IX. PROCESSING AND TRANSMISSION OF INFORMATION

Prof. S. H. Caldwell
Prof. P. Elias
Prof. R. M. Fano
Prof. D. A. Huffman
W. G. Breckenridge
J. Capon
E. Ferretti
E. J. McCluskey, Jr.
A. J. Osborne
R. A. Silverman
J. C. Stoddard
F. F. Tung
S. H. Unger
W. A. Youngblood

A. ENTROPY OF PRINTED ENGLISH

The two curves shown in Fig. IX-1 are upper and lower bounds to the entropy per letter of printed English; they were obtained from a prediction experiment of the type suggested by C. E. Shannon (1). The subjects used in the experiment were students in the graduate course in Information Theory at M.I.T. These instructions were given to them:

"The following experiment is to be carried out as a class project from which useful data will be obtained about the information properties of the English language. The experiment requires the help of another person.

"Select a book, preferably a modern novel, which you have not yet read. Select a number p between 1 and the total number of pages in the book, a number ℓ between 1 and 30 and a number s between 1 and 50. Have the helper find a position in the book on the p^{th} page, the ℓ^{th} line and the s^{th} letter position in the line (spaces included but punctuation marks not included). Try to guess the $(s+1)^{th}$ letter in the line selected (space being considered as a letter). The helper will tell you if your guess is correct; if not correct, guess again until you guess correctly. Write the correct letter and under it the number of guesses that you have made. Then try to guess the next letter, in the same manner, and write finally the correct letter and the number of guesses required to obtain it. Continue this procedure for the following letters up to and including the tenth letter. Have the helper read to you the next nine letters and record them; then try to guess the 20th letter as before. Proceed in the same manner to guess every tenth letter up to and including the 100th.

"Repeat the above procedure 10 times for different texts; use, if possible ten different books. Otherwise, divide the number of pages in the book available in ten parts and select at random a page in each part.

"Record the number of correct first guesses, second guesses, etc. for each letter."

A total of 295 sequences of 100 letters was analyzed.

Let $f_r(n)$ be the number of sequences in which the n^{th} letter was guessed correctly at the r^{th} guess, and let $p_r(n) = f_r(n)/295$. The upper bound was obtained from the equation

$$H_u(n) = -\sum_{r=1}^{27} p_r(n) \log_2 p_r(n)$$

The lower bound was obtained from the equation

$$H_L(n) = \sum_{r=2}^{27} r\left[p_r(n) - p_{r+1}(n)\right] \log_2 r$$

$$= \sum_{r=2}^{27} p_r(n) \log_2 \frac{r^r}{(r-1)^{r-1}}$$

Comparison of the results obtained with those published by Shannon shows Shannon's values to be somewhat lower. It should be noted that Shannon's values for n = 1 and n = 2 were obtained directly from available letter frequencies and digram frequencies, while corresponding values in Fig. IX-1 were obtained from the prediction experiment.

The data were not smoothed other than in drawing the two curves. The size of the sampling errors increases with n, as expected. The lower bound seems to tend asymptotically to a value equal approximately to one bit per letter. The asymptotic value of the upper bound appears to be in the neighborhood of 1.6-1.7 bits per letter.

R. M. Fano

Reference

1. C. E. Shannon, Bell System Tech. J. XXX, No. 1, p. 50 (Jan. 1951).

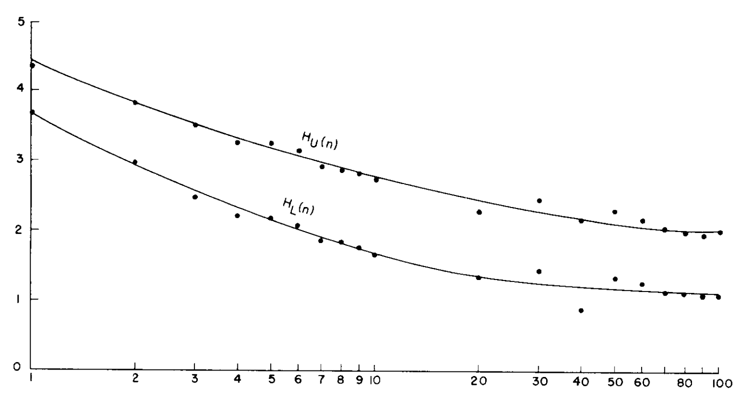

Fig. IX-1

Upper and lower bounds to the entropy of printed English.

B. BOUNDS TO THE ENTROPY OF TELEVISION SIGNALS

During the past year, equipment was built for measuring second-order probability densities of the light intensities of pictures. There is also available a device for measuring the corresponding two-dimensional autocorrelation function. These statistical parameters will be used to compute upper bounds to the entropy of pictures.

Previous investigations in this field were made by E. R. Kretzmer (1) and W. F. Schreiber (2). Kretzmer, at Bell Telephone Laboratories, measured autocorrelation functions of pictures. Schreiber measured second-order probabilities. But these measurements were not made on the same picture. This experiment will differ from the work of the two previous investigators in that both measurements will be made on the same picture and the results compared. With the kind permission of the Cruft Laboratory of Harvard University, part of the equipment used by Schreiber is being used here. Although the measurements are being done in analog, equipment is also available for measuring these quantities digitally (see Sec. IX-C).

The equipment used for the second-order probability measurement is shown in block diagram form in Fig. IX-2.

