IX. PROCESSING AND TRANSMISSION OF INFORMATION*

Prof. E. Arthurs
Prof. P. Elias
Prof. R. M. Fano
Prof. J. B. Dennis
Prof. E. M. Hofstetter
Prof. D. A. Huffman
Prof. R. C. Jeffrey
Prof. H. Rogers, Jr.
Prof. C. E. Shannon
Prof. W. M. Siebert
Prof. J. M. Wozencraft
Dr. M. Eden
Dr. M. P. Schützenberger
D. C. Coll
M. Corradetti
H. A. Ernst
E. F. Ferretti
R. G. Gallager
U. F. Gronemann
F. C. Hennie III
M. Horstein
T. S. Huang
F. Jelinek
T. Kailath
L. Lofgren
B. Reiffen
L. G. Roberts
J. L. Rosenfeld
N. K. H. Tam
W. L. Wells
H. P. Zeiger

A. A METHOD OF PLANAR PREDICTION FOR CODING PICTURES

The digital techniques that were illustrated in earlier reports (1, 2, 3, 4) have been used to determine upper bounds on the amount of information necessary for specifying pictures (5).

The picture plane was divided into strips, each of which contained five scanning lines (see Fig. IX-1). The three initial samples A, B, and C that were used for each strip defined a plane from which the intensities of the adjacent sample points $P_{0,0}$, $P_{1,1}$, $P_{2,0}$, $P_{2,2}$, $P_{3,1}$, $P_{4,0}$ were predicted. For each of these points the difference between the true and predicted values was then determined.

The true ordinates of $P_{1,1}$, $P_{2,2}$, $P_{3,1}$, obtained by correction of their corresponding predicted values, were used to define another plane (see Fig. IX-2) that enabled the prediction of the ordinates of the samples $P_{0,1}$, $P_{1,2}$, $P_{2,3}$, $P_{3,2}$, $P_{4,1}$, as well as the calculation of the corresponding errors (differences between the true and the predicted values).

Such a procedure was continued until the whole strip was covered. From what has been said above, it follows that only the ordinates of the initial points A, B, and C and the prediction errors were necessary for specifying the picture in the strip.

The TX-0 computer was used to perform the operations required for planar prediction and resynthesis of the pictures. A summary of the main operations executed by the computer follows. For simplicity, the reflectance values will be indicated by the same letters as those for the points at which they are measured; for example, the letter "A" will indicate both sample A and its corresponding light intensity.

The average (A+C)/2 yielded the prediction $P'_{2,0}$ of $P_{2,0}$ and, as Fig. IX-3 shows, the knowledge of the differences $(C-P'_{2,0})$ and $(B-P'_{2,0})$ allowed the prediction of the points $P_{4,0}$, $P_{0,0}$, $P_{2,2}$.

---

*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.

The prediction of points $P'_{1,1}$ and $P'_{3,1}$ (Fig. IX-1) was made according to the following expressions:

$$P'_{1,1} = (P'_{0,0}+P'_{2,2})/2$$

$$P'_{3,1} = (P'_{4,0}+P'_{2,2})/2$$

Once the predicted values were known, the calculation of the errors was performed in a very straightforward way. A new prediction plane, determined from the values $P'_{1,1}$, $P'_{2,2}$, $P'_{3,1}$ corrected, and with their corresponding errors, was then ready for extending the decorrelation to other points of the strip.

Two other operations were executed by the computer: limitation of the predicted values, and quantization of the errors. Limitation consists of clipping all of the negative values (a negative value of reflectance has no physical meaning) and positive values larger than the quantum level 63 (64 equally spaced quantum levels are considered enough for still picture representation). These values are replaced by 0 and 63 quantum levels, respectively. Quantization of errors consists in limiting the number of levels used to specify the mistakes in prediction. The range of possible errors is ±63, and therefore we need 125 levels for their complete specification (-0 and +0 are considered coincident). If all these levels are allowed, it is obvious that the picture can be reconstructed at the receiver end in the form in which it was presented to the transmitter.

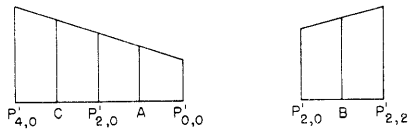

Fig. IX-1. Planar prediction procedure.

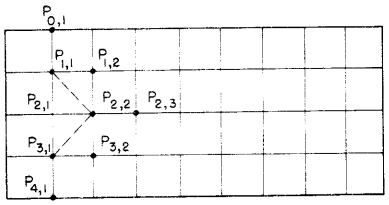

Fig. IX-2. Planar prediction procedure.

