$, giving the information difference $I(\underline{x}, \underline{y})$ at the point h, $0 \leq h \leq 1$, and the other, $d(h)$, giving the distortion $d(\underline{w}, \underline{z})$ at h. The distortion $d(\underline{w})$ in Eq. 1, can be lower-bounded in terms of these functions by

$$d(\underline{w}) \geq \int_0^1 d(h)\, e^{-nI(h)}\, dh. \tag{2}$$

The lower bound to the total average transmission distortion is then the average of this bound over all possible source events.

If the probability function $f(\underline{y})$ and the probability function, $g(\underline{z})$, induced on $\underline{z}^n$ by $f(\underline{y})$ through the optimum decoder function, are used to define the quantities $I(\underline{x}, \underline{y})$ and $d(\underline{w}, \underline{z})$ as random variables, the functions $I(h)$ and $d(h)$ can be seen to be the "inverses" of their cumulative distribution functions. By using estimates to these distribution functions,[3,5] the lower bound in Eq. 2 can be simplified considerably. When the (unknown) $\underline{g}(z)$ is approximated by a probability function factorable into blocks of arbitrary but constant size and then varied to minimize the bound, and $f(\underline{y})$ is also varied to optimize the bound,

it can be shown that the asymptotic form of this approximated lower bound to distortion is

$$d(S) \geq d_c + \frac{a}{n} + 0\left(\frac{1}{n}\right), \tag{3}$$

where

$$a = \frac{1}{2|s|}\left\{\frac{\gamma''}{s^2\mu''} - 1 - \ln\frac{\gamma''}{s^2\mu''} + \frac{\sigma^2}{s^2\mu''}\right\}$$

d_c = distortion at R = C on the rate-distortion curve for the source

C = capacity of the channel

$$\mu(s) = \sum_i q_i \ln \sum_j g_j e^{sd_{ij}} \equiv \sum_i q_i \mu_i(s)$$

$$\gamma(t) = \sum_k C_k \ln \sum_\ell f_\ell^{1+t} p_{kc}^{-t}$$

$\underline{q} = \underline{p}$

$\underline{p}$ = source output probabilities

$\underline{g}$ = output probability on the test channel for the source at the point (d_c, C) on the rate-distortion curve

$\underline{c}, \underline{f}$ = input and output probabilities on the channel when it is used to capacity

σ^2 = variance of $\mu_i(s) - s\mu_i'(s)$ according to $\underline{p}$

t = -1

s satisfies: $\mu(s) - s\mu'(s) = -C$.

The coefficient a can be shown to be a non-negative function of the source and channel statistics that interrelates these statistics in such a way that the particular channel (among those of capacity C) for which a has its minimum value depends upon the source that is used. The reverse is also true. Among those sources that have a common point (d_c, C) on their rate-distortion curves, the particular source that minimizes a is different for different channels. Also, the coefficient a is precisely zero when the source and channel are matched. These properties of a suggest its utility as a measure of "mismatch" between the source and channel; the larger the mismatch, the slower is the approach of the lower bound to its asymptote. Several examples of different types of mismatch have been provided, and a strict lower bound, including the specification of the low-order terms, are to be found in the author's thesis.[1]

2. Upper Bound

A random-coding argument is used to derive the upper bound. That is, an ensemble of encoders and decoders are defined over which the ensemble average transmission distortion is calculated. This, then, upper-bounds the minimum individual average (over source and noise events) transmission distortion in the ensemble and, in turn, upper-bounds the minimum average transmission distortion attainable with any encoding and decoding method.

