XIX. PROCESSING AND TRANSMISSION OF INFORMATION[*]

Prof. R. G. Gallager
Prof. F. C. Hennie III
Prof. R. S. Kennedy
Prof. C. E. Shannon
Prof. J. M. Wozencraft
Dr. R. E. Kahn
D. Chase
J. R. Colton
P. M. Ebert
D. D. Falconer
E. F. Ferretti
G. D. Forney, Jr.
C. J. Johnson
J. Max
R. F. McCann
C. W. Niessen
C. W. Niessen
R. E. Olsen
R. Pilc
J. T. Pinkston III
J. E. Savage
R. N. Spann
W. R. Sutherland
M. G. Taylor

RESEARCH OBJECTIVES AND SUMMARY OF RESEARCH

1. Communications

The work of the group is focused on the dual problems of ascertaining the best performance that can be attained with a communication system, and developing efficient techniques for actually achieving performance substantially this good.

a. Bounds on Error Probability

Bounds on the minimum error probability achievable through coding as a function of channel, transmission rate and code constraint length are important for at least two reasons: they provide a standard against which practical coding schemes can be compared, and they contribute insight towards the design of new coding schemes and associated digitial modems.

A large number of new error probability bounds have been contributed over the past year, including upper bounds for Gaussian noise channels, general time discrete amplitude continuous channels,[1] discrete channels with memory, and networks of channels. New lower bounds for discrete memoryless channels, and a number of upper and lower bounds for block coding with instantaneous feedback on discrete memoryless channels,[2] have also been developed.

Work is continuing on error probability bounds for networks of channels, channels with memory, and continuous channels. Finally, work is starting on developing bounds for restricted classes of codes on discrete memoryless channels.

References

1. R. G. Gallager, A simple derivation of the coding theorem and some applications, IEEE Trans. on Information Theory (in press).

2. E. Berlekamp, Block Coding with Noiseless Feedback, Ph. D. Thesis, Department of Electrical Engineering, M. I. T., August 1964.

b. Instrumentation

The problem of coded system design breaks naturally into questions of coding-decoding and modulation-demodulation. The inter-relation between these two aspects

[*] This work was supported in part by the National Science Foundation (Grant GP-2495), the National Institutes of Health (Grant MH-04737-04), and the National Aeronautics and Space Administration (Grants NsG-334 and NsG-496).

has been clarified in recent work[1,2] and continues as an active subject of theoretical and experimental investigation. In particular, a data-acquisition system has been designed and procured which will permit a variety of sequential decoding algorithms to be tested on signals produced by actual communication systems. For example, the bubble-tank facility provides a convenient time-and-frequency dispersive channel for the study of adaptive decoders that estimate propagation characteristics as well as the transmitted message.[3] The decoding experiments will be performed on the Project MAC PDP-6 computer; a special-purpose compiler language will be developed for use in these experiments.

Equipment complexity and computational speed are the central problems in decoder implementation. A theoretical study of the computational requirements with sequential decoding is discussed in the present report. An improved method of decoding Bose-Chaudhuri-Hocquengem codes is also reported.

References

1. J. M. Wozencraft and R. S. Kennedy, Modulation and demodulation for probabilistic coding (submitted to IEEE Trans. (PTGIT)).

2. J. L. Holsinger, Digital Communication over Fixed Time Continuous Channels with Memory – With Special Application to Telephone Channels, Sc. D. Thesis, Department of Electrical Engineering, M. I. T., October 1964, also Technical Report 430, Research Laboratory of Electronics, M. I. T. (forthcoming).

3. H. L. Yudkin, Channel State Testing in Decoding of Information, Sc. D. Thesis, Department of Electrical Engineering, M. I. T., September 1964.

c. Vocoded Speech

The proper role of coding in conjunction with speech communication is not clear. A preliminary investigation of the deterioration of vocoder intelligibility as a function of noise statistics should be completed by June 1965. Concurrently, an attempt will be made to devise an efficient coding-decoding scheme appropriately matched to listener tolerance. The vocoder is being simulated on the 7094 computer.

d. Optical Communications

A theoretical and experimental study directed toward extending modern communication theory to the treatment of optical communication systems has been initiated. One aspect of the investigation is devoted to the statistical description of the channels that might be used at optical frequencies. These descriptions are sought so that the performance limitations of the channels can be determined. Another aspect is the determination of practical means by which these performance limits may be approached. This determination involves the statistical description of devices that might be employed in optical systems. A collateral phase of the investigation is the development of an equipment facility for experimental studies.

The equipment facility now consists of a 5-mw Helium-Neon gas laser, a propagation path which simulates a long-range free space channel, and a photomultiplier tube which serves as a receiver. The facility includes the electronic equipment necessary for processing the received data at energy levels for which quantum effects are important. The statistical characteristics of this system have been determined by analysis and measurement.[1]

The design of dielectric filters for optical communication systems has also been investigated.[2] The synthesis of a planar dielectric filter with a prescribed transfer characteristic was the central problem considered. This problem reduces to the determination of the required filter permittivity as a function of position. In most previous studies, only layered filters are considered. That is, the permittivity is constrained to be a staircase function of position. The more general nonlayered problem was treated through recourse to the mathematical techniques of inverse scattering theory.

R. G. Gallager, R. S. Kennedy, J. M. Wozencraft

References

1. G. D. Papadopoulis, Photon Channel Statistics, S. M. Thesis, Department of Electrical Engineering, M. I. T., June 1964.

2. J. C. Portinari, Application of the Inverse Scattering Problem to Optical Filter Design, S. M. Thesis, Department of Electrical Engineering, M. I. T., June 1964.

2. Digital Machines and Automata

During the past year, work has been concentrated in two areas: (i) the development of techniques for detecting and correcting faulty behavior in finite-state machines; (ii) the relationships between given information processing requirements and the amounts of time and equipment needed to implement them.

Two major advances have been made in the first area. First, a method has been developed[1] for determining experimentally whether a given sequential circuit is operating correctly, or whether it has suffered one of a certain broad class of malfunctions. Although this method does not necessarily lead to minimum-length experiments, it does yield relatively short experiments and is easy to apply. Second, progress has been made on the problem of providing error-detection capabilities through a combination of coding and redundant circuitry. Some of this work is described later in the present report.