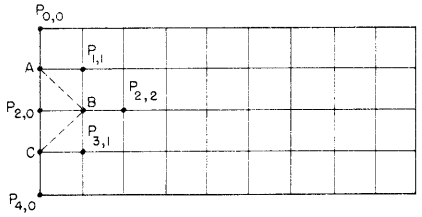

Fig. IX-3. Planar prediction procedure. The prime characterizes the predicted values: $P'_{2,0} = \frac{1}{2}(A+C)$; $P'_{0,0} = A - (C-P'_{2,0})$; $P'_{4,0} = C + (C-P'_{2,0})$; $P'_{2,2} = B + (B-P'_{2,0})$.

(a)

(b)

(c)

Fig. IX-4. Pictures reproduced by computer display. (a) Data processing with criterion a. (b) Data processing with criterion b. (c) Data processing with criterion c.

(a)

(b)

(c)

Fig. IX-5. Pictures reproduced by computer display. (a) Data processing with criterion a. (b) Data processing with criterion b. (c) Data processing with criterion c.

We called this criterion of remapping the data of a picture source "criterion a."

Two other quantization criteria were adopted: criterion b, 13 quantum levels $(0,\pm1,\pm4,\pm7,\pm20,\pm30,\pm40)$; and criterion c, 7 quantum levels $(0,\pm4,\pm14,\pm29)$. With the use of criteria b and c, distortions were introduced in the signal ensemble, with a consequent degradation of the quality of the picture reproduction.

To judge the importance of these distortions, a display subroutine was incorporated into the processing program to provide a high-speed output of the resynthesized

Table IX-1. Source-rate estimation.

Picture	Criterion	Upper Bound on Source Rate (bits/ sample)	Quantization Level
Fig. IX-4	a	2.66	125
Fig. IX-4	b	2.30	13
Fig. IX-4	c	1.95	7
Fig. IX-5	a	4.76	125
Fig. IX-5	b	3.43	13
Fig. IX-5	c	2.30	7

picture (6). Three resyntheses of a picture are shown in Fig. IX-4. The quality of reproduction with criterion b is definitely good, but that with criterion c suffers from the presence of a limited number of quantization levels, mainly at points where the light intensity surface of the picture presents many discontinuities and therefore is less suitable for planar prediction. The results of the same processing criteria for another picture are shown in Fig. IX-5, and similar remarks are valid. The estimated source rates for the two pictures are listed in Table IX-1.

M. Corradetti

References

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

2. W. A. Youngblood, Picture processing, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., July 15, 1958, pp. 134-136.

3. J. E. Cunningham, Recording pictures by generation of lowpass and correction signals, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., July 15, 1958, pp. 136-137.

4. R. S. Marcus, Picture coding by linear interpolation, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., July 15, 1958, pp. 137-140.

5. This report is based on a thesis by M. Corradetti, A method of planar prediction for coding pictures, S.M. Thesis, Department of Electrical Engineering, M.I.T., July 1959.

6. J. E. Cunningham, Reproduction of pictures by computer display, Quarterly Progress Report No. 53, Research Laboratory of Electronics, M.I.T., April 15, 1959, pp. 113-114.

B. AN INEQUALITY OF E. F. MOORE AND C. E. SHANNON

The following inequality on the "Boolian polynomial" h(p) associated with the Boolian function u(a) of the n variates a_i has been given by Moore and Shannon (1):

(I). $F_u = \dfrac{h'^2(p)}{h(p)(1-h(1))} \leq np^{-1}(1-p)^{-1}$

However, a more detailed version of the proof allows us to obtain instead:

(I)'. $F_u \leq A(p)\, p^{-1}(1-p)^{-1}$

where A(p) is, roughly speaking, the average number of bits of information needed for determining the value of u by a "binary decision program," as defined in the theory of C. Y. Lee (2).

1. Definitions

Let u be a Boolian function of the n variates a_i. We say that u is in normal form if, for some a_i,

$$u = a_i u' + \bar{a}_i u'' \qquad (\bar{a}\text{: the negation of } a)$$

where, recursively, u' and u" are in normal form and are independent of a_i. It is assumed that a function that is always equal to 0 (or to 1) is in normal form only if it is written as a constant, that is, not written at all.

For instance, if $u = a_1a_2 + a_1a_3 + a_2a_3a_4$, a normal form u^* of u could be $u^* = a_2(a_1+\bar{a}_1a_3a_4) + \bar{a}_2(a_3a_1)$.

To each normal form is associated a tree whose nodes correspond to the variates a_i, and whose branches form a minimal set of prime implicants (3) of u and of $\bar{u}$. In our example, these prime implicants are:

For u: a_2a_1; $a_2\bar{a}_1a_3a_4$; $\bar{a}_2a_3a_1$.

For $\bar{u}$: $a_2\bar{a}_1\bar{a}_3$; $a_2\bar{a}_1a_3\bar{a}_4$; $\bar{a}_2\bar{a}_3$; $\bar{a}_2a_3\bar{a}_1$.

The tree is given in Fig. IX-6.