The synchronization generator provides the 15.75 kc/sec and 60-cps pulses necessary to trigger the sweeps of the sweep generator. These repetition rates correspond to those used in present-day television transmission. The sweep generator, in turn, deflects the beam in the flying spot scanner, forming a raster on the face of the tube. Two 35-mm transparencies are scanned by means of a beam splitter that consists of a half-aluminized mirror that reflects half of the impinging light and transmits the other

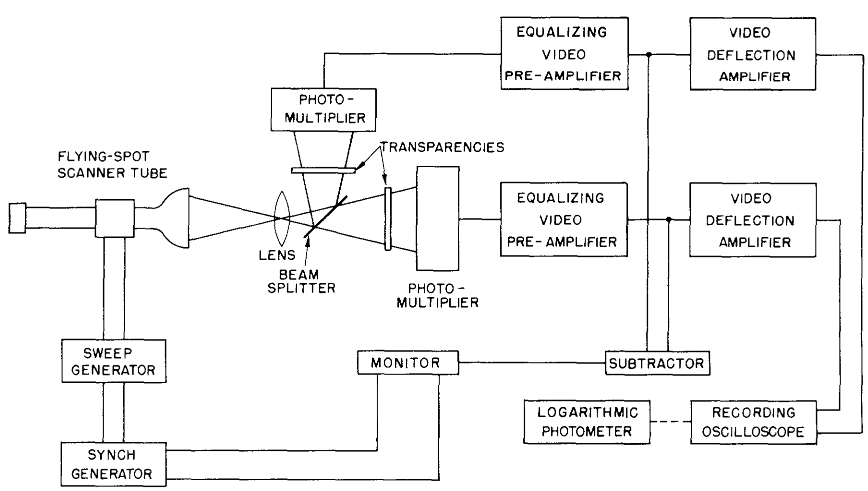

Fig. IX-2
Block diagram for the probability measurement.

half. The two transparencies are contained in holders that may be moved relative to each other. This enables the beam to scan the two pictures in register or slightly out of register. The light signals resulting from the scanning are converted to electric signals in the photomultipliers and then amplified in the video preamplifiers. They are also equalized in the preamplifiers in order to compensate for the persistence of the phosphor on the flying-spot tube. The video signals emerging from the preamplifiers are amplified in the video deflection amplifiers and applied to the recording scope. The preamplifier outputs are also sent into a subtracter, where their difference is obtained and applied to a monitoring circuit. The monitor consists of the video amplifiers and picture tube of a commercial television set which is used to look at the difference of the two video signals. Except for these amplifiers (which probably have a bandwidth of 4 Mc/sec) all units have a bandwidth of 10 Mc/sec.

Now that the components have been described, an idea of how the system functions can be obtained by first considering that the two transparencies are being scanned in register and that two identical video signals are being generated. Assuming that the gain and phase shift through the video amplifiers are the same, identical signals are applied to the recording scope. It is well known that if equal signals are applied to the two pairs of deflection plates of an oscilloscope, the pattern that results is a 45° line. A very sensitive indication of the pictures' being in register is given on the screen of the monitor, since a complete null is observed. If one of the transparencies is displaced by a unit cell, the 45° line should fan out into a scatter diagram. In this experiment it is assumed that a picture is composed of an array of $(525)^2$ cells, 525 vertically and 525 horizontally, corresponding, respectively, to the 525 scanning lines and 525 resolvable picture elements per line of the standard television raster. Thus a unit cell displacement corresponds to 1/525 of the horizontal length being scanned.

Since the beam current is constant, the brightness of any typical element x_iy_j of the scatter diagram shown in Fig. IX-3 is a direct indication of the amount of time that one signal is dwelling at level i while the other signal is dwelling at level j. This, in turn, is proportional to the probability that a pair of adjacent cells are at levels i and j, respectively. The number of levels for i and j is assumed to be 32, so that a total of 1024 probabilities is obtained.

The logarithmic photometer is used to measure the light from the face of the recording scope. It employs a 931-A tube in a feedback circuit (3), which tends to stabilize the output of the photomultiplier. The calibration of the photometer is good to within 5 per cent.

Since noise is introduced in the photomultipliers, a perfect 45° line is not obtained. Instead, the line is slightly blurred – as would be expected. It is this noise that will ultimately limit the precision and accuracy of the results and the extent to which small probabilities can be measured.

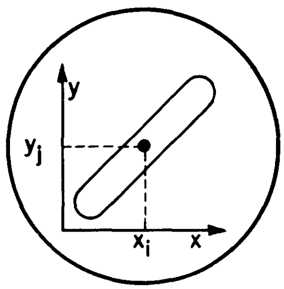

Fig. IX-3
Typical pattern on recording scope.

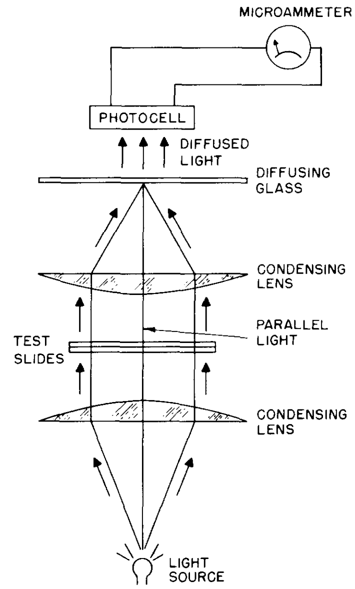

Fig. IX-4
Schematic representation for autocorrelation measurements.

