XII. PROCESSING AND TRANSMISSION OF INFORMATION

Prof. E. Arthurs
Prof. P. Elias
Prof. R. M. Fano
Prof. J. Granlund
Prof. D. A. Huffman
Prof. H. Rogers, Jr.
Prof. C. E. Shannon (absent)
Prof. W. M. Siebert
Prof. J. M. Wozencraft
R. K. Brayton
D. C. Coll
J. E. Cunningham
M. A. Epstein
E. F. Ferretti
R. G. Gallager
T. L. Grettenberg
P. R. Hall
F. C. Hennie III
E. M. Hofstetter
Cynthia H. Hsiao
T. Kailath
W. G. Kellner
D. C. Luckham
R. S. Marcus
J. Max
G. E. Mongold, Jr.
G. E. W. Sparrendahl (visitor)
W. A. Youngblood
H. L. Yudkin

RESEARCH OBJECTIVES

The interest of this group ranges over a wide variety of problems in the processing and transmission of information.

One current activity is the statistical investigation of information sources. The main objective is to estimate the rate at which the sources generate information and to determine how to encode their output economically in order to decrease the channel capacity required for transmission. The group is currently continuing such an investigation on pictures as information sources by recording the pictures in digital form for analysis and processing on a digital computer.

A second current activity is the investigation of channels for information transmission. This includes study of the fundamental theoretical relationships between transmission rate, error probability, delay, and equipment complexity, and the construction of suitable coding procedures, especially for binary channels. Efforts are being made to extend some of the results on binary channels to continuous channels, particularly with respect to communication over a time-varying multipath medium.

An interesting problem in the processing of information is the selection of subsets, called "bibliographies," from a set of stored documents. A bibliography may be heuristically defined as a set of documents pertinent to a particular topic. The model that is being studied consists of assigning a multivariate probability distribution over the set of documents which can be used to measure the coherence of a subset. The probability distribution will be based on such data as the relative frequency of simultaneous occurrence of a reference of a subset in reference lists and the relative frequency of concomitant requests of documents by users of a library. The subsets that yield relative maxima of a coherence function are the bibliographies. The problems of operationally defining the probability distribution and choosing suitable coherence functions in such a way that the bibliographies can be economically generated by computing systems are being investigated.

In the design of coding and decoding devices, finite-state binary logical circuits play a critical role. A binary circuit can be considered as a transducer that transforms an input stream of binary digits into a related output stream. In the process, information may be lost and the coded form of the stream changed drastically. Thus a finite-state circuit is a rather general information channel. Study in this area is aimed at understanding how the loss of information in a finite-state circuit depends upon the logical structure of the circuit and upon the solvability of the equations that show how the output is related to the input.

From an alternative point of view, any finite-state circuit classifies each of the infinitely many possible input sequences of digits into one of the finite number of states. Studies made of minimal descriptions of finite-state circuits and of approximate models thereof are important, in that they show how to describe efficiently the patterns of digits in a sequence. More closely applicable to the pattern-recognition problem is a study of logical circuits whose description is most conveniently made in more than one dimension. For example, the study of circuits whose logical elements can be arranged in a uniform planar array like that formed by the squares on an indefinitely large chessboard is particularly pertinent to possible schemes for filtering and logical processing of an ordinary black and white photograph.

A model for an infinite-state logical circuit is the Turing machine. A study is being made of the dependence of the length of a computation on the capacity of the channel connecting the memory to the control of a class of Turing systems.

P. Elias, R. M. Fano, D. A. Huffman

A. SOLVABILITY CRITERION FOR SIMULTANEOUS LOGICAL EQUATIONS

If we have been given n simultaneous equations that are functions of n binary variables, the question of whether or not a set of solutions exists corresponds directly to the question of whether or not the logical network which produces the functions (as outputs) from the variables (as inputs) is information-lossless. That is (in either case), if a set of values of the functions is given, can the corresponding set of variable values be uniquely determined?

Any binary function can be expressed as the modulo-two sum of terms involving a constant, the single variables, their products taken two at a time, their products taken three at a time, and so forth. For example, any three-variable combinational function can be expressed as

$$f = C_o + C_x X + C_y Y + C_z Z + C_{xy} XY + C_{xz} XZ + C_{yz} YZ + C_{xyz} XYZ$$

where the C's are eight binary coefficients whose values determine which of the $2^{2^3} = 2^8 = 256$ functions of three variables is being described.

If we adopt the notation $\partial f/\partial x$ for $f_{x=0} + f_{x=1}$, we find that

$$\frac{\partial f}{\partial x} = C_x + C_{xy} Y + C_{xz} Z + C_{xyz} YZ$$

That is, as far as the symbology is concerned, we can compute $\partial f/\partial x$ as if f were a function of continuous variables. In the same way, we obtain

$$\frac{\partial^2 f}{\partial x \partial y} = C_{xy} + ZC_{xyz}$$

$$\frac{\partial^3 f}{\partial x \partial y \partial z} = C_{xyz}$$

We adopt the notation

$$\left.\begin{aligned} a_x &= \frac{\partial f}{\partial x} \\ a_{xy} &= \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} + \frac{\partial^2 f}{\partial x \partial y} \\ a_{xyz} &= \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z} + \frac{\partial^2 f}{\partial x \partial y} + \frac{\partial^2 f}{\partial x \partial z} + \frac{\partial^2 f}{\partial y \partial z} + \frac{\partial^3 f}{\partial x \partial y \partial z} \end{aligned}\right\} \quad (1)$$

and similar meanings for a_y, a_z, a_{xz}, and a_{yz}. Now consider three functions a, b, and c of the variables x, y, and z. Let

$$\left.\begin{aligned} S_x &= a_x + b_x + c_x + a_xb_x + a_xc_x + b_xc_x + a_xb_xc_x \\ S_{xy} &= a_{xy} + b_{xy} + c_{xy} + a_{xy}b_{xy} + a_{xy}c_{xy} + b_{xy}c_{xy} + a_{xy}b_{xy}c_{xy} \\ S_{xyz} &= a_{xyz} + b_{xyz} + c_{xyz} + a_{xyz}b_{xyz} + a_{xyz}c_{xyz} + b_{xyz}c_{xyz} + a_{xyz}b_{xyz}c_{xyz}, \text{ etc.} \end{aligned}\right\} \tag{2}$$