In the second area, progress has been made on the problem of determining the amounts of time required to compute functions with Turing machines. The computations performed by a restricted class of Turing machines can now be described[2] by a generalization of the finite-internal-state concept, and this technique makes it possible to obtain good lower bounds on the computation times of such machines. In particular, it can be shown that for this class of Turing machines the reduction from two tapes to one can require a squaring of the computation time.

During the coming year, effort will be devoted to characterizing the class of modular threshold functions and applying these functions to the problem of error detection in combinational logic. Further investigation of computational complexity questions is anticipated, with the ultimate goal of developing more reasonable measures of complexity than that provided by Turing machine computation time. Finally, consideration will be given to the general problem of relating the structure of machines to the processing requirements that they must meet.

F. C. Hennie III

References

1. F. C. Hennie, Fault Detecting Experiments for Sequential Circuits, Proc. Fifth Symposium on Switching Circuit Theory and Logical Design, Princeton, New Jersey, November 1964.

2. F. C. Hennie, One-Tape, Off-Line Turing Machine Computations (to be published in Information and Control).

A. DECODING BOSE-CHAUDHURI-HOCQUENGHEM CODES

1. Introduction

The outstanding problem of coding theory is to devise codes for combatting disturbances that are typical of communication channels. To be useful a code must admit a decoding scheme that is practically realizable, a requirement that has been the severest impediment to use of complicated codes in operational communication systems.

The channel disturbance that has been most studied is that of uncorrelated, equiprobable errors. Substantial attempts have been made to apply algebraic theory to the construction of codes to combat such disturbances. The minimum distance of a code,[1] that is, the minimum number of places in which any two code words differ, has been a central concept and the primary criterion of goodness, for when a code has minimum distance d it is theoretically capable of correcting any number of errors t such that $2t \leq d - 1$.

The outstanding success of this algebraic approach has been the class of binary (2-symbol) codes discovered by Bose and Ray-Chaudhuri,[2,3] and independently by Hocquenghem.[4] This class contains many good codes with a wide range of lengths, rates, and minimum distances. Peterson[5,6] soon discovered an efficient decoding algorithm for such codes, which was suitable for implementation on a special-purpose digital computer. Bartee and Schneider[7] actually constructed such a machine.

Gorenstein and Zierler[8] succeeded in generalizing these codes to the nonbinary case; we shall henceforth call all codes of this type, binary or nonbinary, BCH codes. They were able to outline an efficient error-correction procedure for these codes, consisting essentially of three steps: $d - 1$ parity checks are formed from the received symbols; the locations of the errors are then found by manipulations with these parity checks; and since the parity checks are linear functions of the error values, the t unknown error values are given by the solution of t linear parity-check equations.

Errors occur in a communication channel when the detection apparatus at the receiver 'guesses' that a certain symbol was sent, when it was not. It has been recognized that the performance of a channel could be improved by allowing the detection apparatus not to guess at all whenever the evidence does not clearly indicate one symbol as the most probable.[9,10] The output of the detection apparatus in such an event is called an erasure, or deletion; we shall consider erasures as errors (possibly of value zero) whose locations are known.

A code of minimum distance d is theoretically capable of simultaneously correcting t errors and s erasures whenever $2t + s \leq d - 1$. A technique for correcting erasures and errors with binary BCH codes is mentioned by Peterson,[11] but this technique does not generalize to nonbinary codes.

In this report we resolve the general error-erasure correction problem by introducing a set of parity checks modified according to the positions of the erasures so as not to check the erased symbols. We show, in a generalization of Gorenstein and Zierler's work, that these 'modified cyclic parity checks' can be used to find the locations of the errors, and hence to correct simultaneous erasures and errors for general BCH codes.

In this development, as in Gorenstein and Zierler's, the rank of a certain matrix indicates the number of errors in a received word. As we try changing one symbol, therefore, the behavior of this rank can tell us whether or not that symbol is in error, and by how much, so that, in principle, a kind of step-by-step decoding becomes possible. We show,that, because of the special form of this matrix, explicit solutions for the values of erasures and errors can be obtained in a form that is attractive to implement. In particular, we demonstrate procedures for the correction of erasures with no errors which are simpler than the solution of s equations in s unknowns.

2. Preliminary Definitions

BCH codes are conveniently described in the language of the theory of finite fields, which has been well developed for this purpose by Peterson.[6] A finite, or Galois, field with p^M elements (written $GF(p^M)$) exists if p is a prime, and M is any integer; p is called the characteristic of the field. In any field there is a zero element 0, a unit element 1, and at least one primitive element α, such that any other nonzero element β can be expressed as a power of α. The order of β is the least integer e such that $\beta^e = 1$; a primitive element α has order $p^M - 1$. If M is a factor of N, the elements of $GF(p^M)$ are included in $GF(p^N)$, and the former is said to be a subfield of the latter, or the latter an extension field of the former.

In this language, code words of length n_o are represented by sequences of n_o elements from $GF(p^N)$, which we shall write as $\vec{f} = (f_1, f_2, \dots, f_{n_o})$. If we define the column vector of descending powers of X, $\vec{X}_{(a,b)} \equiv (X^a, X^{a-1}, \dots, X^b)^T$, where X is an indeterminate and T indicates the transpose, then the dot product $\vec{f} \cdot \vec{X}_{(n_o-1,0)}$ is a polynomial in X of degree $n_o - 1$, which we call $f(X)$. Similarly, if we define $\beta^{\vec{m}}_{(a,b)} \equiv (\beta^{ma}, \beta^{m(a-1)}, \dots, \beta^{mb})^T$, where β is any element of $GF(p^N)$ or of an extension or subfield $GF(p^M)$ thereof, we can define $f(\beta^m) \equiv \vec{f} \cdot \beta^{\vec{m}}_{(n_o-1,0)} = \sum_i f_i \beta^{m(n_o-i)}$.

BCH codes of length n_o consist of the set of all $\vec{f}$ such that $f(\beta^m) = 0$, for all m such that $m_o \leqslant m \leqslant m_o + d - 2$, where m_o and d are arbitrary integers and β is a field element of order n_o. It will be found that d is the minimum distance of the code. Information on the number of words in some binary BCH codes ($p=2$, $N=1$) has been given by Peterson. Commonly $\beta = \alpha$, a primitive element of $GF(p^M)$, and m_o equals zero or one.