It should also be mentioned that the light-emitting characteristic of the flying-spot scanner phosphor is not uniform. However, it is felt that this condition can be remedied. By exposing a negative so that its contrast is opposite to that of the phosphor characteristic and placing it in front of the phosphor, we can obtain an over-all uniform transmission.

The apparatus used to measure autocorrelation (4) is shown schematically in Fig. IX-4. The pictures in this case are on glass slides (4 inches by 3.25 inches). The two slides are shifted past each other and the amount of light transmitted is measured. If the autocorrelation function is thought of as $\overline{f(t)\,f(t+\tau)}$, then by substituting for f(t) the transmittance of the picture we see that the light transmitted for any given shift τ is proportional to the value of the autocorrelation function for that shift. From this measurement, an upper bound to the entropy per symbol (5) can be computed, as mentioned earlier.

All of the equipment has been built and tested and is ready for actual measurements.

J. Capon

References

1. E. R. Kretzmer, Bell System Tech. J. 31, 751-763 (1952).
2. W. F. Schreiber, Ph. D. Thesis, Harvard University (Dec. 1952).
3. M. H. Sweet, J. Soc. Motion Picture Televis. Engrs. 54, 34 (Jan. 1950).
4. Kretzmer, loc. cit.
5. C. E. Shannon, The Mathematical Theory of Communication (University of Illinois Press, Urbana, 1949) p. 57.

C. MEASURING SECOND-ORDER PROBABILITY DISTRIBUTIONS BY DIGITAL MEANS

Digital equipment for measuring second-order probability distributions of an input voltage has been built. The input voltage is sampled and quantized, and the joint probability of two adjacent quantized samples is measured.

The equipment was built as part of the research program on the transmission of pictures. A facsimile transmitter is used to convert light intensity variations in a picture into a varying voltage that is then fed to the statistics measuring equipment. The second-order probability distribution of intensity for adjacent samples of a picture can be used to set an upper bound to the communication entropy of pictures and, hence, to the bandwidth which could be saved by a suitable coding scheme. (Equipment for measuring second-order probability distributions by analog is described in Sec. IX-B.)

The digital procedure is basically one of counting relative frequencies. The picture waveform is sampled, quantized, and recorded in pulse-code form on magnetic tape. The tape is played back, and joint occurrences of combinations of intensity in adjacent samples are counted. For example, the occurrence of level five followed by level five would be counted. One would expect this to be a more common occurrence than the occurrence of level five followed by level twenty-five. The equipment quantizes the signal into thirty-two levels. The complete measurement of a second-order probability distribution involves the counting of the occurrences of 32^2, or 1024, joint events. Since it is impractical to have that many counters, the tape has to be played back many times and the occurrences of different events counted each time. By recording the sampled and quantized signal on the tape, the input signal does not have to be reproduced many times.

A block diagram of the recording equipment is shown in Fig. IX-5. The analog-to-digital converter is an experimental coding tube developed at the Bell Telephone Laboratories (1). It was made available to us through the courtesy of Mr. R. W. Sears (1) of that Laboratory. The tube is a special cathode-ray tube equipped with a perforated target. The electron gun forms a horizontal beam on the target. The signal to be quantized

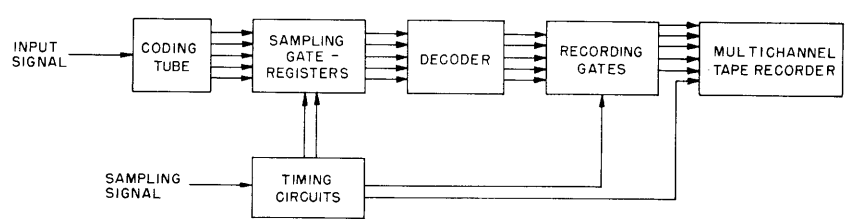

Fig. IX-5
Recording equipment.

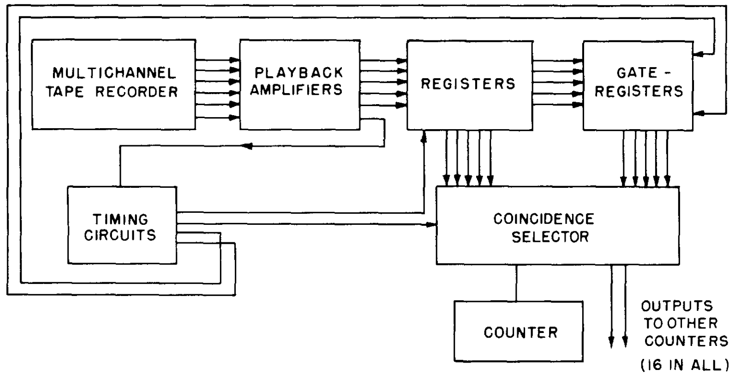

Fig. IX-6
Playback equipment.

is applied to a pair of deflection plates that position the line on the target vertically. Current collectors are positioned behind the target plate so that the collectors receive or do not receive electrons through the perforations, depending on the vertical position of the beam and hence on the value of the applied signal. The apparatus uses 2^5, or 32, quantum levels, although the coding tube works well for 2^6, or 64, levels.