Obviously, for a given u there exist, in general, several different normal forms and corresponding sets of prime implicants.

Let us now assume that the a_i's are random independent variates with $Pr(a_i=1) = p = 1 - q$ for all i's. Under this hypothesis, for a given normal form the average length of the prime implicants is a polynomial A(p) in p. It is easily seen that, for any p and any normal form u^*, we have $A(p) \leq n$ (=the number of variates). Indeed, if u is not a

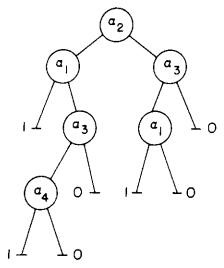

Fig. IX-6. Tree associated with a normal form u^* of $u = a_1a_2 + a_1a_3 + a_2a_3a_4$.

modulo 2 sum of the $\tilde{a}_i$'s ($\tilde{a}_i = a_i$ or $\bar{a}_i$), this relation is always a strict inequality.

Under the same hypothesis, h(p) is defined as Pr(u=1).

2. Verification of the Inequality

Let us suppose that $p = p_o$ or $= p_1$, two fixed values, and consider $W(p_o, p_1)$ the average value of the logarithm of the likelihood ratio (Wald information) when $p = p_o$.

For the observation of a single a_i, we have

$$W_a = p_o \log p_o/p_1 + q_o \log q_o/q_1$$

Consequently, if the a_i's are observed sequentially according to the tree describing some normal form u^*, we have

$$W_{u^*} = A(p_o)\, W_a$$

since, at every node, the probability that the corresponding a_i has the value 1 is independent of the result of the previous observations.

Finally, for the observation of the sole value of u we have

$$W_u = h(p_o) \log \frac{h(p_o)}{h(p_1)} + (1-h(p_o)) \log \frac{1 - h(p_o)}{1 - h(p_1)}$$

From the fact that, by construction, the knowledge of any prime implicant associated with u^* gives the value of this function, we obtain the inequality

(I)". $W_u \leq W_{u^*} = A(p_o)\, W_a \leq nW_a$

This ends the proof because, classically, the Fisher information, F(p), is equal to $\lim_{\epsilon \to 0} \epsilon^{-2} W(p, p+\epsilon)$.

Here, as is well known,

$$F_a = p^{-1}(1-p)^{-1} \text{ and } F_u = h'^2(p)(h(p))^{-1}(1-h(p))^{-1}$$

and (I)' results from (I)" by inserting these values in (I)".

M. P. Schützenberger

References

1. E. F. Moore and C. E. Shannon, J. Franklin Inst. 262, 191-208; 281-297 (1956).

2. C. Y. Lee, Bell System Tech. J. 38, 985-1000 (1959).

3. W. V. Quine, Am. Math. Month. 62, 627-631 (1955).

C. A CHARACTERISTIC PROPERTY OF CERTAIN POLYNOMIALS OF E. F. MOORE AND C. E. SHANNON

Let $\overline{L}_n$ be the set of all Boolian functions, and L_n the subset of all Boolian functions not involving the negation operation, in the n variates $\alpha(i)$ $(i=1, 2, \ldots, n)$. For any $\mu \in \overline{L}_n$, if the $\alpha(i)$ are random independent variates with $Pr(\alpha(i) = 1) = p$, then $Pr(\mu=1)$ is a polynomial (1), $h(\mu)$ in p. We give an elementary necessary and sufficient condition for the existence of at least a $\lambda \in L_n$ for which $h(\mu) = h(\lambda)$. As is well known (2), there is a natural one-to-one correspondence between L_n and the set of all simplicial complexes with, at most, n vertices. Consequently, this condition is also a characterization of the sequences of integers $\{a_j\}$ that can be the number of j-simplexes contained in a complex and its boundary. Because of this interpretation, it is unlikely that the condition is new, but I have not been able to find any relevant reference to it.

With the help of this condition and of the corresponding extremal functions $\omega(g)$ and $\Sigma P_{n-j}^{a_j}$, defined below, more elementary proofs can be given for Yamamoto's inequality (2) on the number of prime implicants of $\lambda \in L_n$ and for the Moore-Shannon lower bound (1) on the value of the derivative of $h(\lambda)$ $(\lambda \in L_n)$.