The equations that relate a, b, and c to x, y, and z can be solved if and only if

$$S_xS_yS_zS_{xy}S_{xz}S_{yz}S_{xyz} = 1 \tag{3}$$

The conditions in Eqs. 1, 2, and 3 (which involve the partial "derivatives" of each of the functions with respect to the variables) serve the same function for n simultaneous logical equations as does the conventional Jacobian test for equations in continuous variables.

The correctness of these manipulations should not mislead the reader into believing that, even notationally speaking, all operations involving partial "derivatives" are performed on modulo-two equations in the same way that they are performed on equations with continuous variables. For example, in the binary case we have

$$\frac{\partial(a \cdot b)}{\partial x} = a\frac{\partial b}{\partial x} + b\frac{\partial a}{\partial x} + \frac{\partial a}{\partial x} \cdot \frac{\partial b}{\partial x}$$

D. A. Huffman

B. ZERO ERROR CAPACITY FOR LIST DETECTION

Shannon (1) has defined C_o, the zero error capacity of a channel, as the least upper bound to rates of transmission that are attainable with blocks of length N and zero error probability. He does not evaluate C_o in general. He also defined C_{oF}, the zero error capacity for block coding when a noiseless feedback channel is available, and evaluates it explicitly as follows.

a. $C_{oF} = 0$ if, given any pair of input letters, there is at least one output letter that both can cause (i.e., to which both are connected on the graph of the channel).

b. Otherwise, $C_{oF} = -\log P_o$, where

$$P_o = \min_{P_i} \max_{j} \sum_{S_j} P_i$$

Here S_j is the set of input letters connected to output letter j, $1 \leqslant j \leqslant s$, and P_i is an assignment of probabilities to the input letters i, $1 \leqslant i \leqslant t$. Thus P_o is the largest amount of input probability to which any output letter is connected, when the P_i are selected so as to minimize this largest amount.

The purpose of this note is to show that if list detection (2, 3) is used, then C_o for block coding without feedback is also $-\log P_o$ as in expression b, but without restriction a. We define $C_{oL}(k)$ for a block code as the least upper bound of rates for codes in which the transmitter selects one of kM input words, each of length N input letters, and transmits it, and the receiver produces a list of k input words which is guaranteed to include the one which was in fact sent. The rate for such a scheme is then

$$R \geqslant \frac{\log kM - \log k}{N} = \frac{\log M}{N}$$

Next, we define C_{oL} and obtain the following theorem.

THEOREM: $C_{oL} = \lim\limits_{k\to\infty} C_{oL}(k) = -\log P_o$.

PROOF: Shannon (1) has an argument that shows that $C_{oF} \leqslant -\log P_o$, which still applies in this case. Thus we need only show that there are block codes for which the rate with list decoding and zero error probability approaches $-\log P_o$ for sufficiently large k and N. We use a random coding argument.

Consider the ensemble of codes constructed by selecting kM sequences of N input letters each, using the probabilities P_i of expression b, and selecting each letter of each word with statistical independence from this distribution (or from one of the minimizing distributions if there is more than one). We shall bound the probability that a set of $k+1$ of these words has first letters, to all of which some received letter is connected.

First, consider any particular received letter, say $j = 1$. This letter will distinguish between the given $k+1$ words unless all of their first letters fall in the set $S_j = S_1$, which is an event of probability in the ensemble $\leqslant P_o^{k+1}$, by the definition of P_o in expression a. This result obviously holds for any j: each particular received symbol can cause confusion only in a fraction $\leqslant P_o^{k+1}$ of the cases, and one or more of the s possible received symbols will cause confusion only in $\leqslant sP_o^{k+1}$ of the cases.

Now, for $k+1$ input words of length N to have a received word in common requires that all N sets of $k+1$ corresponding letters have a received letter in common. In the ensemble the probability of this event is $\leqslant \left(sP_o^{k+1}\right)^N$. And there are only

$$\binom{kM}{k+1} \leqslant (k+1)^{k+1}M^{k+1}/(k+1)!$$

different sets of $k+1$ words to consider. The probability P_c that one or more of these sets has a code word in common is then bounded.

$$P_c \leqslant \frac{(k+1)^{k+1} M^{k+1} \left(s P_o^{k+1}\right)^N}{(k+1)!} \tag{1}$$

Now, if $P_c < 1$, there is a positive probability of picking a code with the property that no set of $k + 1$ of its input words has an output word in common. Then there must be at least one such code, and it is obviously a zero-error code for list decoding at list size k: Given any received word, k at most, of the possible transmitted words are consistent with it. To make $P_c < 1$, or $\log P_c < 0$, take logarithms in Eq. 1 above and divide by $(k+1)N$. This gives $P_c < 1$ if

$$\frac{\log M}{N} < -\log P_o - \frac{1}{k+1} \log s - \frac{1}{N}\left[\log(k+1) - \frac{\log(k+1)!}{k+1}\right]$$

Q. E. D.

P. Elias

References

1. C. E. Shannon, The zero error capacity of a noisy channel, Trans. IRE, vol. IT-2, no. 3, pp. 8-19 (Sept. 1956).
2. P. Elias, List decoding for noisy channels, Technical Report 335, Research Laboratory of Electronics, M. I. T. (in press).
3. J. M. Wozencraft, Sequential decoding for reliable communication, Technical Report 325, Research Laboratory of Electronics, M. I. T., Aug. 9, 1957.