Transmission of a code word through a noisy channel results in a reception represented by the vector $\vec{r} \equiv (r_1, r_2, \ldots, r_{n_o})$. If no errors occur, $\vec{r} = \vec{f}$. If t errors occur, there will be t places in which $\vec{r}$ differs from $\vec{f}$. If s erasures occur, then there will be s places in which $\vec{r}$ generally differs from $\vec{f}$, but in the case of erasures these places will be known to the receiver, rather than unknown as in the case of errors. If the j^{th} error is in the i^{th} place, then we shall call $X_j \equiv \beta^{n_o-i}$ the locator of the error, and $e_j \equiv r_i - f_i$ the value of the error. Similarly, if the k^{th} erasure is in the i^{th} place, then we shall call $Y_k \equiv \beta^{n_o-i}$ the locator of the erasure, and $d_k \equiv r_i - f_i$ the value of the erasure (possibly zero). The decoding problem is to find the e_j, X_j, and d_k, $1 \leq j \leq t$, $1 \leq k \leq s$, when $\vec{r}$ and the Y_k are known. The decoding algorithm that we shall give solves this problem whenever $2t + s \leq d - 1$.

The starting point in decoding BCH codes is to calculate from $\vec{r}$ a set of parity checks that depend only on the error-erasure pattern, and not on the particular code word sent. In this case the appropriate parity checks S_m are defined by

$$S_m \equiv r(\beta^m), \qquad m_o \leq m \leq m_o + d - 2$$

$$= \sum_{i=1}^{n_o} r_i \beta^{m(n_o-i)}.$$

It follows from the definitions of e_j, X_j, d_k, and Y_k that

$$r(\beta^m) = f(\beta^m) + \sum_{j=1}^{t} e_j X_j^m + \sum_{k=1}^{s} d_k Y_k^m.$$

But $f(\beta^m) = 0$, $m_o \leq m \leq m_o + d - 2$, whatever the code word $\vec{f}$. Therefore

$$S_m = \sum_{j=1}^{t} e_j X_j^m + \sum_{k=1}^{s} d_k Y_k^M.$$

We shall find it convenient in the sequel to define the column vectors $\vec{S}_{(a,b)} = (S_a, S_{a-1}, \ldots, S_b)^T$, $m_o \leq a \leq b \leq m_o + d - 2$, $\vec{X}_{j(a,b)} \equiv \left(X_j^a, X_j^{a-1}, \ldots, X_j^b\right)^T$, and $\vec{Y}_{k(a,b)} \equiv \left(Y_k^a, Y_k^{a-1}, \ldots, Y_k^b\right)^T$. Evidently,

$$S_{(a,b)} = \sum_{j=1}^{t} e_j \vec{X}_{j(a,b)} + \sum_{k=1}^{s} d_k \vec{Y}_{k(a,b)}.$$

Finally, let us consider the polynomial $\sigma(Z)$ defined by

$$\sigma(Z) \equiv (Z-Z_1)(Z-Z_2)\ldots(Z-Z_L),$$

where Z is an indeterminate and $Z_1, Z_2, \ldots, Z_L$ are members of a field. Clearly $\sigma(Z) = 0$ if and only if Z equals one of the Z_ℓ. Expanding $\sigma(Z)$, we get

$$\sigma(Z) = Z^L - (Z_1+Z_2+\ldots+Z_L)Z^{L-1} + \ldots + (-1)^L (Z_1Z_2\ldots Z_L).$$

The coefficient of $(-1)^{L-\ell}Z^\ell$ in this expansion is defined as the L-ℓth elementary symmetric function $\sigma_{L-\ell}$ of $Z_1, Z_2, \ldots, Z_L$; note that σ_o is always one. We define $\vec{\sigma}$ as the row vector $\left(\sigma_o, -\sigma_1, \ldots, (-1)^L \sigma_L\right)$; then the dot product

$$\vec{\sigma} \cdot \vec{Z}_{(L,0)} = \sigma(Z).$$

3. Modified Cyclic Parity Checks

The S_m are not the only parity checks that could be formed; in fact, any linear combination of the S_m is also a valid parity check. We look for a set of $d - s - 1$ independent parity checks that, unlike the S_m, do not depend on the erased symbols but still retain the general properties of the S_m.

Define $\sigma_d(Z) \equiv (Z-Y_1)(Z-Y_2)\ldots(Z-Y_s)$, and let $\vec{\sigma}_d$ then be the vector of the symmetric functions σ_{dk} of the erasure locators Y_k, as above. We define the modified cyclic parity checks T_n by

$$T_n \equiv \vec{\sigma}_d \cdot \vec{S}_{(m_o+n+s, m_o+n)}. \tag{1}$$

Since we must have $m_o \leq m_o + n$ and $m_o + n + s \leq m_o + d - 2$, the range of n is $0 \leq n \leq d - s - 2$. In the case of no erasures, $T_n = S_{m_o+n}$.

Now, since

$$\vec{S}_{(m_o+n+s, m_o+n)} = \sum_{j=1}^{t} e_j \vec{X}_{j(m_o+n+s, m_o+n)} + \sum_{k=1}^{s} d_k \vec{Y}_{k(m_o+n+s, m_o+n)}$$
$$= \sum_{j=1}^{t} e_j X_j^{m_o+n} \vec{X}_{j(s,0)} + \sum_{k=1}^{s} d_k Y_k^{m_o+n} \vec{Y}_{k(s,0)}, \tag{2}$$

we have

$$T_n \equiv \vec{\sigma}_d \cdot \vec{S}_{(m_o+n+s, m_o+n)} = \sum_{j=1}^{t} e_j X_j^{m_o+n} \vec{\sigma}_d \cdot \vec{X}_{j(s,0)} + \sum_{k=1}^{s} d_k Y_k^{m_o+n} \vec{\sigma}_d \cdot \vec{Y}_{k(s,0)}$$
$$= \sum_{j=1}^{t} e_j X_j^{m_o} \sigma_d(X_j) X_j^n + \sum_{k=1}^{s} d_k Y_k^{m_o+n} \sigma_d(Y_k)$$
$$= \sum_{j=1}^{t} E_j X_j^n. \tag{3}$$

Here, we have defined $E_j \equiv e_j X_j^{m_o} \sigma_d(X_j)$ and used the fact that $\sigma_d(Y_k) = 0$ (because Y_k is one of the erasure locators upon which σ_d is defined). Because the modified cyclic parity checks can be expressed as the simple function of the error locators given by Eq. 3 we may solve for the error locators exactly as if there were no erasures and the minimum distance were d - s.

