XI. PROCESSING AND TRANSMISSION OF INFORMATION[*]

Prof. E. Arthurs
Prof. P. Elias
Prof. R. M. Fano
Prof. J. Granlund
Prof. D. A. Huffman
Prof. R. C. Jeffrey
Prof. H. Rogers, Jr.
Prof. C. E. Shannon
Prof. W. M. Siebert
Prof. J. M. Wozencraft
Prof. W. A. Youngblood
D. C. Coll
J. E. Cunningham
J. B. Dennis
H. A. Ernst
E. F. Ferretti
R. G. Gallager
T. L. Grettenberg
F. C. Hennie III
E. M. Hofstetter
M. Horstein
T. S. Huang
F. Jelinek
T. Kailath
R. A. Kirsch
G. E. Mongold, Jr.
B. Reiffen
L. G. Roberts
W. L. Wells
H. L. Yudkin
H. P. Zeiger

A. PICTURE PROCESSING

The modified facsimile system for scanning pictures and producing digitalized samples along scan lines, described in an earlier report (1), has been used for further study of picture coding methods.

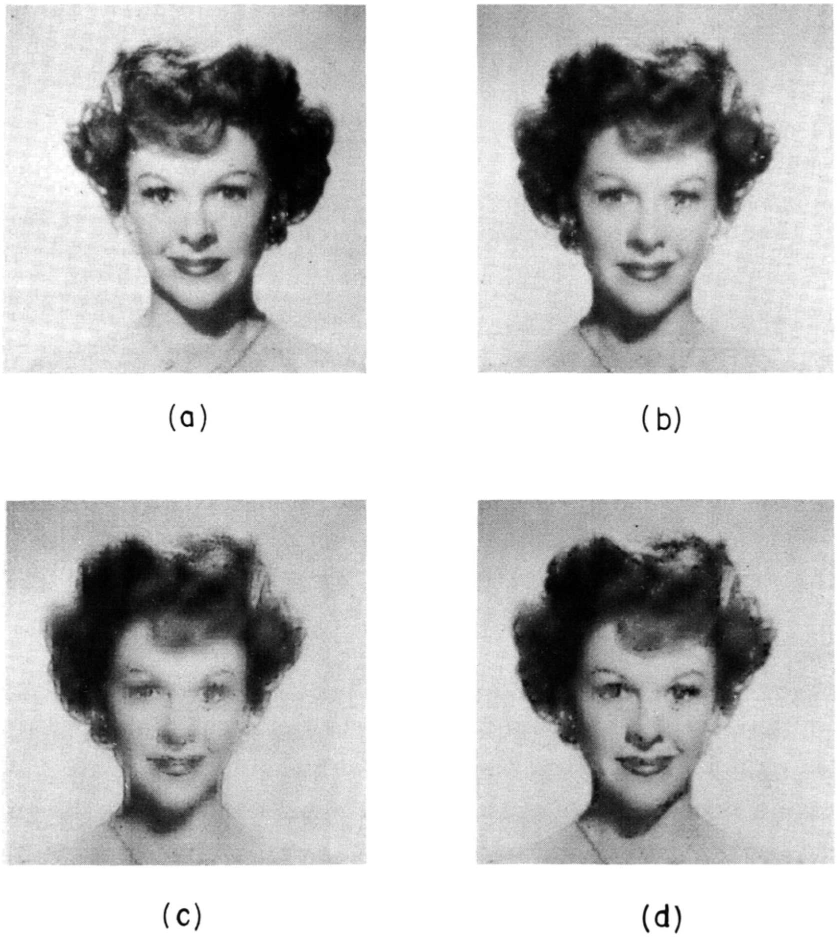
The TX-0 computer has been programmed to process the picture data and produce pictures on the display oscilloscope (2). The programs (3) average the intensity level of blocks of data, which are either 3 samples by 3 samples or 5 × 5. This averaging operation produces a lowpass picture that need be specified only at the centers of the blocks; intermediate values of intensity are obtained by algebraic interpolation of the surrounding values.