C. LIST DECODING

The computation of bounds on the probability of error behavior of codes designed for communication over binary symmetric channels has been carried out in great detail by Elias (1, 2, 3). This report covers an independent derivation, in weaker but less complicated form, of one of Elias' results (3).

The specific problem under consideration is that of obtaining an upper bound to the ensemble average probability $P_m(e)$ that a message actually transmitted will not be included when the receiver prepares a list of the m most-probable messages. It is assumed that block coding and maximum-likelihood decoding procedures are used. The average value $P_m(e)$ is computed over the ensemble of all possible block codes n digits long; we are then assured that at least one code exists for which the actual probability of error is as small as this.

In a typical experiment, let M_k be the number of possible transmitter messages that differ from the received message in k or fewer digits. Then, averaging first over the ensemble of codes, and next over the ensemble of binary-symmetric-channel transmission-error patterns, we have

$$P_m(e) = \sum_{k=0}^{n} P[d_o = k]\, P[M_k \geqslant m] \tag{1}$$

where d_o is the actual number of errors introduced by the binary symmetric channel.

First, we bound $P[M_k \geqslant m]$. If the number of possible transmitter messages is S, then over the ensemble of all such message sets we have

$$P[M_k \geqslant m] < P_k^m \binom{S-1}{m} \tag{2}$$

where P_k is the probability that any particular incorrect message differs from the received message in k or fewer digits. As Elias points out (3), this bound follows from the facts that retention by the receiver of at least m incorrect messages certainly implies the retention of exactly m, and that the right-hand side of Eq. 2 represents the sum of all possible disjoint ways in which the latter event can occur. Expanding the binomial coefficient in Eq. 2, bounding, and applying Stirling's approximation, we obtain the successive inequalities

$$P[M_k \geqslant m] < \frac{(SP_k)^m}{m!} < \frac{1}{(2\pi m)^{1/2}} \left(\frac{eSP_k}{m}\right)^m \tag{3}$$

Evaluation of the right-hand side of Eq. 3 is straightforward. As Shannon first showed (4), application of the Chernov bound, specialized to binary variables (5), leads to the result that

$$P_k \leqslant 2^{-n\left[1-H\left(\frac{k}{n}\right)\right]} \tag{4}$$

where H is the usual entropy function.

$$H(x) = -x \log_2 x - (1-x) \log_2(1-x) \qquad (0 \leqslant x \leqslant 1) \tag{5}$$

For convenience, we now define an auxiliary parameter p_t in terms of the message set size S and word length n, by means of Eq. 6:

$$S = 2^{n[1-H(p_t)]} \qquad \left(p_t < \frac{1}{2}\right) \tag{6}$$

Finally, substituting Eqs. 4 and 6 in Eq. 3, we obtain the desired bound on $P[M_k \geqslant m]$.

$$\left.\begin{aligned} P[M_k \geqslant m] &< \frac{1}{(2\pi m)^{1/2}} \left(\frac{e}{m}\right)^m 2^{-nm\left[H(p_t)-H\left(\frac{k}{n}\right)\right]} && \left(k \leqslant np_t < \frac{n}{2}\right) \\ &\leqslant 1 && \text{(otherwise)} \end{aligned}\right\} \quad (7)$$

Next, we require a bound on $P[d_o = k]$. Let p_o be the binary-symmetric-channel transition probability. Then

$$P[d_o = k] = p_o^k\, q_o^{n-k} \binom{n}{k} \qquad (q_o = 1 - p_o) \qquad (8)$$

When we apply Stirling's approximation to Eq. 8, we obtain (5), for $p_o \neq 0, 1$,

$$\left.\begin{aligned} P[d_o = k] &\leqslant \frac{1}{\left(2\pi(n-k)\frac{k}{n}\right)^{1/2}}\, 2^{-n\left[H(p_o)-H\left(\frac{k}{n}\right)+\left(\frac{k}{n}-p_o\right)\log_2 \frac{q_o}{p_o}\right]} && (k > np_o) \\ &\leqslant 1 && \text{(otherwise)} \end{aligned}\right\} \quad (9)$$

At this time we also note for future reference that Chernov's bound also leads (5) to the similar result:

$$P[d_o > k] = \sum_{j=k+1}^{n} P[d_o = k] \leqslant 2^{-n\left[H(p_o)-H\left(\frac{k}{n}\right)+\left(\frac{k}{n}-p_o\right)\log_2 \frac{q_o}{p_o}\right]} \qquad (k > np_o) \quad (10)$$

The final step is to substitute Eqs. 7, 9, and 10 in Eq. 1, and thus break the summation into three ranges

$$P_m(e) = \sum_{k=0}^{n} P[d_o = k]\, P[M_k \geqslant m] \qquad (1)$$

$$P_m(e) = \sum_{k=0}^{np_o} P[M_k \geqslant m] + \sum_{k=np_o+1}^{np_t} P[d_o = k]\, P[M_k \geqslant m] + \sum_{k=np_t+1}^{n} P[d_o = k] \qquad (11)$$

The largest term in the first summation occurs for $k = np_o$, and a sum is bounded by the product of its largest term and the number of terms. Therefore,

$$\sum_{k=0}^{np_o} < \frac{np_o}{(2\pi m)^{1/2}} \left(\frac{e}{m}\right)^m 2^{-nm[H(p_t)-H(p_o)]} \qquad (12)$$