4. Determining the Number of Errors

If d - s is odd, the maximum number of errors that can be corrected is $t_o = \frac{1}{2}(d-s-1)$, while if d - s is even, up to $t_o = \frac{1}{2}(d-s-2)$ errors are correctable, and $t_o + 1$ errors are detectable.

We now show that the actual number of errors t is the rank of a certain $t_o \times t_o$ matrix M, whose components are modified cyclic parity checks, as long as $t \leq t_o$. In order to do this we use the theorem of algebra in which the rank of a matrix is t if and only if there is at least one $t \times t$ submatrix with a nonzero determinant, and all $(t+1) \times (t+1)$ submatrices have zero determinants. We also use the fact that the determinant of a matrix that is the product of square matrices is the product of the determinants of the square matrices.

THEOREM (after Gorenstein and Zierler[8]): If $t \leq t_o$, then M has rank t, where

$$M \equiv \begin{bmatrix} T_{2t_o-2} & T_{2t_o-3} & \cdots & T_{t_o-1} \\ T_{2t_o-3} & T_{2t_o-4} & \cdots & T_{t_o-2} \\ \vdots & \vdots & & \vdots \\ T_{t_o-1} & T_{t_o-2} & \cdots & T_0 \end{bmatrix}.$$

Since $2t_o - 2 < d - s - 2$, all the T_n in this matrix are available.

PROOF: First consider the $t \times t$ submatrix M_t formed by the first t rows and columns of M. Using Eq. 3, we can write M_t as the product of three $t \times t$ matrices as follows:

$$M_t \equiv \begin{bmatrix} T_{2t_o-2} & T_{2t_o-3} & \cdots & T_{2t_o-t-1} \\ T_{2t_o-3} & T_{2t_o-4} & \cdots & T_{2t_o-t-2} \\ \vdots & \vdots & & \vdots \\ T_{2t_o-t-1} & T_{2t_o-t-2} & \cdots & T_{2t_o-2t} \end{bmatrix}$$

$$= \begin{bmatrix} X_1^{t-1} & X_2^{t-1} & \cdots & X_t^{t-1} \\ X_1^{t-2} & X_2^{t-2} & \cdots & X_t^{t-2} \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} E_1X_1^{2t_o-2t} & 0 & 0 \\ 0 & E_2X_2^{2t_o-2t} & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & E_tX_t^{2t_o-2t} \end{bmatrix} \begin{bmatrix} X_1^{t-1} & X_1^{t-2} & \cdots & 1 \\ X_2^{t-1} & X_2^{t-2} & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ X_t^{t-1} & X_t^{t-2} & \cdots & 1 \end{bmatrix}.$$

This may be checked by direct multiplication.

The center matrix is diagonal, and therefore its determinant is $\prod_j E_j X_j^{2t_o - 2t}$. Since $E_j = e_j X_j^{m_o} \sigma_d(X_j)$ and $X_j \neq Y_k$, $e_j \neq 0$, this determinant is nonzero. The first matrix is Van der Monde, with the determinant $\prod_{i>j} (X_i - X_j)$, which is nonzero because the error locators are distinct; the third is the transpose of the first, and has the same determinant. The determinant $|M_t|$ is the product of three nonzero factors, and is therefore itself nonzero. Thus the rank of M is t or greater.

Now consider any of the (t+1) × (t+1) submatrices of M, which will have the general form

$$\begin{bmatrix} T_{a_o+b_o} & T_{a_o+b_1} & \cdots & T_{a_o+b_t} \\ T_{a_1+b_o} & T_{a_1+b_1} & \cdots & T_{a_1+b_t} \\ \vdots & \vdots & & \vdots \\ T_{a_t+b_o} & T_{a_t+b_1} & \cdots & T_{a_t+b_t} \end{bmatrix}$$

$$= \begin{bmatrix} X_1^{a_o} & X_2^{a_o} & \cdots & X_t^{a_o} & 0 \\ X_1^{a_1} & X_2^{a_1} & \cdots & X_t^{a_1} & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ X_1^{a_t} & X_2^{a_t} & \cdots & X_t^{a_t} & 0 \end{bmatrix} \begin{bmatrix} E_1 & 0 & \cdots & 0 & 0 \\ 0 & E_2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & E_t & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} X_1^{b_o} & X_1^{b_1} & \cdots & X_1^{b_t} \\ X_2^{b_o} & X_2^{b_1} & \cdots & X_2^{b_t} \\ \vdots & \vdots & & \vdots \\ X_t^{b_o} & X_t^{b_1} & \cdots & X_t^{b_t} \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

in which the a_i and b_i are the chosen row and column numbers (read backwards). That this threefold decomposition can be made may again be checked by direct multiplication with the use of Eq. 3. Each of the three factor matrices has a zero determinant; therefore all (t+1) × (t+1) submatrices of M have zero determinants. Thus the rank of M can be no greater than t; but then it is t. Q. E. D.

5. Locating the Errors

Still following Gorenstein and Zierler, we now consider the vector $\vec{\sigma}_e$ of elementary symmetric functions σ_{ej} of the X_j, and its associated polynomial $\sigma_e(X) = \vec{\sigma}_e \cdot \vec{X}_{(t,0)}$. If we can find the components of $\vec{\sigma}_e$, we can determine the error locators by finding the t distinct roots of $\sigma_e(X)$. Defining $T_{(a,b)} \equiv (T_a, T_{a-1}, \ldots, T_b)^T$, $0 \leqslant b \leqslant a \leqslant d - s - 2$, we have

$$\vec{\sigma}_e \cdot \vec{T}_{(n'+t,n')} = \sum_{j=1}^{t} E_j X_j^{n'} \sigma_e(X_j) = 0, \qquad 0 \leq n' \leq d-s-t-2.$$

We know that σ_{eo}, the first component of σ_e, equals one. Since the range of n' is at least $2t_o - t$, we have a set of $2t_o - t$ equations in t unknowns. Remembering that $t \leq t_o$ by assumption, we take the t equations specified by $2t_o - t - 1 \geq n' \geq 2t_o - 2t$, which become in matrix form

$$-\begin{bmatrix} T_{2t_o-1} \\ T_{2t_o-2} \\ \vdots \\ T_{2t_o-t} \end{bmatrix} = \begin{bmatrix} T_{2t_o-2} & T_{2t_o-3} & \cdots & T_{2t_o-t-1} \\ T_{2t_o-3} & T_{2t_o-4} & \cdots & T_{2t_o-t-2} \\ \vdots & \vdots & & \vdots \\ T_{2t_o-t-1} & T_{2t_o-t-2} & \cdots & T_{2t_o-2t} \end{bmatrix} \begin{bmatrix} -\sigma_{e1} \\ \sigma_{e2} \\ \vdots \\ (-1)^t \sigma_t \end{bmatrix}.$$

Or, defining $\vec{\sigma}'_e \equiv (-\sigma_{e1}, \sigma_{e2}, \ldots, (-1)^t \sigma_{et})$, we have

$$-\vec{T}_{(2t_o-1, 2t_o-t)} = \vec{\sigma}'_e M_t. \tag{4}$$

Since $0 \leq 2t_o - 2t$ and $2t_o - 1 \leq d - s - 2$, all of the T_n needed to form these equations are available.