In addition, the program compares the values of the lowpass samples with the corresponding actual sample values. If the magnitude of the difference exceeds a preset criterion, a fixed value of correction signal is added to the lowpass signal; if a second criterion is exceeded, a larger fixed value is added; the sign of the correction signal is determined by the sign of the difference. A value of 0 is added when the first criterion is not exceeded; thus the correction alphabet consists of 5 symbols. The magnitude of the correction signals are chosen to be 2 times the first criterion and 1.5 times the second criterion.

The original picture samples were 6-bit binary numbers, representing the 64 allowable intensity values; thus 6 bits per sample is required to specify the picture in this form. To specify the lowpass picture, only 1 sample per block is required, thus the source-rate estimate on the basis of this type of coding process is the sum of the information required to represent the lowpass picture plus the entropy $(\Sigma - p_i \log_2 p_i)$ of the correction-signal alphabet. The p_i of the correction alphabet is obtained by counting the number of each type of correction signal sent.

In Fig. XI-1 some representative results of the study are shown. (Proper viewing

[*]This research was supported in part by Purchase Order DDL-B222 with Lincoln Laboratory, a center for research operated by M.I.T., which is supported by the U.S. Air Force under Air Force Contract AF19(604)-5200.



(a)

(b)

(c)

(d)

Fig. XI-1. Pictures reproduced by computer display.

distance is approximately 25 inches because of the size of the scanning aperture.) Figure XI-1a is a reproduction in which the full 6 bits per sample are used from the data tape. Figure XI-1b is a processed version of a 3 × 3 averaging block, with criteria of 5 and 10; the source rate estimate is 0.99 bits per sample. Figure XI-1c is the same picture with processing by a 5 × 5 scanning block, with criteria of 5 and 10; the source rate estimate is 0.63 bits per sample. Figure XI-1d is processed with a 5 × 5 scanning block, but with the error-correction signal determined by difference of interpolated and actual values rather than by difference of average and actual values, as in the two previous examples; the source rate estimate for criteria of 5 and 10 is 1.05 bits per sample.

This two-dimensional processing avoids some of the objectionable artifacts found in the usual line-by-line processing (1) because distortions are not so highly correlated.

J. E. Cunningham

References

1. W. A. Youngblood, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., Jan. 15, 1958, pp. 95-100.

2. J. E. Cunningham, Quarterly Progress Report No. 53, Research Laboratory of Electronics, M.I.T., April 15, 1959, pp. 113-114.

3. J. E. Cunningham, A study of some methods of picture coding, S.M. Thesis, Department of Electrical Engineering, M.I.T., June 1959.

B. SAMPLING THEOREMS FOR LINEAR TIME-VARIANT SYSTEMS

This work, which has been completed, will be presented in Technical Report 352, "Sampling Models for Linear Time-Variant Filters."

T. Kailath

C. PARITY-CHECK CODES WITH LIMITED CONSTRAINTS PER DIGIT

1. Introduction

This report considers a class of parity-check codes with arbitrarily long block length but with a fixed number of constraints on each digit and a fixed rate. It is shown that these codes have a probability of error that decreases exponentially with block length, but at a slightly slower rate than the exponent for optimum codes. A decoding scheme for these codes that is not optimum, but for which the computation per digit appears to be independent of block length, is described.

When signals are transmitted through noisy channels, it is desirable to add a certain amount of redundancy to the signal which enables it to be correctly reproduced with high probability at the receiver. With binary symmetric channels, parity-check codes provide a simple means of adding the necessary redundancy. The code words of a parity check code are formed by transmitting a certain number of digits of the message followed by a number of redundant digits. Each redundant digit is the modulo 2 sum of a prespecified set of the message digits. We shall call the redundant digit plus its set of

$$\begin{array}{c|ccccccc|}
 & \multicolumn{7}{c}{n} \\
 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ \hline
 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\
n(1-R) & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\
 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ \hline
\end{array}$$

$$x_5 = x_1 \oplus x_2 \oplus x_3$$
$$x_6 = x_1 \oplus x_2 \oplus x_4$$
$$x_7 = x_1 \oplus x_3 \oplus x_4$$