The outputs from the coding tube are continuous binary voltages that are sampled by sampling gate-registers that hold the samples so that long pulses may be used for more efficient recording without danger of error caused by an input changing a level during the longer pulse. The decoder is inserted between the sampling gate-registers and the recording gates to convert the reflected binary code from the output of the coding tube to the conventional binary code. The conventional code is recorded so that the quantized and sampled picture can be recovered and so that one or more digits representing the signal can be eliminated to cut short the measurement process. The recording gates supply the actual pulse current to the tape heads. The sampling rate was chosen to be 1250 cps, twice the highest frequency present in the output of the facsimile transmitter. This rate may be reduced for other signal inputs, or increased with some modification of the equipment.

The tape recorder is a commercial unit modified for 35-mm tape and equipped with eight heads for simultaneous recording of as many as eight signals (2).

A block diagram of the equipment used to play back the tape and count the probabilities is shown in Fig. IX-6. The registers are used to store the present sample. The gate-registers are used to store the previous sample so that the two sets of binary digits, representing the present and the previous sample, are available simultaneously. The coincidence selector is a device that gates out a pulse when the desired combination of past and present binary digits occurs. The particular set of binary digits (out of the 1024 possible sets) that causes this gating-out of a pulse is determined by switches on the selector and by the output employed. In actual use, one counter is connected to each of the sixteen outputs of the selector, so that sixteen of the 1024 possible combinations of events can be counted each time the tape is played back. The actual second-order probabilities are obtained from the counts of joint events by dividing the count by total number of events.

The counters themselves are of standard-design using a binary scale of 256 in front of an electromechanical register.

All equipment has been built and the individual components tested. The system is being tested and calibrated, in preparation for actual measurements on pictures.

J. C. Stoddard

References

1. R. W. Sears, Bell System Tech. J. 27, 44-57 (Jan. 1948).
2. L. Dolansky, S.M. Thesis, Department of Electrical Engineering, M.I.T. (May 1949).

D. CODING FOR NOISY CHANNELS

Considerable progress has been made in the study of the transmission of information over the noisy binary channel. This work originated in an attempt to improve the performance of the error-free codes discussed earlier (1, 2). Some new check-symbol codes were discovered. These codes transmit at rates arbitrarily close to the channel capacity, with error probability as small as the receiver cares to set it; the decrease of error probability with delay is exponential. In order to evaluate these codes it was necessary to compute the best obtainable error probability for any code for this channel. The results on error probability are presented first and are followed by results showing that error-free coding can be accomplished by a stationary procedure with essentially no sacrifice, from either a channel-capacity or an error-probability point of view, as compared to the best possible code. Finally, some results are given for a channel with binary input and ternary output, the zero-x-one channel. More precise statements and demonstrations of all of these results will be given in Technical Report 291 (to be published).

1. The Binary Symmetric Channel

Given a binary symmetric channel with transmission error probability p_o and $q_o = 1 - p_o$, the equivocation E_o and the capacity C_o are given by

$$E_o = -p_o \log p_o - q_o \log q_o \qquad (1)$$
$$C_o = 1 - E_o$$

Using a transmission rate $C_1 < C_o$ and a redundancy of transmission $E_1 > E_o$, define p_1 as the upper bound of the transmission error probabilities for which transmission at this rate can be accomplished, and define $q_1 = 1 - p_1$, where p_1 is determined by

$$\begin{aligned} & p_1 < \frac{1}{2} \\ & E_1 = E(p_1) = -p_1 \log p_1 - q_1 \log q_1 \end{aligned} \tag{2}$$

since a plot of E(p) or C(p) is monotonic for $0 \leqslant p \leqslant 1/2$.

The average probability of making an error when decoding a block of N adjacent symbols is denoted by $P_e(N, p_o, p_1)$. It turns out that the error probability P_e is bounded not only above but below by exponentials in N, and that for a considerable range of channel and code parameters the exponents of the two bounds agree. The error exponent for the best possible code is defined as

$$a_{opt}(N, p_o, p_1) = \frac{-\log P_e(N, p_o, p_1)}{N} \tag{3}$$

and $a_{avg}(N, p_o, p_1)$ is defined as the same function of the average of the error probabilities of all codes.

An additional probability value is also needed, along with the values of a, C, and E which go with it:

$$\left.\begin{aligned} & p_{crit} = \frac{p^{1/2}}{p^{1/2} + q^{1/2}}, \quad q_{crit} = 1 - p_{crit} \\ & E_{crit} = E(p_{crit}), \quad C_{crit} = 1 - E_{crit} \\ & a_{crit} = \lim_{N \to \infty} a_{opt}(N, p_o, p_{crit}) \end{aligned}\right\} \tag{4}$$

Finally, the margin in error probability and the margin in channel capacity need labeling:

$$\begin{aligned} & \delta = p_1 - p_o \\ & \Delta = C_o - C_1 \end{aligned} \tag{5}$$

For a binary symmetric channel with capacity C_o and transmission rate C_1, the following statements hold.