First, two distortion values, d_R and d^*, are chosen to satisfy

$$d_c < d_R < d^* \leq d_{max}. \tag{4}$$

Since a valid upper bound results for any two such choices, d_R and d^* are considered as parameters to be optimized later. The monotonicity of the rate-distortion curve and the inequalities in Eq. 4 provide the following inequalities among the corresponding values of rate on this curve:

$$C > R > R^*. \tag{5}$$

For each choice of d_R and d^*, and each coding block length n, the ensemble of codes is generated by picking, according to some probability distribution $p(\underline{x}, \underline{z})$, M independent pairs $(\underline{x}, \underline{z})$ from $X^n Z^n$. Thus, if in X there are J channel input letters and in Z there are K decoder output letters, there is a total of $(JK)^{nM}$ codes, each with the associated probability

$$\Pr(\text{code}) = \prod_{i=1}^{M} p(\underline{x}_i, \underline{z}_i).$$

The particular distribution that was used factors as $p(\underline{x}, \underline{z}) = p(\underline{x})\, g(\underline{z})$, in which $p(\underline{x})$ is the channel input probability distribution that uses the channel to capacity, and $g(\underline{z})$ is the output probability distribution on the test channel for the source at the point (d^*, R^*) on its rate-distortion curve.

The encoding and decoding is done in the following way. When a source output $\underline{w}$ occurs, the encoder chooses any member in its set of M permissible decoder words, say $\underline{z}_o$, which satisfies

$$d(\underline{w}, \underline{z}_o) \leq d^*. \tag{6}$$

If there is no such member, it chooses any word in the set, say $\underline{z}_1$. Because in each ensemble member there is a particular pairing defined between the M decoder output words and the M channel input words, there corresponds to $\underline{z}_o$, or $\underline{z}_1$, a particular channel input word, $\underline{x}_o$, which is used for transmission. From the received channel output word, $\underline{y}$, the decoder first decodes to one of the M possible channel input words, and

from this, through the pairings defined by the code, to a decoder output word.

Clearly, if no channel error occurs and if the set of decoder words does contain a member satisfying Eq. 6, the transmission distortion must be upper-bounded by d^*. In any other event, the distortion can be upper-bounded by d_{max}. By using the union bound, the total average (the average over source events, noise events, and the ensemble) can therefore be upper-bounded by

$$\bar{d}(\text{ens.}) \leq d^* + \left(d_{max}-d^*\right)[\Pr(\exists' \underline{z}_o \text{ in code}) + \Pr(\text{channel error})], \tag{7}$$

in which the symbol $\exists'$ is used for "there does not exist."

The first probability in Eq. 7 was calculated by conditioning events on the occurrence of $\underline{w}$, finding the probability of codes lacking a decoder word satisfying Eq. 6, and then averaging over the source space W^n. The result is an exponentially decreasing function of n, with the exponent starting from zero and increasing monotonically in the difference $R-R^*$. This is analogous to the second probability, which is known also to be an exponentially decreasing function of n, but has an exponent starting from zero and increasing monotonically in the difference C-R. Thus the upper bound in Eq. 7 converges exponentially to the level d^* which, from Eq. 4, is strictly greater than d_c. This bound alone would not be satisfactory, since Shannon has shown that the level d_c can be approached.

As the bound in Eq. 7 is valid for each d^* and R satisfying Eqs. 4 and 5, the lower envelope to the set of bounds corresponding to all such choices of d^* and R is also a valid upper bound. It can be seen that the optimum choice of d^* (corresponding to that bound to which the lower envelope is tangent) must decrease toward d_c with increasing n and, from Eqs. 4 and 5, that R^* (and R) must increase toward C. The result of this is that the exponents in the probabilities of Eq. 7 must decrease toward zero, with the further consequence that the exponential terms in this equation decay more slowly as n increases. (For this reason, a choice of d^* marginally above d_c is not optimum for all block lengths.) The asymptotic form of the lower envelope, which is our upper bound to the average transmission distortion, is found to be

$$d(S) \leq d_c + b\sqrt{\frac{\ln n}{n}}\,[1+0(1)], \tag{8}$$

in which

$$f(n) \triangleq 0(1) \quad \text{if} \lim_{n\to\infty} f(n) = 0$$

$$b = \frac{1}{|s|}\left[E_s''(R^*)^{-1/2} + E''(C)^{-1/2}\right].$$

In Eq. 8, E(R) is the reliability function for the channel, and $E_s(R)$ is given by