Fig. XI-2. Example of parity-check matrix.

message digits a parity-check set. A particular parity-check code may be specified by a matrix such as Fig. XI-2. Each row in Fig. XI-2 represents a parity-check constraint, and the positions of the 1's in a row represent a parity-check set. To decrease the error probability of a code, either the block length must be increased, or the rate, which is the ratio of message digits to block length, must be decreased. In fact, Elias (1) has shown that the probability of decoding error decreases exponentially with block length for a code whose check-digit matrix is randomly chosen and whose rate is fixed at less than channel capacity.

Although increasing the length of a parity check code increases the reliability, it also complicates the decoding problem. The usual decoding scheme for a parity-check code is as follows: First, compute the parity checks. These are defined as the modulo 2 sums of the digits in the associated parity-check sets. Second, construct a code book associating each sequence of parity checks with the most likely error sequence. Unfortunately, the size of this code book grows exponentially with the number of parity checks. The object of the work described in this paper is to achieve the very small error probabilities that are possible with long codes and to avoid the excessive computation associated with code-book decoding.

2. Codes with j Constraints Per Digit

Consider an n by n(1-R) parity-check matrix in which each of the n columns contains a small fixed number, j, of 1's and contains 0's elsewhere. Note that the last n(1-R) columns are no longer diagonalized and the code can no longer be thought of as separated into "message" digits and redundant digits. We shall show that with any constant number, $j \geq 3$, of 1's per column and constant rate, there exist codes whose decoding error probabilities decrease exponentially with increasing block length. Thus the probability of decoding error can be made as small as desired while the number of parity-check sets containing any particular digit is held fixed, and consequently the average size of each parity-check set is held fixed.

One immediate advantage of these codes is that with such small check sets, each parity check provides much more information about the individual digits in the parity-check set. Thus, although the total information in a check digit is no greater than it was before, it can be more easily used and interpreted. This suggests a variety of computationally efficient decoding schemes, one of which has already been tested with promising results on the IBM 704 computer.

Finally, coding is no more difficult for these codes than for ordinary parity-check codes because we can find an equivalent representation for each of these codes in ordinary diagonalized parity-check-code form. Typically, each message digit will then be contained in approximately one half of the parity-check sets.

3. Minimum Distance

An expression will be derived for the probability of decoding error for codes with j constraints per digit by taking an average over the whole ensemble of codes. The method is similar to that used by Elias (1) for ordinary parity-check codes, except that with only j constraints per digit, the "bad" codes dominate the average probability of error expression. However, these "bad" codes can be removed from the ensemble before computing the average probability of error over the rest of the ensemble. Actually, minimum distance will be used as a criterion for which codes to remove; this does not correspond exactly to probability of error, but it is close enough for our purposes. The minimum distance for one of these codes is the smallest number of 1's in any nonzero code word. We must first precisely define the ensemble of codes with j constraints per digit, and then calculate the probability that a code drawn from this ensemble will have a certain minimum distance. For mathematical simplicity, the ensemble will actually allow j or fewer 1's per column, but this is of no practical importance.

The ensemble of codes of block length n with n(1-R) parity-check sets and j constraints per digit is defined in terms of an n by n(1-R) parity-check matrix. Consider putting j distinguishable 1's into each column, where each 1 has a probability $\frac{1}{n(1-R)}$ of being in each row. If more than one 1 occurs in some position of the matrix, then their modulo 2 sum is placed in that position. This defines our ensemble of codes.