For the second summation, we have

$$\sum_{k=np_o+1}^{np_t} = \sum_{k=np_o+1}^{np_t} \frac{1}{\left(2\pi(n-k)\frac{k}{n}\right)^{1/2}} \frac{1}{(2\pi m)^{1/2}}\left(\frac{e}{m}\right)^m 2^{-n\left[H(p_o)-H\left(\frac{k}{n}\right)+\left(\frac{k}{n}-p_o\right)\log_2\frac{q_o}{p_o}\right]}$$

$$\cdot\, 2^{-nm\left[H(p_t)-H\left(\frac{k}{n}\right)\right]} \qquad (13)$$

Over the range of summation,

$$\frac{1}{\left[2\pi(n-k)\left(\frac{k}{n}\right)\right]^{1/2}} < \frac{1}{(2\pi\, np_oq_o)^{1/2}} \qquad (14)$$

and we have

$$\sum_{k=np_o+1}^{np_t} < \frac{\left(\frac{e}{m}\right)^m}{2\pi(nm\, p_oq_o)^{1/2}}\, 2^{-n\left[H(p_o)-p_o\log\frac{q_o}{p_o}-mH(p_t)\right]}$$

$$\cdot \sum_{k=np_o+1}^{np_t} 2^{+n(1+m)H\left(\frac{k}{n}\right)-k\log_2\frac{q_o}{p_o}} \qquad (15)$$

By differentiation, the exponent in Eq. 15 is found to have a maximum value at $k = np_{\text{crit}_m}$, where we define

$$\frac{q_{\text{crit}_m}}{p_{\text{crit}_m}} = \left(\frac{q_o}{p_o}\right)^{1/1+m} \qquad q_{\text{crit}_m} = 1 - p_{\text{crit}_m} \qquad (16)$$

The exponent is monotonically increasing for $k/n < p_{\text{crit}_m}$. Accordingly, if $p_t \leqslant p_{\text{crit}_m}$, the largest term is the last one, and

$$\sum_{k=np_o+1}^{np_t} < \frac{\left(\frac{e}{m}\right)^m n(p_t-p_o)}{2\pi(nm\, p_oq_o)^{1/2}}\, 2^{-n\left[H(p_o)-H(p_t)+(p_t-p_o)\log_2\frac{q_o}{p_o}\right]} \qquad \left(p_t \leqslant p_{\text{crit}_m}\right) \quad (17)$$

The third summation has already been bounded by Eq. 10. All that remains to be done is to note that the exponential factor in Eq. 17 must be larger than that in Eq. 12. This follows from the fact that the exponent of Eq. 13 reduces to the exponent of Eq. 12 for $k = np_o < np_t$. We therefore loosen the bound in Eq. 12, and write, instead,

$$\sum_{k=0}^{np_o} < \frac{np_o}{(2\pi m)^{1/2}} \left(\frac{e}{m}\right)^m 2^{-n\left[H(p_o)-H(p_t)+(p_t-p_o)\log_2\frac{q_o}{p_o}\right]} \tag{18}$$

Finally, substituting Eqs. 10, 17, and 18 in Eq. 11, and simplifying, we obtain the ultimate result

$$P_m(e) < \left[1 + \left(p_o + \frac{p_t - p_o}{(2\pi np_oq_o)^{1/2}}\right) \frac{n}{(2\pi m)^{1/2}} \left(\frac{e}{m}\right)^m\right] 2^{-n\left[H(p_o)-H(p_t)+(p_t-p_o)\log_2\frac{q_o}{p_o}\right]}$$

$$\left(\text{for } p_o < p_t \leq p_{crit_m} < \frac{1}{2}\right) \tag{19}$$

where

$$S = 2^{n[1-H(p_t)]} \tag{6}$$

and

$$\frac{q_{crit_m}}{p_{crit_m}} = \left(\frac{q_o}{p_o}\right)^{\frac{1}{1+m}} \tag{16}$$

The significance of this result arises from comparison with the bounds on the probability of error previously obtained by Elias (1, 2) for the case in which m = 1 (that is to say, in which the receiver selects only the single most-probable message). The first of these earlier results bounds the ensemble average probability of error.

$$P_1(e) \leq A_t\, 2^{-n\left[H(p_o)-H(p_t)+(p_t-p_o)\log\frac{q_o}{p_o}\right]} \qquad \left(p_o < p_t \leq p_{crit_1} < \frac{1}{2}\right) \tag{20}$$

where

$$\frac{q_{crit_1}}{p_{crit_1}} = \left(\frac{q_o}{p_o}\right)^{\frac{1}{2}} \tag{21}$$

and A_t is a function that decreases (slowly) with n as $1/n^{1/2}$. The second result gives for large n an asymptotic lower bound to the best possible error probability, $P_1(e)_{opt}$, that is obtainable with a given channel, code-word length, and number of possible messages.

$$P_1(e)_{opt} \underset{\sim}{>} A_{opt}\, 2^{-n\left[H(p_o)-H(p_t)+(p_t-p_o)\log\frac{q_o}{p_o}\right]} \tag{22}$$

where A_{opt} also decreases as $1/n^{1/2}$. For the symbol $\underset{\sim}{>}$, read "is asymptotically greater than...for large n."

It is seen that all of these expressions have the same exponential behavior. But when the number of possible messages S is so small that p_t is greater than p_{crit_1}, Elias (1, 2) shows that the exponential behavior of the ensemble average error probability $p_1(e)$ is not so good as that of $P_1(e)_{opt}$. Essentially, for high transmission rates $\left(p_t \leq p_{crit_1}\right)$, the probability of error is dominated by the probability that too many channel errors may occur during transmission. Appositively, the ensemble average error probability for low transmission rates $\left(p_t > p_{crit_1}\right)$ is dominated by the probability that two or more messages in a randomly selected transmitter set may differ too slightly from each other.