We have already shown that M_t has rank t, so that these equations are soluble for $\vec{\sigma}'_e$ and hence $\vec{\sigma}_e$. Then, since $\sigma_e\left(\beta^{n_o-i}\right)$ is zero if and only if β^{n_o-i} is an error locator X_{j_o}, calculation of $\sigma_e\left(\beta^{n_o-i}\right)$ for each i will reveal in turn the positions of all t errors.

6. Remark

In Peterson,[6] first the rank of M is found, and then a set of t equations in t unknowns is solved, as here. We remark that with the definition of M_t we have made, these two steps may be combined into one. (This feature is implicit in Gorenstein and Zierler.) For consider the equations

$$-\vec{T}_{(2t_o-1, t_o)} = \vec{\sigma}''_e M \tag{5}$$

where $\vec{\sigma}''_e \equiv (-\sigma_{e1}, \sigma_{e2}, \ldots, (-1)^t \sigma_{et}, 0, \ldots, 0)$. An efficient way of solving Eq. 5 is by Gauss-Jordan reduction to upper triangular form. Since the rank of M is t, this reduction will leave t nontrivial equations, the last $t_o - t$ equations being simply 0 = 0. But M_t is the upper left-hand corner of M, so that the upper left-hand corner of the reduced M will be the reduced M_t. We can therefore set the last $t_o - t$ components of $\vec{\sigma}''_e$ to zero, and get a set of equations equivalent to Eqs. 4, which can be solved for $\vec{\sigma}'_e$. Thus we need only one reduction, not two; since Gauss-Jordan reductions are tedious, this may be a significant saving.

7. Solving for the Values of the Erased Symbols

Once the errors have been located, they can be treated as erasures. We are then interested in the problem of determining the values of s + t erased symbols, given that there are no errors in the remaining symbols. To simplify notation, we consider the problem of finding the d_k given Y_k, $1 \leq k \leq s$, and $t = 0$.

Since the parity checks are linear functions of the erasure values, Peterson notes that we could solve s parity-check equations for the s d_k. There is another approach, however, which is more efficient.

Suppose we want to find d_{k_o}. As an aid to understanding this approach, let us think of treating the remaining s - 1 erasures as erasures, but making a stab at guessing d_{k_o}. This would give us a word with s - 1 erasures and either one or (on the chance of a correct guess) zero errors. The rank of the matrix M_1 would therefore be either zero or one; but M_1 is simply a single modified cyclic parity check, formed from the elementary symmetric functions of the s - 1 remaining erasure locators. Its vanishing would therefore tell us when we had guessed d_{k_o} correctly.

To derive an explicit equation for d_{k_o}, let $\overrightarrow{{}_{k_o}\sigma_d}$ be the vector of elementary symmetric functions of the s - 1 erasure locators $Y_1, Y_2, \ldots, Y_{k_o-1}, Y_{k_o+1}, \ldots, Y_s$. Since $t = 0$, we have from Eq. 2

$$\vec{S}_{(m_o+d-2, m_o+d-s-1)} = \sum_{k=1}^{s} d_k Y_k^{m_o+d-s-1} \vec{Y}_{k(s-1,o)}$$

and therefore

$$\begin{aligned}
{}_{k_o}T_{d-s-1} &\equiv \overrightarrow{{}_{k_o}\sigma_d} \cdot \vec{S}_{(m_o+d-2, m_o+d-s-1)} \\
&= d_{k_o} Y_{k_o}^{m_o+d-s-1} \, {}_{k_o}\sigma_d\left(Y_{k_o}\right) + \sum_{k \neq k_o} d_k Y_k^{m_o+d-s-1} \, {}_{k_o}\sigma_d(Y_k) \\
&= d_{k_o} Y_{k_o}^{m_o+d-s-1} \, {}_{k_o}\sigma_d\left(Y_{k_o}\right),
\end{aligned}$$

since ${}_{k_o}\sigma_d(Y_k) = 0$, $k \neq k_o$. Thus

$$d_{k_o} = \frac{{}_{k_o}T_{d-s-1}}{Y_{k_o}^{m_o+d-s-1}\ {}_{k_o}\sigma_d\left(Y_{k_o}\right)}.$$

This gives us an explicit formula for d_{k_o}, which is valid for any s:

$$d_{k_o} = \frac{S_{m_o+d-2} - {}_{k_o}\sigma_{d1} S_{m_o+d-3} + {}_{k_o}\sigma_{d2} S_{m_o+d-4} - \dots}{Y_{k_o}^{m_o+d-2} - {}_{k_o}\sigma_{d1} Y_{k_o}^{m_o+d-3} + {}_{k_o}\sigma_{d2} Y_{k_o}^{m_o+d-4} - \dots}. \tag{6}$$

We can find all erasure values in this way; each requires calculation of the symmetric functions of a different set of s - 1 locators. Alternatively, after finding d_1, we could modify all parity checks to account for this information $\left(\vec{S}'_{(m_o+d-2,m_o)} = \vec{S}_{(m_o+d-2,m_o)} - d_1\vec{Y}_{1(m_o+d-2,m_o)}\right)$, and solve for d_2 in terms of these new parity checks and the remaining s - 2 erasure locators, and so forth.

G. D. Forney, Jr.

References

1. R. W. Hamming, Error detecting and error correcting codes, Bell System Tech. J. 29, 147-160 (1950).

2. R. C. Bose and D. K. Ray-Chaudhuri, On a class of error-correcting binary group codes, Inform. Contr. 3, 68-79 (1960).