The probability, P(d ≤ m), that a code drawn from this ensemble will have a minimum distance of m or less is simply the probability that some sequence containing m or fewer 1's is a code word. Since the probability of a union of events is less than the sum of the probability of the individual events,

$$P(d \leq m) \leq \sum_{\ell=1}^{m} \binom{n}{\ell} P_{\ell}$$

where P_{ℓ} is the probability that a particular word containing ℓ 1's is a code word. But a word with ℓ 1's is a code word if, and only if, those ℓ columns in the matrix have an even total number of 1's in each row. Altogether, there are ℓj 1's in these ℓ columns, each of which has a probability $\frac{1}{n(1-R)}$ of being in each row. The probability of an even number of 1's in each row can be found from the following theorem to be

$$P_{\ell} = 2^{-n(1-R)} \sum_{i=0}^{n(1-R)} \binom{n(1-R)}{i} \left[1 - \frac{2i}{n(1-R)}\right]^{\ell j} \tag{1}$$

$$P(d \leq m) \leq 2^{-n(1-R)} \sum_{\ell=1}^{m} \sum_{i=0}^{n(1-R)} \binom{n}{\ell} \binom{n(1-R)}{i} \left[1 - \frac{2i}{n(1-R)}\right]^{\ell j} \tag{2}$$

THEOREM: Consider an experiment with A equally likely outcomes that is performed independently B times. Then the probability, P(even), that each outcome occurs an even number of times is

$$P(\text{even}) = 2^{-A} \sum_{i=0}^{A} \binom{A}{i} \left(1 - \frac{2i}{A}\right)^{B}$$

PROOF: If B is odd, at least one outcome must occur an odd number of times, and P(even) = 0. The theorem is correct in this case because the A-i term cancels the i term for every i in the summation. For B even, we proceed by induction on A. For A = 1, the theorem is correct by inspection.

Assume that the theorem is correct for A-1. Let P_j be the probability that the first outcome occurs j times, and let $P_j(\text{even})$ be the probability that each of the other outcomes occurs an even number of times, given that the first outcome appeared j times. Thus, by the binomial theorem,

$$P_j = \binom{B}{j} \left(\frac{1}{A}\right)^{j} \left(\frac{A-1}{A}\right)^{B-j}$$

and, by inductive assumption,

$$P_j(\text{even}) = 2^{-A+1} \sum_{i=0}^{A-1} \binom{A-1}{i} \left[1 - \frac{2i}{A-1}\right]^{B-j}$$

$$P(\text{even}) = \sum_{\substack{j=0\\ j \text{ even}}}^{B} P_j\, P_j(\text{even}) = \sum_{\substack{j=0\\ j \text{ even}}}^{B} \binom{B}{j}\left(\frac{1}{A}\right)^{j} \left(\frac{A-1}{A}\right)^{B-j} 2^{-A+1} \sum_{i=0}^{A-1} \binom{A-1}{i} \left[1 - \frac{2i}{A-1}\right]^{B-j}$$

$$= 2^{-A+1} \sum_{i=0}^{A-1} \binom{A-1}{i} \sum_{\substack{j=0\\ j \text{ even}}}^{B} \binom{B}{j} \left(\frac{1}{A}\right)^{j} \left[\frac{A-1}{A}\left(1 - \frac{2i}{A-1}\right)\right]^{B-j}$$

$$P(\text{even}) = 2^{-A} \sum_{i=0}^{A-1} \binom{A-1}{i} \left[\frac{A-1}{A}\left(1 - \frac{2i}{A-1}\right) + \frac{1}{A}\right]^{B} + 2^{-A} \sum_{i=0}^{A-1} \binom{A-1}{i} \left[\frac{A-1}{A}\left(1 - \frac{2i}{A-1}\right) - \frac{1}{A}\right]^{B} \tag{3}$$

Equation 3 can be verified by a binomial expansion of both of its terms and by canceling the odd terms. Now, in the second term, substitute i-1 for i to obtain

$$P(\text{even}) = 2^{-A} \sum_{i=0}^{A-1} \binom{A-1}{i} \left[1 - \frac{2i}{A}\right]^{B} + 2^{-A} \sum_{i=1}^{A} \binom{A-1}{i-1} \left[\frac{A-1}{A}\left(1 - \frac{2i}{A} + \frac{2}{A}\right) - \frac{1}{A}\right]^{B}$$

$$= 2^{-A} \sum_{i=0}^{A-1} \binom{A-1}{i} \left[1 - \frac{2i}{A}\right]^{B} + 2^{-A} \sum_{i=1}^{A} \binom{A-1}{i-1} \left[1 - \frac{2i}{A}\right]^{B}$$