The purpose of the derivation presented here is to show that by allowing the receiver to select the m most probable messages, instead of only the single most probable one, this deficiency of random coding is removed. By making m sufficiently large, random coding can always be made exponentially as effective as optimum coding, in the limit of large code-word length. Furthermore, as shown by Eq. 16, in usual practice m need not be large in order to achieve this result.

J. M. Wozencraft

References

1. P. Elias, Coding for noisy channels, IRE Convention Record, Vol. 3, Part 4, 1955, pp. 37-44.

2. P. Elias, Coding for two noisy channels, Information Theory, Third London Symposium, September 12-16, 1955, edited by C. Cherry (Academic Press, Inc., New New York; Butterworths Scientific Publications, London, 1956).

3. P. Elias, List decoding for noisy channels, Technical Report 335, Research Laboratory of Electronics, M.I.T. (in press).

4. C. Shannon, Certain results in coding theory for noisy channels, Information and Control 1, 6 (Sept. 1957).

5. J. M. Wozencraft, Sequential decoding for reliable communication, Technical Report 325, Research Laboratory of Electronics, M.I.T., Aug. 9, 1957.

D. PICTURE PROCESSING

A large general-purpose digital computer provides a flexible means for studying characteristics of pictures. In order to exploit the capabilities of the Whirlwind computer, equipment has been constructed to scan a black-and-white photograph, sample the scanner output, and record the samples in digital form on paper tape. Tapes can also be reconverted to photographic images on photosensitive paper. Specifications for the terminal equipment are as follows:

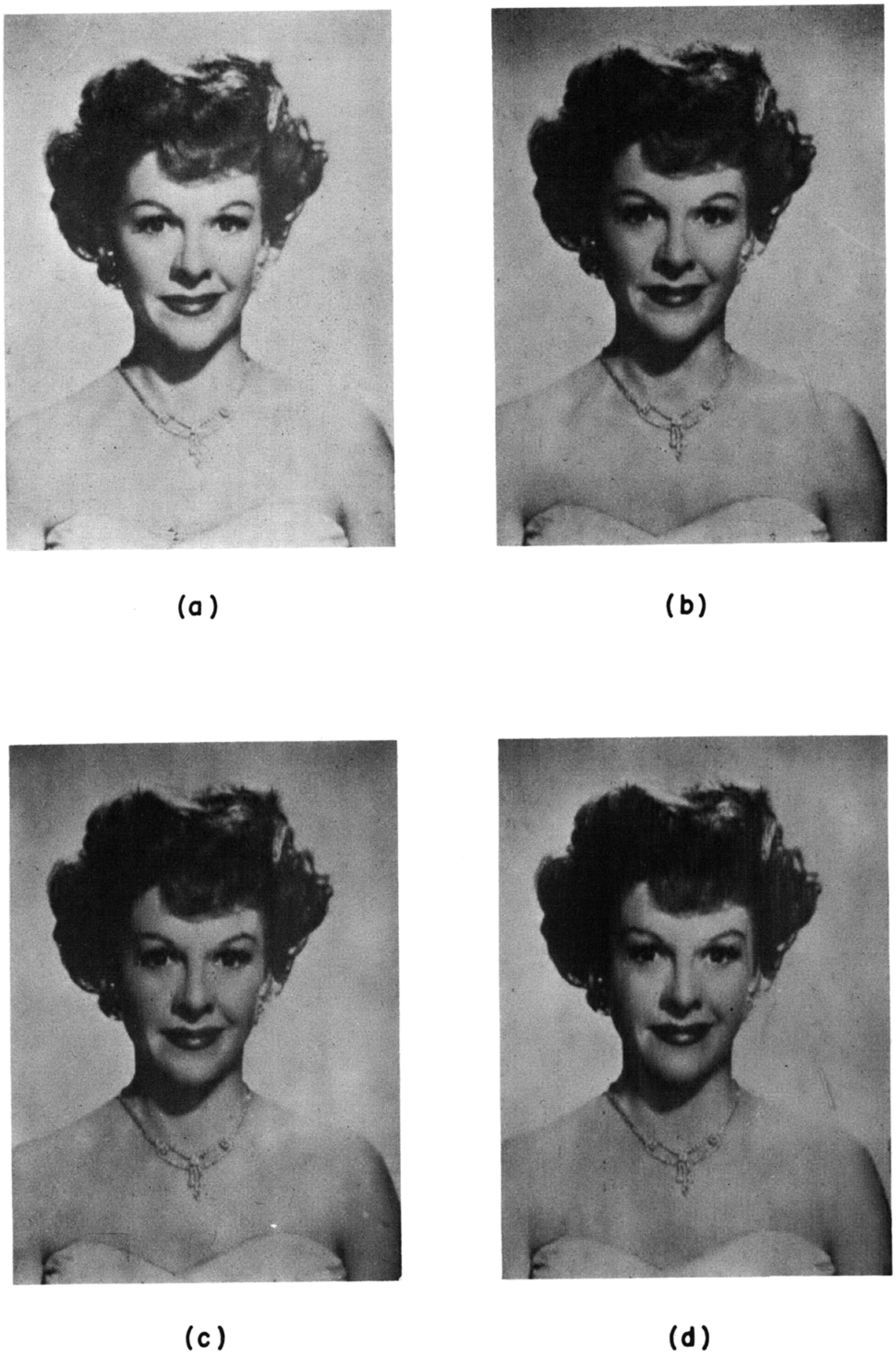
Scanning rate: 96 lines/inch
Sample rate: 93/inch
Scanner aperture: 0.01 inch square, approximate (for both recording and reproducing)
Quantization: 64 levels.