THEOREM 1: "(a) For $p_o < p_1 < p_{crit}$, $C_o > C_1 > C_{crit}$, the average code is essentially as good as the best code:

$$a(p_o, p_1) = \lim_{N \to \infty} a_{opt}(N, p_o, p_1) = \lim_{N \to \infty} a_{avg}(N, p_o, p_1) = -\Delta - \delta \log \frac{p_o}{q_o} \tag{6}$$

(b) For $p_{crit} < p_1 < 1/2$, the average code is not necessarily optimum; for p_1 near 1/2 it is certainly not. Specifically,

$$\alpha_{avg}(p_o, p_1) = \lim_{N\to\infty} \alpha_{avg}(N, p_o, p_1) = \alpha_{crit} + C_{crit} - C_1 \tag{7}$$

where α_{crit} is the $\alpha(p_o, p_1)$ of Eq. 4 with $p_1 = p_{crit}$, while for α_{opt} there are two upper and two lower bounds:

$$\liminf \alpha_{opt}(N, p_o, p_1) \geq \begin{cases} \alpha_{crit} + C_{crit} - C_1 \\ \dfrac{p_1}{2} \log \dfrac{1}{4pq} - C_1 \end{cases} \tag{8}$$

$$\limsup \alpha_{opt}(N, p_o, p_1) \leq \begin{cases} -\Delta - \delta \log \dfrac{p_o}{q_o} \\ \dfrac{E_1}{4} \log \dfrac{1}{4pq} \end{cases} \tag{9}$$

As p_1 approaches 1/2, the second bound in Eq. 8 approaches the second bound in Eq. 9;

$$\lim_{N\to\infty} \lim_{p_1\to 1/2} \alpha_{opt}(N, p_o, p_1) = \frac{1}{4} \log \frac{1}{4pq} \tag{10}$$

which is always greater than

$$\alpha_{avg}\left(p_o, \frac{1}{2}\right) = \alpha_{crit} + C_{crit}." \tag{11}$$

The content of this theorem is illustrated by Fig. IX-7. This is a plot of the channel capacity C(p) vs. transmission error probability p for a binary symmetric channel. A dashed line is drawn tangent to the curve at the point given by the channel parameters p_o, C_o. This tangent line has the slope $\log (p_o/q_o)$. The critical point p_{crit}, C_{crit} is the point at which the slope of the curve is $(1/2) \log (p_o/q_o)$. For $p_o < p_1 < p_{crit}$, the $\alpha(p_o, p_1)$ of Eq. 6, which is both the average and the optimum error exponent, is the length of a vertical dropped from the channel capacity curve to the tangent line at the ordinate p_1.

At $p_1 = p_{crit}$, the dotted line that determines $\alpha_{avg}(p_o, p_1)$ diverges from the tangent line. For $p_{crit} < p_1 < 1/2$ the exact value of $\alpha_{opt}(N, p_o, p_1)$ is not known, but is given by the length of a vertical at ordinate p_1, dropped from the channel capacity curve and terminating in the shaded region. The upper and lower bounds of this region provide lower and upper bounds, respectively, on the value of α_{opt}. These bounds converge to $(1/4) \log (1/4pq)$ at $p_1 = 1/2$, and near this point α_{opt} is definitely greater than α_{avg}.

The value of α given by the tangent line at $p_1 = 1/2$, although not approached for

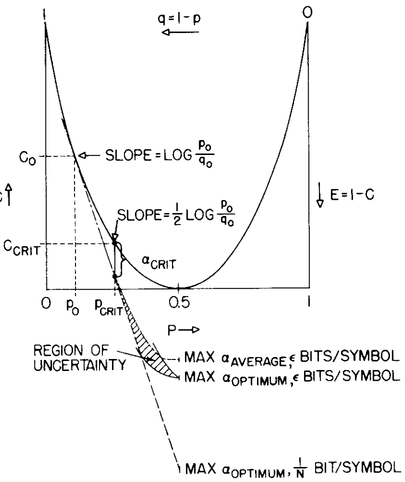

Fig. IX-7

The error exponent α_{optimum}.

the transmission of information at any nonzero rate, is the correct value of α_{opt} for transmission of one bit per block of N symbols.

The following results apply to four kinds of codes of increasing simplicity from the point of view of realization.

A check-symbol code of block length N is a code in which the 2^{NC_1} signal sequences have in their initial NC_1 positions all 2^{NC_1} possible combinations of symbol values. The first NC_1 positions will be called information positions and the last NE_1 will be called check positions. The signal corresponding to a message sequence is that one of the signal sequences whose initial symbols are the message.

A parity check-symbol (pcs) code is a check-symbol code in which the check positions are filled in with digits each of which completes a parity check of some of the information positions. Such codes were first discussed in detail by Hamming (3), who calls them systematic codes. A pcs code is specified by an $NC_1 \times NE_1$ matrix of zeros and ones, the ones in a row giving the locations of the information symbols whose sum modulo two is the check digit corresponding to that row. Such a code requires $NC_1 \times NE_1 = N^2C_1E_1 < \frac{1}{4}N^2$ binary digits in its codebook, these being the digits in the check-sum matrix.

A sliding pcs code is defined as a pcs code in which the check-sum matrix is constructed from a sequence of N binary symbols by using the first NC_1 of them for the first row, the second to $(NC_1 + 1)$st for the second row ..., the NE_1th to the Nth for the NE_1th row. This code requires only an N-binary-digit codebook.

Finally, a convolutional pcs code is defined as one in which check symbols are interspersed with information symbols, and the check symbols check a fixed pattern of the preceding NC_1 information positions if $C_1 \geqslant 1/2$; if $E_1 > 1/2$, the information symbols add a fixed pattern of zeros and ones to the succeeding NE_1 check positions. Such a code requires $\max(NC_1, NE_1) \leqslant N$ binary digits in its codebook.

THEOREM 2: "All of the results of Theorem 1 apply to check-symbol codes and to pcs codes. The results of part (a) of that theorem apply to sliding pcs codes."

In reading Theorem 1 into Theorem 2, the average involved in α_{avg} is the average of all codes of the appropriate type; that is, all combinations of check symbols for the check-symbol codes, all check-sum matrices for the pcs codes, all sequences of N binary digits for the sliding pcs code.