Finally, by using the equality

$$\binom{A}{i} = \binom{A-1}{i} + \binom{A-1}{i-1}$$

we obtain

$$P(\text{even}) = 2^{-A} \sum_{i=0}^{A} \binom{A}{i} \left[1 - \frac{2i}{A}\right]^B$$

and the theorem is proved.

Equation 2 can be shown to be bounded by

$$P(d \leqslant m) \leqslant \alpha n^2 e^{\beta n} + \begin{cases} \gamma n^{-j/2+1} & j \text{ even} \\ \gamma n^{-j+2} & j \text{ odd} \end{cases}$$

where α, β, γ are functions of j, m/n, and R but not of n.

For fixed rate R and any fixed $j \geqslant 3$, β can be shown to be a strictly increasing function of m/n. If we define $\mu(j, R)$ as that value of m/n for which $\beta = 0$, then we see that for any fixed ratio, $m/n < \mu$, we have

$$\lim_{n \to \infty} P(d \leqslant m) = 0$$

Thus it is reasonable to think of μn as being the typical minimum distance of this ensemble of codes. Also, for any $\epsilon > 0$ and n sufficiently large, we can remove all codes from the ensemble with minimum distance less than $(\mu-\epsilon)n$ and still have most of the ensemble left. Figure XI-3 compares the typical minimum distance for these codes with those of ordinary parity-check codes. It can be seen that these codes quickly

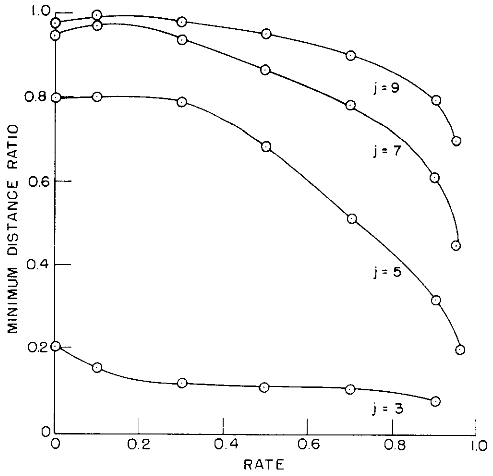


Fig. XI-3. Ratio of typical minimum distance for j constraint code to typical minimum distance for ordinary parity-check code for long block length.

approach the behavior of ordinary parity-check codes for quite small values of j.

For j = 2, it can be shown by another argument that the minimum distance of a code is bounded above by log n times a function of rate, and so we exclude codes with j = 2.

4. Probability of Error with Optimum Decoding

For a binary symmetric channel with transition probability p, the probability of k errors in a block of n digits is

$$\binom{n}{k} p^k (1-p)^{n-k}$$

Thus the ensemble average probability of decoding error, P(e), is given by

$$P(e) = \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} P_k(e) \tag{4}$$

where $P_k(e)$ is the ensemble average probability of decoding error, given k errors in transmission. If we use optimum code-book decoding, a decoding error can only occur when the actual error sequence has the same parity-check sequence as another error sequence with fewer or the same number of 1's. But if two error sequences have identical parity checks, then their modulo 2 sum must be a code word. It can be shown that for ℓ even, there are

$$\sum_{i=\ell/2}^{k} \binom{k}{i} \binom{n-k}{\ell-i}$$

sequences containing k or less 1's which, when they are added modulo 2 to a particular sequence containing k 1's, produce a sequence containing ℓ 1's. Over the whole ensemble, each of these produces a decoding error whenever the sum is a code word. This is an event of probability P_ℓ as given in Eq. 1. The fraction of codes, δ, with minimum distance m, or less, is given by $P(d \leq m)$. After these codes are removed, the probability becomes