3. R. C. Bose and D. K. Ray-Chaudhuri, Further results on error-correcting binary group codes, Inform. Contr. 3, 279-290 (1960).

4. A. A. Hocquenghem, Codes correcteurs d'erreurs, Chiffres 2, 147-156 (1959).

5. W. W. Peterson, Encoding and error-correction procedures for the Bose-Chaudhuri codes, IRE Trans. Vol. IT-6, pp. 459-470, 1960.

6. W. W. Peterson, Error-Correcting Codes (The M. I. T. Press, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1961).

7. T. C. Bartee and D. I. Schneider, An electronic decoder for Bose-Chaudhuri-Hocquenghem error-correcting codes, IRE Trans. Vol. IT-8, pp. S17-S24, 1962.

8. D. Gorenstein and N. Zierler, A class of cyclic linear error-correcting codes in p^m symbols, J. SIAM 9, 207-214 (1961).

9. F. J. Bloom, S. S. L. Chang, B. Harris, A. Hauptschein, and K. C. Morgan, Improvement of binary transmission by null-zone reception, Proc. IRE 45, 963-975 (1957).

10. P. Elias, Coding and Decoding, *Lectures on Communication System Theory*, edited by E. J. Baghdady (McGraw-Hill Publishing Company, New York, 1961).

11. W. W. Peterson, *Error-Correcting Codes*, *op. cit.*, p. 181.

B. MODULAR THRESHOLD FUNCTIONS

This report presents some preliminary results regarding a special class of multiple threshold functions. Multiple threshold functions have been treated by Ercoli and Mercurio.[1] The functions to be treated here are called modular threshold functions (MTF) and are defined as follows.

DEF 1. A Boolian function $F(x_1, \ldots, x_n)$ is a modular threshold function iff there exists a set of real numbers $w_1, \ldots w_n$, A, B and T such that the following constraints are satisfied.

1) $B > 0$

2) $A \leqslant T < A + B$

3) For each combination of variable values x_i the integer p is so chosen that

$$A \leqslant \sum_{i=1}^{n} w_i x_i + pB < A + B.$$

If the quantity $\sum_{i=1}^{n} w_i x_i + pB$ is compared with T, then

$$\sum_{i=1}^{n} w_i x_i + pB \geqslant T \leftrightarrow F(x_1, \ldots, x_n) = 1$$

$$\sum_{i=1}^{n} w_i x_i + pB < T \leftrightarrow F(x_1, \ldots, x_n) = 0.$$

Let us indicate some properties of the class of MTF. The proofs for some of the following theorems are not complete, but enough information is provided so that the reader can finish the proof.

LEMMA 1. All linearly separable[2] functions are MTF.

PROOF. Consider a realization of a linearly separable function $F(x_1, \ldots, x_n)$ with weights q_i and threshold T'. Then the MTF realization of F will have

$$A = \min \sum_{i=1}^{n} q_i x_i$$

$$B = \max \sum_{i=1}^{n} q_i x_i - \min \sum_{i=1}^{n} q_i x_i + \alpha; \quad \alpha > 0$$

$$T = T'$$

$$w_i = q_i, \quad 1 \leqslant i \leqslant n.$$

With these parameters the value of p in Definition 1 will always be zero and the comparisons with the threshold will be identical to those of the linearly separable realization.

LEMMA 2. For all $n > 1$ there exists at least one $F(x_1, \ldots, x_n)$ that is MTF and not linearly separable.

PROOF. For $n > 1$ consider the nonlinearly separable function.

$$F(x_1, \ldots, x_n) = x_1 \oplus x_2 \oplus \ldots \oplus x_n.$$

With the choice $w_i = 1$, $A = 0$, $B = 2$, the value of $\sum\limits_{i=1}^{n} w_i x_i + pB$ satisfying Definition 1 is such that

$$\sum_{i=1}^{n} w_i x_i + pB = x_1 \oplus x_2 \oplus \ldots \oplus x_n.$$

With $T = 1$, $F(x_1, \ldots, x_n)$ is MTF.

THEOREM 1. The class of linearly separable functions is properly contained in the class of modular threshold functions. This follows directly from Lemmas 1 and 2.

THEOREM 2. The MTF class is closed under functional complementation, that is, if $F(x_1, \ldots, x_n)$ is MTF so is $\overline{F}(x_1, \ldots, x_n)$.

PROOF. We shall give the parameters for an MTF realization of $\overline{F}$. Assume that the realization of F has w_i, A, B, and T. The realization of $\overline{F}$ will have w_i', A', B', and T as follows:

$$w_i' = w_i$$

$$A' = T$$

$$B' = B$$

$$T' = A + B.$$

THEOREM 3. The MTF class is closed under single-variable complementation, that is, if $F(x_1, \ldots, x_k, \ldots, x_n)$ is an MTF so is $F(x_1, \ldots, \overline{x}_k, \ldots, x_n)$.

PROOF. Assume that the realization of $F(x_1, \ldots, x_k, \ldots x_n)$ has the parameters w_i, A, B, T. Then the realization of $F(x_1, \ldots, \overline{x}_k, \ldots x_n)$ will have the parameters w_i', A', B', and T' as follows:

$$A' = A - w_k$$

$$B' = B$$

$$T' = T - w_k$$

$$w_k' = -w_k$$

$$w_i' = w_i, \quad 1 \leq i \leq n, \quad i \neq k.$$

COROLLARY 1. The MTF class is closed under dualization, that is, if $F(x_1, \dots, x_n)$ is MTF so is $F_d(x_1, \dots, x_n) = \overline{F}(\overline{x}_1, \dots, \overline{x}_n)$.

THEOREM 4. Assume $F(x_o, x_1, \dots, x_n) = x_o f_1(x_1, \dots, x_n) + \overline{x}_o f_2(x_1, \dots, x_n)$ is MTF. Then both f_1 and f_2 are MTF.

PROOF. Let the realization of F have the parameters w_i, $0 \leq i \leq n$, A, B and T. Let $x_o = 1$. Then $F = f_1$, and f_1 is realized with $w_i^1 = w_i$, $1 \leq i \leq n$, $A^1 = A - w_o$, $B^1 = B$ and $T^1 = T - w_o$. Let $x_o = 0$. Then $F = f_2$, and f_2 is realized with $w_i^2 = w_i$, $1 \leq i \leq n$, $A^2 = A$, $B^2 = B$ and $T^2 = T$.