THEOREM 3: "The results of part (a) of Theorem 1 apply to convolutional pcs codes, if $P_e(N, p_o, p_1)$ is interpreted as the error probability per decoded symbol. For infinite memory (each check symbol checking a set of prior information symbols extending back to the start of transmission over the channel) the N in $P_e(N, p_o, p_1)$ for a particular decoded information symbol is the number of symbols which have been received since it was received."

This theorem shows that error-free coding can be attained at no loss either in channel capacity or in error probability, a question raised by the author when the first error-free code was introduced (2). By waiting long enough, the receiver can obtain as low a probability of error per digit as is desired without a change of code. By gradually reducing the ratio of check to information symbols towards E_o/C_o, using the law of the iterated logarithm for binary sequences, it can be shown that in an infinite sequence of message digits transmission is obtained at average rate C_o with probability one of no errors in the decoded message.

2. The Zero-x-One Channel

Let a channel with binary input (0, 1) and ternary output (0, x, 1) be specified by a symmetrical error probability p_o that an input symbol (0 or 1) will turn into an output x, and probability zero that an input 0 will turn into an output 1, or vice versa. Let $q_o = 1 - p_o$. The capacity C_o and equivocation E_o of this channel are defined by

$$\begin{aligned} C_o &= q_o \\ E_o &= p_o \end{aligned} \tag{12}$$

Given a signaling rate $C_1 < C_o$, define

$$\left.\begin{aligned} q_1 &= C_1 \\ E_1 = p_1 &= 1 - q_1 \\ \delta &= p_1 - p_o \end{aligned}\right\} \tag{13}$$

Define q_{crit} and p_{crit} by

$$p_{crit} + q_{crit} = 1$$
$$\frac{p_{crit}}{q_{crit}} = 2\frac{p_o}{q_o} \tag{14}$$

Define $P_e(N, p_o, p_1)$, $\alpha_{opt}(N, p_o, p_1)$, $\alpha_{avg}(N, p_o, p_1)$ and $\alpha_{crit}(p_o, p_1)$ as in Eqs. 3 and 4 above, with the interpretations of Eqs. 12 and 13 for p_o and p_1.

THEOREM 4: "(a) The results of part (a) of Theorem 1 apply, as written, to the zero-x-one channel. Their extension in Theorems 2 and 3 to pcs, sliding pcs, and convolutional codes also apply as written, as does the error-free behavior of convolutional codes.

(b) For $p_{crit} < p_1 < 1$, random coding is not necessarily optimum. For message coding and for pcs coding, for p_1 sufficiently near 1, it is certainly not."

As applied to this channel, pcs codes give to the statement of the channel capacity theorem a particularly piquant flavor. Given any $\delta > 0$, it is possible to add to a sequence of $N(q_o - \delta)$ information symbols a sequence of $N(p_o + \delta)$ check symbols in such a fashion that although the channel strikes out about Np_o symbols of both types at random, the information symbols can all be reconstructed with an error probability that decreases exponentially with N.

P. Elias

References

1. Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., April 15, 1954, p. 51.
2. P. Elias, Error-free coding, Trans. I.R.E. (PGIT) 4, 30-37 (1954).
3. R. W. Hamming, Error detecting and error correcting codes, Bell System Tech. J. 31, 504-522 (1952).

E. LINEAR BINARY SEQUENCE FILTERS

Recent methods proposed for the generation of coded sequences for binary channels involve obtaining a continuous parity check over digits in positions fixed relative to the present time rather than parity checks only at the end of a block of digits. Consequently, a general study has been made of methods for the synthesis of circuits that accomplish these results.

A linear binary sequence filter is a switching circuit whose present output is the sum modulo two (result of a parity check) over selected past values of its own input and output sequences. For example, if we let the symbol "D" represent a delay operator, and

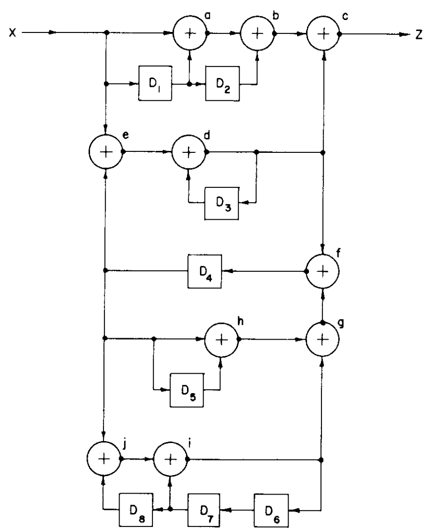

Fig. IX-8
A linear binary filter.