$$P'_\ell = 0 \qquad \ell \leq m$$

$$P'_\ell \leq \frac{2^{-n(1-R)}}{1-\delta} \sum_{i=0}^{n(1-R)} \binom{n(1-R)}{i} \left[1 - \frac{2i}{n(1-R)}\right]^{\ell j} \qquad \ell > m$$

where the $1/(1-\delta)$ results from the increase in probability of each remaining code. It can be seen from Eq. 1 that if ℓ and j are odd, then $P_\ell = 0$. This gives us, finally, for j odd

$$P_k(e) \leq \frac{2^{-n(1-R)}}{1-\delta} \sum_{\frac{\ell}{2}=\frac{m}{2}}^{k} \left\{ \sum_{i=\frac{\ell}{2}}^{k} \binom{k}{i} \binom{n-k}{\ell-i} \right\} \left\{ \sum_{i=0}^{n(1-R)} \binom{n(1-R)}{i} \left[1 - \frac{2i}{n(1-R)}\right]^{\ell j} \right\} \tag{5}$$

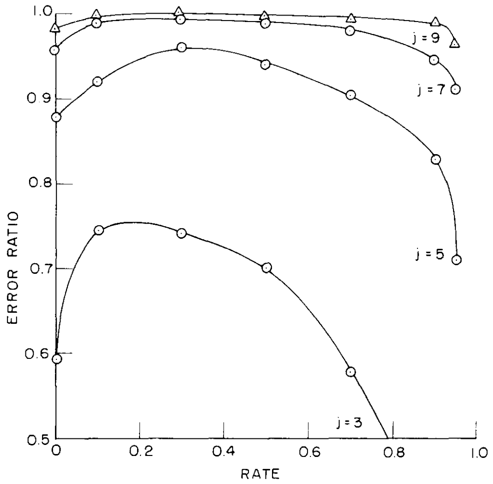


Fig. XI-4. Ratio between j constraint code and ordinary parity-check code of the number of errors that can be corrected with high probability for long block length.

For j even, the result is similar but less neat. Define k_o as the smallest value of k for which the right-hand side of Eq. 5 is greater than or equal to the integer 1. The parameter k_o is very important because it represents the error-correcting breakpoint of the code. For large n, if the fraction of transmission errors is less than k_o/n, we expect to be able to decode with high probability, and if it is greater than k_o/n, we expect to make a decoding error. Combining Eqs. 4 and 5, we obtain the general expression for probability of error for codes with j constraints per digit, j odd,

$$P(e) \leqslant \frac{2^{-n(1-R)}}{1-\delta} \sum_{k=1}^{k_o-1} \sum_{\frac{\ell}{2}=\frac{m}{2}}^{k} \binom{n}{k} p^k (1-p)^{n-k} \left\{ \sum_{i=\frac{\ell}{2}}^{k} \binom{k}{i} \binom{n-k}{\ell-i} \right\}$$

$$\times \left\{ \sum_{i=0}^{n(1-R)} \binom{n(1-R)}{i} \left[1 - \frac{2i}{n(1-R)}\right]^{\ell j} \right\} + \sum_{k=k_o}^{n} \binom{n}{k} p^k (1-p)^{n-k} \qquad (6)$$

No simple general approximation has yet been found for this, except for sufficiently high channel transition probabilities, in particular, when

$$\frac{k_o}{n-k_o} > \frac{p}{1-p} > \frac{k_o - \frac{m}{2}}{n - k_o - \frac{m}{2}}$$

In that case, it can be shown that

$$P_e \leqslant \frac{A}{\sqrt{n}} e^{-n[T_p(k_o/n)-H(k_o/n)]} \qquad (7)$$

where A is independent of n and

$$T_p\left(\frac{k_o}{n}\right) = -\frac{k_o}{n}\ell n\, p - \left(1 - \frac{k_o}{n}\right)\ell n\, (1-p)$$

$$H\left(\frac{k_o}{n}\right) = -\frac{k_o}{n}\ell n\, \frac{k_o}{n} - \left(1 - \frac{k_o}{n}\right)\ell n\, \left(1 - \frac{k_o}{n}\right)$$

This is the same as the expression for ordinary parity-check codes, except that k_o/n is different. Figure XI-4 compares k_o for these codes with k_o for ordinary parity-check codes.