The effect of scanning without sampling or quantization is shown in Figs. XII-1a and XII-2a. In Figs. XII-1b and XII-2b the pictures have been reproduced from digital data recorded on paper tape. In the examples presented there are 240 lines in the shorter dimension and 320 samples per line in the longer dimension, the direction of scanning.

The first computer program used in this work was designed to generate a piecewise-linear approximation to the amplitude of the picture signal along scan lines. Relations between samples in the direction normal to scanning are not considered in this operation. For all computations, the picture is assumed to be surrounded by a completely black region with an initial sample value of zero for each line.

The operation of the linear approximation is explained with reference to Fig. XII-3. From a straight line with zero as origin and the second picture sample as terminus, a new quantized value is computed for the intervening point. The actual sample value is compared with that from the straight line. If the magnitude of error at the intervening point is less than the criterion chosen (3 levels for the example), the straight-line samples are temporarily stored. Then samples are computed for a new line with the same origin but with its terminus moved one sample interval in the direction of scanning. The procedure is repeated until the error at one or more interval points of the line first exceeds the criterion. When failure occurs, the samples from the last successful line are permanently recorded. A new straight-line approximation is now started with the terminus of the previous line as its origin. In order to limit the computation, the maximum length of line allowed is 32 sample intervals. The total number of runs of each length is counted as computation proceeds. From the frequency of occurrence of the various line lengths, a probability distribution of run length can be computed. The distribution in Fig. XII-4 is an example.

The term "source rate" is defined here as the minimum channel capacity required for transmission of the data for a given reproduction. An estimate of source rate is easily calculated by using the formula: Source-rate estimate = $\frac{H_r + N}{\bar{r}}$ bits/sample, where H_r is the entropy of run-length distribution in bits; N is the number of binary digits used to specify sample value at the end of the run; and $\bar{r}$ is the average run length in sample intervals.



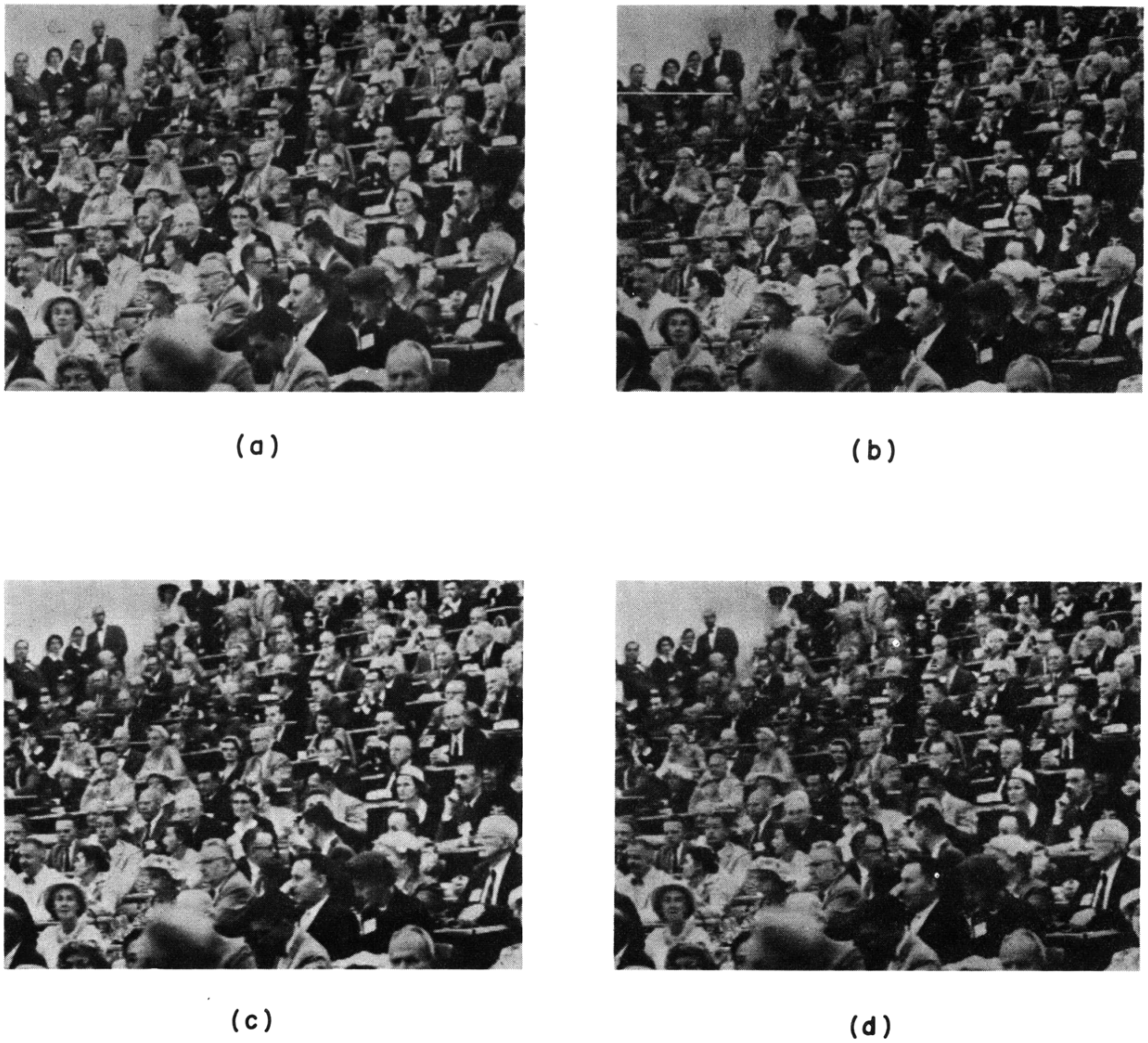
(a)

(b)

(c)

(d)

Fig. XII-1. Tests from picture with little detail: (a) scanned; (b) scanned, sampled, quantized, 6 bits; (c) processed, criterion 3, 1.29 bits; (d) processed, criterion 5, 0.97 bits.