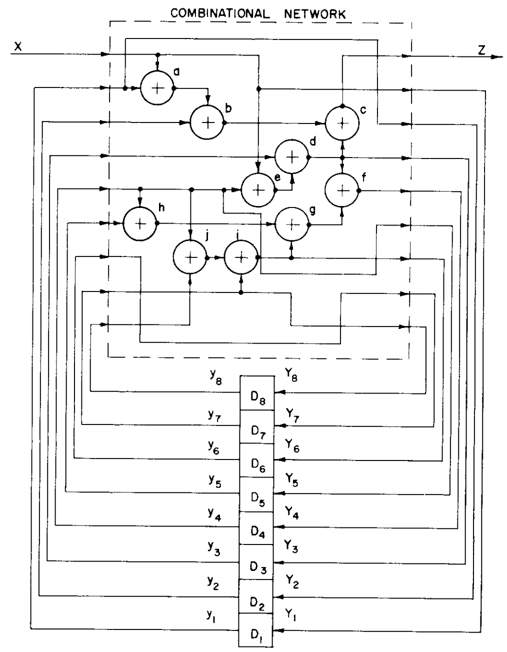

Fig. IX-9
The filter of Fig. IX-8 redrawn as a switching circuit with feedback.

if X and Z represent the input and output sequences of the filter, the equation

$$Z = (D^8 \oplus D^6 \oplus D^3 \oplus D^2 \oplus D)\,X \oplus (D^6 \oplus D^5 \oplus D^4 \oplus D^3)Z \tag{1}$$

describes a filter whose present output is found by adding modulo two the values of the input that occurred eight, six, three, two, and one unit of time into the past, together with the sixth, fifth, fourth, and third previous values of the output. Equation 1 may be rewritten as

$$(D^6 \oplus D^5 \oplus D^4 \oplus D^3 \oplus 1)Z = (D^8 \oplus D^6 \oplus D^3 \oplus D^2 \oplus D)X \tag{2}$$

or, finally, the relationship of Z to X may be expressed as a transfer function that is a ratio of polynomials in D:

$$\frac{Z}{X} = \frac{D^8 \oplus D^6 \oplus D^3 \oplus D^2 \oplus D}{D^6 \oplus D^5 \oplus D^4 \oplus D^3 \oplus 1} \tag{3}$$

The synthesis of a network with this transfer function is quite similar to that of a linear reactive filter, except that the two fundamental building blocks are modulo-two adders and idealized unit delays (or "escapement" circuits).

Many canonic forms have been found for the synthesis of such networks. For example, ladder networks associated with extended fraction forms of the transfer ratio

Z/X, . . . , or series networks associated with the partial fraction expansion of Z/X.

Of interest is a realization that utilizes as few unit delays as possible; in the example of Eq. 3, at least eight delays are necessary. A realization of this transfer ratio with just eight delay circuits is shown in Fig. IX-8. This circuit was obtained by a synthesis method that uses a simple iterated reduction procedure.

If the circuit of Fig. IX-8 is rearranged in the manner shown in Fig. IX-9 it becomes evident that it is in the form of a sequential switching circuit that is composed of a function generator and feedback loops with associated delays; however, the combinational network in the case shown here has outputs, Z, Y_1, Y_2, . . . , which are linear functions of the inputs x, y_1, y_2, . . . , rather than the more general nonlinear generator.

In a forthcoming technical report a thorough discussion of the properties of linear binary sequence filters will be presented. Some of the topics will be: physical realizability, characteristics of circuits inverse to a given circuit, natural frequencies, impulse response, steady-state and transient behavior, factorability of polynomials in D, canonic forms for synthesis, multiterminal filters, and treatment by matrix techniques.

D. A. Huffman

F. PROCEDURE FOR THE MINIMIZATION OF SWITCHING FUNCTIONS

Usually, the first step in designing a combinational switching circuit is to write the desired transmission function as a sum of product terms, using as few terms as possible, and then to factor. The form before factorization is called a minimum sum. Several procedures for determining minimum sums have been presented (1, 2, 3). These procedures are all incomplete, since, in effect, they merely present a list of product terms (called prime implicants by Quine) that can be used in forming a minimum sum. From this list it is necessary to select those prime implicants that must occur in the minimum sum. For many functions, this selection is straightforward; for others, a prohibitively large number of alternate selections must be considered before a minimum sum is obtained.

A method has been discovered that forms a minimum sum from a list of prime implicants without resorting to a lengthy enumeration procedure. A matrix is formed in which each column corresponds to a row of the table of combinations for which the function is to have the value "one," and each row corresponds to a prime implicant. If a prime implicant has the value one for a row of the table of combinations, a "1" is entered at the intersection of the corresponding row and column of the matrix. All other spaces are left blank. (This matrix is similar to Quine's Table of Prime Implicants.) A first "guess" is now made at a minimum sum and the rows corresponding to the terms of the "guess" placed at the top of the matrix. The matrix is partitioned into two submatrices: one contains the rows corresponding to the "guess"; the other, the remaining rows. Each column of the "guess" submatrix must have at least one entry, and those columns that contain only one entry should have that entry circled.

It is now necessary to consider the possibility of replacing two rows of the "guess" submatrix by one of the remaining rows, three rows by two rows, and so on. (It is necessary to maintain at least one entry in each column.) Rules have been developed which, through simplifying these considerations, lead quickly to a modified "guess" submatrix having the fewest possible number of rows. A minimum sum is formed directly by using the prime implicants corresponding to the rows of this matrix.

The details of this procedure for obtaining the minimum sum from the prime implicants, as well as a speedy method for getting the list of prime implicants, will be presented in a forthcoming technical report.

E. J. McCluskey, Jr.

References

1. Synthesis of Electronic Computing and Control Circuits, Staff of the Computation Laboratory of Harvard University (Harvard University Press, Cambridge, Mass., 1951).
2. W. V. Quine, Amer. Math. Mon. 59, No. 8, 521-531 (Oct. 1952).
3. M. Karnaugh, Trans. A.I.E.E. 72, Part 1, Commun. and Electronics, 593-599 (Nov. 1953).