A multiple-threshold function is defined by a set of weights w_i and a set of thresholds $T_1, T_2, \dots, T_m$. The usual convention is that $T_1 > T_2 > \dots > T_m$. It is also necessary to specify the value of the function in some range, usually for $\Sigma\, w_i x_i \geq T_1$. The function value for $T_j \leq \Sigma\, w_i x_i < T_{j-1}$ is then the complement of its value for $T_{j-1} \leq \Sigma\, w_i x_i < T_{j-2}$.

LEMMA 3. The class of two-threshold threshold-realizable functions (2T-TR) is included in the class MTF.

PROOF. Consider a 2T-TR function $F(x_1, \dots, x_n)$ with weights w_i, $1 \leq i \leq n$ and thresholds T_1 and T_2 with $T_1 > T_2$.

Case I.

Assume $F = 0$ for $\sum\limits_{i=1}^{n} w_i x_i \geq T_1$. The MTF realization of $F(x_1, \dots, x_n)$ will have

$$A = \min\left(\sum_{i=1}^{n} w_i x_i\right) - \left[\max\left(\sum_{i=1}^{n} w_i x_i\right) - \overset{\min}{\sum_{i=1}^{n}} > T_1\left(\sum_{i=1}^{n} w_i x_i\right)\right]$$

$$B = \max\left(\sum_{i=1}^{n} w_i x_i\right) - \min\left(\sum_{i=1}^{n} w_i x_i\right)$$

$$w_i^{MTF} = w_i, \quad 1 \leq i \leq n$$

$$T = T_2.$$

Case II.

A 2T-TR has value $F = 1$ for $\sum\limits_{i=1}^{n} w_i x_i \geq T$. Find the MTF realization of $\overline{F}$ and note that the class MTF is closed under functional complementation. Q. E. D.

THEOREM 5.[3] A function $G(x_1, \dots, x_n)$ is 2T-TR iff there exists a decomposition $G = f_1 + f_2$ such that $\overline{x}_o f_1 + x_o \overline{f}_2$ is linearly separable.

THEOREM 6. A function $G(x_1, \dots, x_n)$ is MTF if there exists a decomposition $G = f_1 + f_2$ such that $G(x_1, \dots, x_n) = \overline{x}_o f_1 + x_o \overline{f}_2$ is linearly separable.

LEMMA 4. There exist MTF functions for which the above-mentioned decomposition does not exist.

PROOF. It can be shown that $G(x_1, \ldots, x_n) = x_1 \oplus x_2 \oplus \ldots \oplus x_n$ for $n \geq 3$ cannot be decomposed as in Theorem 6. By Lemma 2, G is MTF.

THEOREM 7. The class of 2T-TR functions is properly included in the class of modular threshold functions.

We can make some informal observations regarding modular threshold functions. First, if the quantities A and B are associated with an MTF realization, the same function will be realized if A is replaced by $A+KB$, where K is an integer. Also no weight associated with the MTF realization need have magnitude greater than or equal to B. Also, all weights can be positive. Thus for any MTF realization, $0 \leq w_i < B$. This in turn bounds the quantity $\sum_{i=1}^{n} w_ix_i$, that is, $\sum_{i=1}^{n} w_ix_i < nB$. By using this bound it is possible to show that there is a multiple-threshold realization of any MTF with a maximum of 2n thresholds.

The MTF characterization allows any modular threshold function to be represented in compact form. This form is an (n+3)-tuple giving the parameters w_i, A, B, and T. (It is not known by the author whether all functions can be so represented.)

We shall indicate how the concept of modular threshold functions can be used to obtain some error-correction ability for noisy combinational logic circuits. Let us restrict our attention to MTF realizations in which all of the parameters w_i, A, B, and T are non-negative integers. (Note that any realization having all rational parameters can be scaled to satisfy this requirement.) The input combinations of an integer realization of a function F can be grouped according to the value of the quantity $\sum_{i=1}^{n} w_ix_i + pB$ associated with them. An ordering of these groups is indicated in Fig. XIX-1, where each K_i is a non-negative integer, $K_i < K_j$ if $i < j$, and $K_m < B$. Assume that the threshold for the function F is $T = A + K_j$ as shown.

Figure XIX-2 shows a set of functions F_i, $1 \leq i \leq m-1$ that are derived by using the weights associated with F and varying the threshold. In particular, F_i is the function obtained by using $A + K_{i+1}$ as the threshold. We shall now assign output weights q_i to the functions F_i so that for all input combinations and their corresponding output combinations

$$\sum_{i=1}^{n} w_ix_i + pB - \sum_{i=1}^{m-1} q_iF_i = \text{constant.}$$

It can be shown that the choice

$$q_i = K_{i+1} - K_i$$

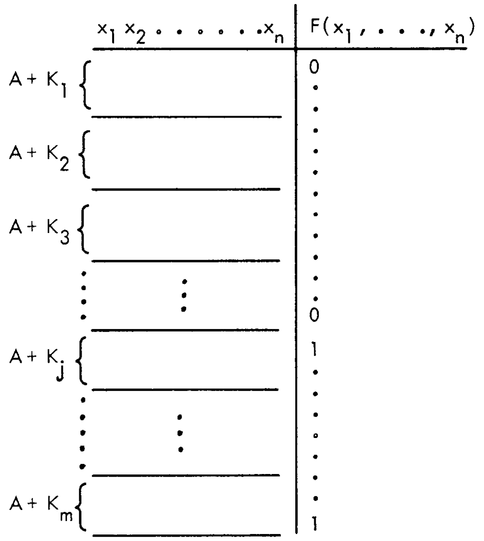

Fig. XIX-1. Modular threshold function.