(a)

(b)

(c)

(d)

Fig. XII-2. Tests from picture with detail: (a) scanned; (b) scanned, sampled, quantized, 6 bits; (c) processed, criterion 3, 3.34 bits; (d) processed, criterion 5, 2.79 bits.

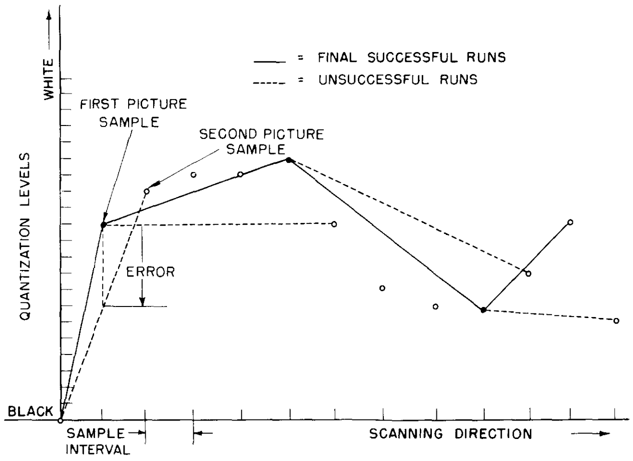


Fig. XII-3. Piecewise-linear approximation. (Maximum magnitude for quantized approximation error is 3 levels.)

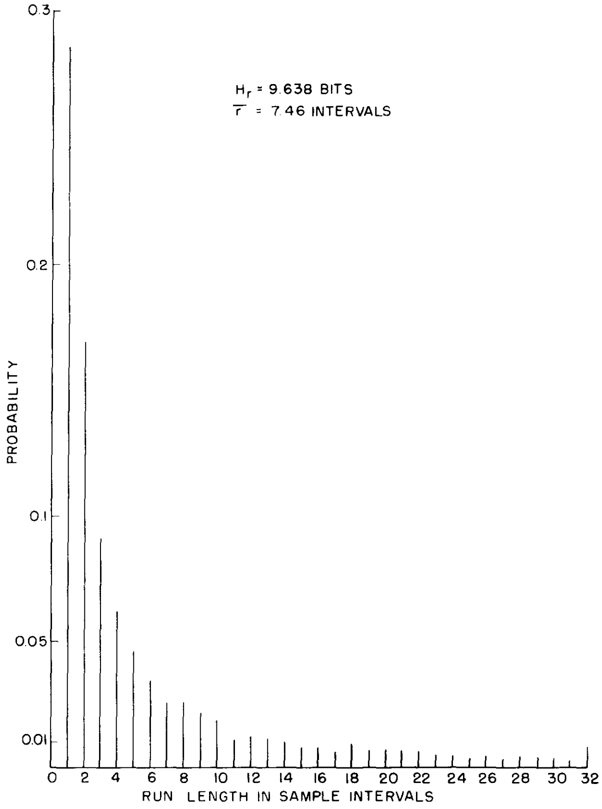


Fig. XII-4. Typical run-length probability distribution. (This example was obtained from Fig. XII-1c.)

Tests have been performed with criteria of 3 and 5 for computing approximations to the quantized pictures shown in Figs. XII-1b and XII-2b. Source-rate estimates for the reproductions of Figs. XII-1c, 1d, XII-2c, and 2d are given in Table XII-1.

Table XII-1. Source-Rate Estimates.

Figure	Picture	Source-Rate Estimate (bits)	Average Run Length (samples)
	Quantized Pictures	6	1
XII-1c	Girl (Criterion 3)	1.29	7.46
XII-1d	Girl (Criterion 5)	0.97	10.12
XII-2c	Crowd (Criterion 3)	3.34	2.48
XII-2d	Crowd (Criterion 5)	2.74	3.21

The effects of other methods of approximation for picture transmission are being investigated.

W. A. Youngblood

E. ALGEBRAIC DECODING FOR THE BINARY ERASURE CHANNEL

An algebraic decoding procedure for the binary erasure channel is being investigated. This procedure is interesting because the application of this scheme to convolution parity-check codes requires an average amount of decoding computation per digit which is bounded by a given number, even though the code length is arbitrarily large and the probability of error is arbitrarily small.

The transmitter of a binary erasure channel can transmit only the binary symbols 0, 1. The receiver of the channel either receives the transmitted digit correctly or the digit is erased and an X is received. To decode, the transmitted values of the erased digits must be determined.

Elias (1) has shown that parity-check codes can be coded with an amount of computation per digit that is proportional to code length and that the probability of error for codes with a given rate decreases exponentially with code length. If, in addition, the algebraic procedure is used to decode, we have a complete coding-decoding procedure which requires little computation and behaves almost as well as the best possible code.

For the purpose of illustration, we can apply the procedure to a block parity-check code. In a block parity-check code of length n and rate k, nR information digits are encoded into a transmitted message of n digits as follows. The first nR digits of the message are the information digits (see Fig. XII-5). The last n(1-R) digits of the message are the check digits and they are determined by Eq. 1 and the matrix a_{ij}, whose elements are either zero or one.

$$\sum_{j=1}^{nR} a_{ij} I_j = C_i \pmod 2 \qquad i = 1, 2, \ldots, n(1-R) \tag{1}$$

In transmission the digits are erased at random with a certain probability p that is characteristic of the channel (see Fig. XII-6). Usually, some information and some check digits will be erased. The decoding is done by means of the received message and the parity-check equations. The values of the unerased digits of the received message are inserted into each parity-check equation to form the constant, and the erased digits are treated as unknowns. When this has been done, we have n(1-R) linear modulo-two equations in the erased digits. By means of the modulo-two analog of the Gauss-Jordan procedure, we can find out if the erased digits are determined by the parity-check equations. If the equations determine the erased digits, the procedure will solve for the values of these digits (see Fig. XII-7). The average amount of computation for decoding a block of n digits is proportional to n^3. This means that the average number of computations per digit is proportional to the square of the block length, so that as the block length becomes infinite the average number of computations per digit becomes infinite.