G. A THEOREM ON FEEDBACK IN SEQUENTIAL SWITCHING CIRCUITS

It will be proved here that for a large class of sequential switching circuits there is a separate feedback loop associated with each nonsuperfluous secondary relay (or equivalent device). A more precise statement appears later in the report.

A sequencer is defined as a device with these characteristics:

1. The input and the output are voltages, or some other physical quantity, with magnitudes E and R, respectively. A binary variable Y, called the excitation, is associated with E, taking on the values 1 or 0, depending on whether E does or does not exceed some specified level. A two-valued variable y, called the response, is similarly related to R. In the steady state, $y = Y$.

2. The device can amplify whatever physical quantity it operates on. A relay, for example, can amplify conductance. If we consider the input E to be the conductance applied between ground and one terminal of the relay coil (the other terminal being at a suitable potential above ground), then, when the relay is being held in the operated state by one of its own contacts in series with one or more other contacts, it is evident that E is less than the conductance of the contact, which is the output R.

3. The sequencer exhibits an inertial effect in responding to changes in excitation. In other words, its output will not change unless the excitation change persists for some

minimum time interval.

4. The output R is distinctly two-valued and when it changes, it changes sharply from one of its permissible values to the other, regardless of the shape of the input function. This is the property of requantization.

It has been shown (1) that any sequential problem can be physically realized with devices having these properties and elements that can realize combinational functions. Many systems include sequencers that need not possess all of the listed properties. For example, property 3 is unnecessary if there are no hazards.

Typical sequencers are electromechanical relays and Schmitt trigger circuits (1, 2). Since Eccles-Jordan flip-flops can be shown to be built up from sequencers they need not be considered separately.

Flow tables can also be realized with inherently two-state devices such as magnetic cores. Although the terminal behavior of a magnetic core or other type of toggle – a device whose present output is a function of the last input (3) – can be simulated with the aid of sequencers, the question is still unresolved of whether or not it is possible to select internal physical parameters which might somehow indicate the actual existence of sequencers within all such devices. In this report only systems that use sequencers will be considered.

In what follows, it will be assumed that the reader is familiar with the concepts and terminology developed in reference 1.

THEOREM: If a sequential circuit with n sequencers operates without cycles or critical race conditions, and if the outputs during unstable states are never required to be different from the outputs during the succeeding stable states, and if the input is never changed while the system is in an unstable state, then any sequencer S_i whose excitation Y_i is independent of its response y_i may be eliminated from the system without altering the pertinent external system behavior.

Thus, under the stated hypothesis, each essential sequencer is contained in a feedback loop that does not pass through any other sequencers.

Proof: Consider any two rows of the system's Y-matrix which differ only in the values of y_i. For any particular input column, the Y_i entries will be the same for both rows, since Y_i is, by hypothesis, independent of y_i. Designate each such pair of rows (the 2^n rows are divided into 2^{n-1} such pairs) by an index k, and for each input column j denote by a_{jk} the state in the k^{th} row pair for which Y_i is stable. Call the corresponding state for which Y_i is unstable "b_{jk}."

Now assume that S_i responds much faster than any of the other sequencers, so that it wins all races. This assumption leaves invariant the over-all system operation, since all races are noncritical. It now follows that whenever the system enters a state b_{jk}, a transition to a_{jk} will occur, regardless of the entries in the b_{jk} position in the Y-matrix for the sequencers other than S_i. This must occur because S_i is unstable in state b_{jk} and will therefore change its response before anything else can happen.

We can therefore replace all Y entries in the b_{jk} position in the Y-matrix, except the one pertaining to S_i, with arbitrary values. In particular, let us replace them with the corresponding values from a_{jk} so that in states a_{jk} and b_{jk} the Y values are the same. This process is carried out for all j's and all k's, so that we generate a new Y-matrix that describes a system operationally identical with the original one (with the fast S_i), although all the sequencer excitations are now independent of y_i.

Since all b_{jk} states are unstable, we can replace all entries Z in the b_{jk} positions of the output matrix with the corresponding entries for the a_{jk} positions without altering the outputs during stable states and without introducing false outputs during unstable states. This transformation makes the Z's independent of the state of S_i.

But if all Y's and Z's are independent of y_i, we can eliminate S_i by merging the members of each of the row pairs k. The only effect of this step is to eliminate meaningless transitions from b-states to a-states. The operation of the resulting circuit is otherwise identical with that of a circuit with the original Y and Z matrices and the fast-moving S_i. Since the latter is, as was previously pointed out, essentially equivalent to a circuit whose sequencers operate at arbitrary speeds, the reduced circuit must also be equivalent to the original one. This proves the theorem.

It is not difficult to find counterexamples that show that the conclusion is not generally true if the input is permitted to change during unstable states, or if special momentary output changes are required during unstable states.

Although the idea that sequential circuits entail feedback around various circuit elements is certainly not new (for example, see ref. 4) the fact that the excitation of each sequencer must be, in part, directly dependent on its own response, without the intervention of other sequencers, has not, to my knowledge, been previously established.

S. H. Unger

References

1. D. A. Huffman, Jour. Franklin Inst. 257, 161-190 (March 1954); 257, 275-303 (April 1954).

2. W. C. Elmore and M. Sands, Electronics (McGraw-Hill Book Company, Inc., New York, 1954).

3. T. H. Meisling, Report, Division of Electrical Engineering, University of California, Dec. 1952.

4. J. von Neumann, Lecture Notes, California Institute of Technology (Jan. 1952) p. 12.