5. Decoding Schemes

The simplest decoding scheme applicable to these codes is to: Change the digits for which all or most of the associated parity checks are 1; then recompute the parity checks, change more digits, and so forth. If the number of errors is sufficiently small, then after several repetitions all the parity checks will be 0 and the sequence is decoded. This scheme works rather poorly for any appreciable number of errors because the j parity checks associated with a digit do not furnish enough information about that digit to warrant any decision. These parity checks would furnish more information if we knew more about the other digits in the parity check sets.

One way to handle this situation is, by using, first, only the parity checks associated with a digit, to estimate the probability that that digit is correct. Then the estimation of each digit can be refined by using not only the parity checks associated with that digit but also the previous estimate of the other digits in those parity-check sets. This procedure can be repeated as often as desired, and if the number of errors is not too great, the estimates should converge to indicate which digits are incorrect.

The amount of computation per digit per repetition in schemes of this kind is clearly independent of block length. It appears, although proof is still lacking, that the average number of repetitions necessary to decode is bounded by a quantity independent of block length. If this is true, the average computation per digit is also bounded. Naturally, the storage capacity and delay necessary for decoding increase linearly with block length.

A decoding scheme based on these principles has been partially tested on the IBM 704 computer for a code of block length 512, rate 1/2, and $j = 5$. The program seems to decode well up to, say, 30 errors, and this might be improved with some modifications. When optimum decoding with the same code is used, the breakpoint (from Eq. 5) is approximately 53 errors, and with an optimum code of the same rate and length, the breakpoint would be approximately 56 errors. Such a scheme can also be used to decode from binary input, multioutput channels, such as binary signals disturbed by Gaussian noise. This provides a way to avoid the information loss that accompanies the making of a binary decision on each digit before decoding. No definite results are now available for these more general channels.

R. G. Gallager

References

1. P. Elias, Coding for noisy channels, IRE Convention Record, Part 4, 1955, pp. 37-46.

D. SWITCHING CIRCUITS WITH MULTIPLE-HOLE MAGNETIC CORES

Magnetic cores have long been used in memory arrays. However, they have found limited use in logic circuits, for two reasons: First, in all core circuits diodes had to be used to prevent backward flow of information. Since diodes are much less reliable than cores, the inherent reliability of cores goes to waste in circuits of this type. Second, design techniques have been too complicated for quick acceptance in the field, and circuit operation is too involved for easy maintenance.

Recent work in the field of multiple-hole cores has added a new degree of freedom to core circuit design. In particular, we can now eliminate diodes from core logic circuits.

The purpose of this research is to produce a design method for the realization of general switching functions of many variables by means of circuits that use only magnetic cores and connecting wire, and are simple and uncritical in their operation.

The first system that we developed operated on a continuous ac carrier. At any point in the logic circuit the presence of this carrier indicates a "1", its absence a "0" . This method permitted the realization of any switching function of three variables through the use of three toroidal cores, and four magnetic elements shaped like three-rung ladders. This system is simple and uncritical in its operation, but suffers from two major drawbacks: (a) The loss in signal level across a logic circuit necessitates the use of transistor amplifiers in any large system. (b) Rate of operation is slow. A carrier frequency of 50 kc has been used, and frequencies higher than approximately 200 kc seem unlikely to be realized with existing cores. Nevertheless, the extreme simplicity, reliability, and compactness of this system may make it useful for certain special applications.

Work continues on a method to combine the advantages of the system that has been described with the inherent speed and gain of pulse-operated circuits.

H. P. Zeiger