$$E_s(R) = \min_i \ln \left[\left(\frac{p_i}{p_i \pm \Delta} \right)^{p_i \pm \Delta} \left(\frac{1 - p_i}{1 - p_i \mp \Delta} \right)^{1 - p_i \mp \Delta} \right],$$

where Δ is the largest number for which

$$\mu(s) - s\mu'(s) \geq -R$$

$$\mu'(s) = d^*$$

$$q_i = p_i \pm \Delta$$

$$\sum_i q_i = 1$$

are all satisfied. Another form of the function $E_s(R)$, which is more difficult to work with but provides a tighter bound, has been found.[1]

In this derivation, we were forced to use a coding ensemble in which the signal set in each ensemble member is limited to $M < e^{nC}$ points, since no more general code could be found that provided the correct asymptote, d_c. The restriction to such a signal set, in effect, introduces an interface between the source and channel. This causes the coefficient b not to reveal the mismatch properties that the coefficient a brings about in the lower bound, since the set of source and channel statistics that minimize b are each independent of the other. We can, though, interpret b as (the reciprocal of) a type of stretch factor similar to those studied by Shannon[6] and by Wozencraft and Jacobs.[7]

With the restriction to a signal set with $M < e^{nC}$, we have also found a lower bound to distortion that (for noisy channels) has the asymptotic form

$$d(S) \geq d_c + a_1 n^{-1/2}.$$

Thus one can conclude that it is necessary to have a signaling set larger than e^{nC} if one is to attain the 1/n rate of approach to d_c that appears in the lower bound in Eq. 3. Although we cannot exhibit such a coding scheme, the author conjectures that one does exist, and that the lower bound in Eq. 3 more correctly expresses the behavior of the performance curve.

For the special case of a noiseless channel, upper and lower bounds to the average transmission distortion have been found which, asymptotically, behave the same. Their form is

$$d_c + \frac{1}{2} \frac{\ln n}{|s| n} [1+0(1)] \leq d(s) \leq d_c + \left(\frac{1}{2} + \epsilon \right) \frac{\ln n}{|s| n} [1+0(1)],$$

in which s is equal to the slope of the rate-distortion curve at (d_c, C), and ϵ is an arbitrarily small positive constant. The lower bound is similar to one derived by Goblick[8] (the bound in Eq. 3 is not applicable as $a = \infty$). The upper bound is derived by using essentially the same procedure as that used to obtain the noisy-channel upper bound. The significant difference is the replacement of the threshold encoder (Eq. 6) with an optimum encoder, that is, choosing for $\underline{z}_0$ that permissible decoder output word which minimizes $d(\underline{w}, \underline{z})$.

R. J. Pilc

References

1. R. J. Pilc, "Coding Theorems for Discrete Source-Channel Pairs," Ph. D. Thesis, Department of Electrical Engineering, M. I. T., 1966.
2. C. E. Shannon, "Coding Theorems for a Discrete Source with a Fidelity Criterion," IRE National Convention Record, Part 4, 1959, p. 142.
3. C. E. Shannon, "Notes for Seminar in Information Theory at M. I. T.," 1956 (unpublished).
4. R. M. Fano, The Transmission of Information (The M. I. T. Press, Cambridge, Mass., 1961).
5. R. G. Gallager, "Lower Bounds on the Tails of Probability Distributions," Quarterly Progress Report No. 77, Research Laboratory of Electronics, M. I. T., April 15, 1965, p. 277.
6. C. E. Shannon, "Communication in the Presence of Noise," Proc. IRE 37, 10 (1949).
7. J. M. Wozencraft and I. M. Jacobs, Principles of Communication Engineering (John Wiley and Sons, Inc., New York, 1965).
8. T. J. Goblick, "Coding for a Discrete Information Source with a Distortion Measure," Ph. D. Thesis, Department of Electrical Engineering, M. I. T., 1962.