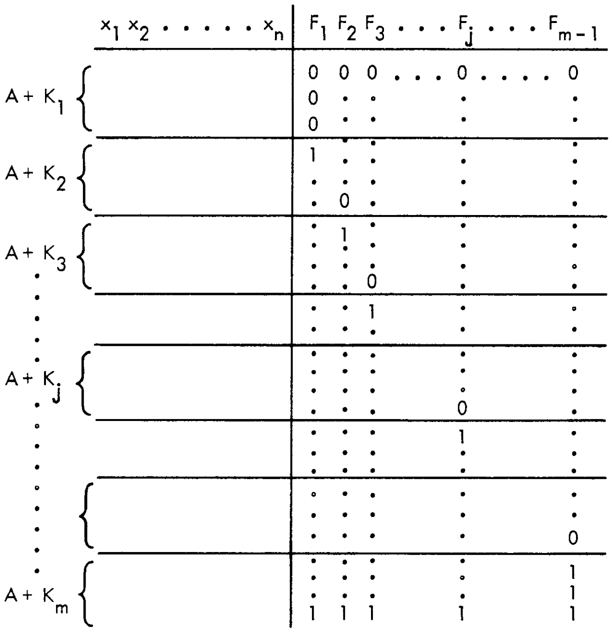

Fig. XIX-2. Modular threshold set.

yields

$$\sum_{i=1}^{n} w_i x_i + pB - \sum_{i=1}^{m-1} q_i F_i = A + K_1.$$

Figure XIX-3 shows a circuit in which the combinational logic realizes the set of functions F_i described above. In addition, this same combinational logic realizes a function F_m such that $x_1 \oplus x_2 \oplus \ldots \oplus x_n = F_1 \oplus F_2 \oplus \ldots \oplus F_m$. The inputs to this logic (x_i') come from block encoders associated with the over-all inputs x_i. These encoders have two properties: (i) the same single-error-correcting code is imposed on the r-bit outputs of all the encoders; (ii) for every s-bit input sequence they produce an r-bit output sequence such that $\sum_r w_i x_i' = N$. The r-bit output sequences of the combinational logic are processed by a decoder that produces s-bit sequences. These Z_i sequences should, in the absence of noise, be the same as those resulting from sequential operation of the combinational circuit on the original s-bit x_i sequences. It is assumed that the encoding and decoding circuits are not subject to error during their operation, although the combinational logic may be. During the operation of the combinational logic one or more of the output functions may have a value equal to the complement of the correct output value.

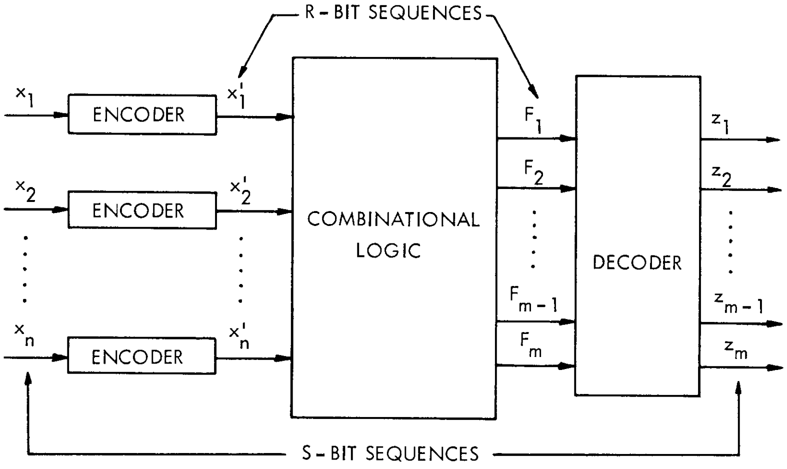

Fig. XIX-3. Logic and coding circuits.

If, during the processing of some block of information, the combinational logic is error-free then

$$\sum_{i=1}^{n} N_i + p_1 B = \sum_{r} \sum_{i=1}^{m-1} q_i F_i + r(A+K_1) + p_2 B,$$

where the integers p_1 and p_2 are chosen to make the left- and right-hand sides of this equation greater than or equal to A and less than A + B. Also the r-bit word formed at the decoder by computing $F_1 \oplus F_2 \oplus \dots \oplus F_m$ for each operation will satisfy the same parity equations as are imposed by the encoders.

Assume now that a single error occurs in some F_i during one of the r operations in a block. The quantity $\sum_{i=1}^{m-1} q_iF_i + r(A+K_1)$ will be changed by an amount $\pm q_i$, the sign of q_i being determined by the type of error. The parity equations on $F_1 \oplus F_2 \oplus \dots \oplus F_m$ will also give the location of the error within the r-bit sequence. Knowledge of the value of the actual output at this position and the value of

$$E = \sum_{i=1}^{n} N_i - \left[\sum_{r} \sum_{i=1}^{m-1} q_iF_i + r(A+K_1)\right] + pB, \qquad A \leqslant E < A + B$$

is sufficient to correct the error. This can best be demonstrated by showing that no situation occurs in which two different correct outputs can be corrupted, by at most a single error, to yield the same actual output and the same value of E. This proof will not be presented.here.

Thus the concept of modular threshold functions provides a new approach to single-error correction for multiple output circuits. If it can be determined that the set of functions to be realized is a set of the type described, the error correction technique can be applied with only the addition of F_m. Alternatively, if one desires a realization of a single MTF, the technique describes a method for adding redundant functions so as to achieve some degree of error-correcting ability.

R. N. Spann

References

1. P. Ercoli and L. Mercurio, Threshold Logic with One or More than One Threshold, Proceedings of the IFIPS Congress, 1962 (North-Holland Publishing Company, Amsterdam, 1962), pp. 741-745.

2. C. L. Coates and P. M. Lewis II, Linearly separable switching functions, J. Franklin Inst. 272, 360-410 (1961).

3. D. Haring, Private communication, M. I. T., 1964.

C. THE COMPUTATION PROBLEM WITH SEQUENTIAL DECODING

This report summarizes a thesis that will be submitted to the Department of Electrical Engineering, M. I. T. on February 5, 1965 in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

The Fano Sequential Decoding procedure is a technique for communicating at a high information rate and with a high reliability over a large class of channels. Its use is limited, however, by equipment cost and by variation in the time required to decode successive transmitted digits. It is with the latter issue that this work is concerned.

Others have shown that the average processing time per decoded digit is small if the information rate of the source is less than a rate R_{comp}. In this thesis the probability distribution of the processing time random variable is studied. The results of this examination are applied to the buffer overflow probability, i.e., the probability that the decoding delay forces incoming data to fill and overflow a finite buffer. It is shown that the overflow probability is relatively insensitive to the buffer storage capacity and to the computational speed of the decoder, but that it is quite sensitive to information rate. In particular, halving the source rate more than squares the overflow probability. It is found that these sensitivities are basic to sequential decoding, and arise because the computation per decoded digit is large during an interval of high channel noise and grows exponentially with the length of such an interval.

A conjecture is presented concerning the exact behavior of the overflow probability with information rate. This conjecture agrees well with the (limited) experimental evidence that is now available.

J. E. Savage