The most important result of this investigation arises from the application of this new algebraic decoding procedure to convolution parity-check codes. A convolution parity-check code of length n is a code in which information and check digits are intermingled, and the check digits are determined by parity checks on some of the information digits in the preceding n digits. Figure XII-8 illustrates a convolution parity-check matrix and Eq. 2 represents the equations for such a code, in which p(i) is defined as the position of the i^{th} check.

$$d_{p(i)} = \sum_{j=p(i)-n}^{p(i)-1} a_{ij} d_j \pmod 2 \qquad i = 1, 2, \ldots \tag{2}$$

In the decoding procedure for the convolution code digit 1 is decoded first, then digit 2, then digit 3, and so on. This decoding procedure is similar to the decoding procedure developed by Wozencraft (2) for the binary symmetric channel. Because digits are decoded in sequence, digits 1, 2, ..., m-1 are known, and all terms in the parity-check equations that contain these digits are known, when digit m is being decoded. This fact leads to the important result that decoding a given digit is essentially the same as decoding any other digit. Therefore, the following discussion is limited to the problem of decoding a typical digit.

A given digit is decoded in a series of steps. Thus, we attempt, first, to decode with one equation, then, with two equations, and so on. The process terminates when the digit is decoded. In the step in which m equations are used the computational difficulty is the same as for a block code of length m/1-R and the number of computations

$$\begin{array}{c} 1 \quad 0 \quad 1 \\ \text{Information Digits} \end{array} \qquad \begin{array}{c} \boxed{\begin{matrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{matrix}} \\ \text{Parity-Check Matrix} \end{array} \qquad \begin{array}{c} \begin{array}{cccccc} I_1 & I_2 & I_3 & C_1 & C_2 & C_3 \\ \hline 1 & 0 & 1 & 0 & 1 & 1 \end{array} \\ \text{Transmitted Message} \end{array}$$

Fig. XII-5. Coding at the transmitter.

$$\begin{array}{c} \begin{array}{cccccc} I_1 & I_2 & I_3 & C_1 & C_2 & C_3 \\ \hline 1 & 0 & 1 & 0 & 1 & 1 \end{array} \\ \text{Transmitted Message} \end{array} \longrightarrow \boxed{\begin{array}{c} \text{Binary} \\ \text{Erasure} \\ \text{Channel} \end{array}} \longrightarrow \begin{array}{c} \begin{array}{cccccc} I_1 & I_2 & I_3 & C_1 & C_2 & C_3 \\ \hline 1 & X & 1 & 0 & 1 & X \end{array} \\ \text{Received Message} \end{array}$$

Fig. XII-6. Transmission through the channel.

$$\begin{array}{c} \begin{array}{cccccc} I_1 & I_2 & I_3 & C_1 & C_2 & C_3 \\ \hline 1 & X & 1 & 0 & 1 & X \end{array} \\ \text{Received Message} \\ I_2,\ C_3 \text{ Erased} \end{array} \qquad \begin{array}{c} 1 \oplus X_1 \oplus 1 = 0 \\ 0 \oplus 0 \oplus 1 = 1 \\ 1 \oplus X_1 \oplus 0 = X_2 \\ \text{Decoding Equations} \\ X_1 = I_2;\ X_2 = C_3 \end{array} \qquad \begin{array}{c} X_1 = 0 \\ X_2 = 1 \\ \text{Solution} \end{array} \qquad \begin{array}{c} 1 \quad 0 \quad 1 \\ \text{Information Digits} \end{array}$$

Fig. XII-7. Decoding at the receiver.

$$\begin{array}{|cccccccccc} \multicolumn{1}{c}{d_1} & d_2 & d_3 & d_4 & d_5 & d_6 & d_7 & d_8 & d_9 & d_{10} \\ \multicolumn{1}{c}{I_1} & C_1 & I_2 & C_2 & I_3 & C_3 & I_4 & C_4 & I_5 & C_5 \\ \hline 1 & 1 & & & & & & & & \\ 1 & & 1 & 1 & & & & & & \\ & & 0 & & 1 & 1 & & & & \\ & & & & 1 & & 0 & 1 & & \\ & & & & & & 0 & & 1 & 1 \end{array}$$

Fig. XII-8. Convolution matrix of length 4. (Blank spaces are zero.)

required is proportional to m^3. With the help of Elias' (1) results, we can show that the probability of reaching step m decreases exponentially and can be bounded by ce^{-am}. Thus the average number of computations needed to decode one digit is equal to the sum, over the maximum number of steps possible, of the average number of computations needed for each step and is bounded by $\Sigma\, cm^3\, e^{-am}$. This sum is bounded by some finite number B, which is independent of the code length but depends on the channel capacity and the rate of transmission, since these factors determine the constants "a" and "c" in the exponential bound. Hence, the average number of computations needed to decode any given digit is bounded by B, regardless of the convolution code length.

M. A. Epstein

References

1. P. Elias, Coding for two noisy channels, Information Theory, paper presented at Third London Symposium, September 12-16, 1955, edited by C. Cherry (Academic Press, Inc., New York; Butterworths Scientific Publications, London, 1956).

2. J. M. Wozencraft, Sequential decoding for reliable communication, Technical Report 325, Research Laboratory of Electronics, M.I.T., Aug. 9, 1957.