1. Notations

i. Let P_m^x be the set of the x first products of m of the variates $\alpha(i)$ when these products are taken in lexicographic order with respect to the indices i. We write P_m, instead of P_m^x, when x has its maximal value $\begin{bmatrix} n \\ m \end{bmatrix}$ and $P = \bigcup_m P_m$. For any subset $P' \subset P$, $\Sigma P'$ denotes the Boolian function (belonging to L_n) which is the sum of all the products, β, satisfying $\beta \leqslant \beta'$ for some $\beta' \in P'$. Conversely, for any $\lambda \in L_n$, $P_m\lambda$ is defined as the set of all the $\beta \in P_m$ that are such that $\beta \leqslant \lambda$. Thus, $\lambda = \Sigma P\lambda$ for any $\lambda \in L_n$. The set of all products, $\beta \in P'$, of the form $\beta = \beta'\alpha(i)$, with $\beta' \in P'$ and with $\alpha(i)$ not a factor of β', is denoted by $\Delta P'$ [cf. Yamamoto (2)].

ii. To every pair of positive integers x and m there corresponds one and only one strictly decreasing sequence of $m' \leqslant m$ positive integers: $y_1, y_2, \ldots, y_{m'}$ with the property that

$$x = \begin{bmatrix} y_1 \\ m \end{bmatrix} + \begin{bmatrix} y_2 \\ m-1 \end{bmatrix} + \ldots\ldots + \begin{bmatrix} y_{m'} \\ m - m' + 1 \end{bmatrix}$$

Consequently, the function

$$D_m(x) = \begin{bmatrix} y_1 \\ m-1 \end{bmatrix} + \begin{bmatrix} y_2 \\ m-2 \end{bmatrix} + \ldots\ldots + \begin{bmatrix} y_{m'} \\ m-m' \end{bmatrix}$$

is well determined for all non-negative integers x and m if we define $D_m(0)$ and $D_0(x)$ as zero.

For any x and m, $x + D_m(x) \geq D_{m+1}(x)$ (with a strict inequality if and only if $x > m + 1$). For all x, $D_m(x) + D_{m-1}(x') \geq D_m(x+x')$ if and only if $x' \leq D_m(x)$.

iii. For any $\mu \in \overline{L}_n$, we define the polynomial $g(\mu)$ as the product by $(1+t)^n$ of the function obtained when $(1+t)^{-1}$ is substituted for p in $h(\mu)$. The coefficient a_j of t^j in $g(\mu)$ is the number of monomials with n - j asserted, and j negated, variates $\alpha(i)$ in the canonical expansion of μ; when $\mu \in L_n$, a_j is also the number of elements in $P_{n-j}\mu$.

2. Statement of the Condition

For any $\mu \in \overline{L}_{n'}$, a necessary and sufficient condition that there exist a $\lambda \in L_n$ for which $g(\mu) = g(\lambda) \left(=a_0+a_1t+\ldots+a_mt^m\right)$ is that

$$\begin{bmatrix} n \\ j-1 \end{bmatrix} \geq a_{j-1} \geq D_j(a_j), \qquad \text{for all } j > 0$$

3. Verification

The condition is sufficient, since, for any polynomial $g(\mu)$ that fulfills it, we can define a function $\omega(g) \in L_n$ as

$$\omega(g) = \Sigma\left(\bigcup_j P_{n-j}^{a_j}\right)$$

and $\omega(g)$ satisfies $g(\omega(g)) = g$ because $\Delta P_m^x = P_{m+1}^{x'}$, when $x' = D_{n-m}(x)$.

It can be remarked that the functions $\Sigma P_{n-j}^{a_j}$ are the only functions in L_n for which $a_{j'-1} = D_{j'}(a_{j'})$ for all $j' \leq j$.

The condition is necessary. The first inequality is obvious. With respect to the proof of the second inequality it is enough to consider a truncated function $\lambda = \Sigma P_{n-j}\lambda$ with a_j prime implicants. Let α and α' be any two $\alpha(i)$'s. Then, λ can be written as $\alpha\alpha'A + \alpha(B+C) + \alpha'(B+C') + D$, where A, B, C, C', and D are sums of products not involving α and α', and where, furthermore, $P_{n-j+2}C$ and $P_{n-j+2}C'$ are disjoint sets. It is readily checked that the function $\lambda' = \alpha\alpha'A + \alpha(B+C+C') + \alpha'(B) + D$ is such that the set $P_{n-j}\lambda'$ has a_j elements and that the set $P_{n-j+1}\Delta\lambda'$ has, at most, as many elements as $P_{n-j+1}\Delta\lambda$. By taking successively $\alpha = \alpha(i)$ and $\alpha' = \alpha(i+1)$ for all i, we can reduce the function λ to a function $\Sigma P_j^{a_j}$ and the result is proved.

M. P. Schützenberger

References

1. E. F. Moore and C. E. Shannon, J. Franklin Inst. 262, 191-208; 281-297 (1956).
2. K. Yamamoto, J. Math. Soc. Japan 6, 343-353 (1954).

D. REDUNDANCY BOUNDS FOR w-ERROR CORRECTING CONTACT NETWORKS FOR A BOOLIAN FUNCTION

1. Introduction

In a well-known paper, von Neumann (11) shows that it is possible to design a combinational network for an arbitrarily accurate computation of a Boolian function, even though each component (Sheffer stroke) of the network has a certain probability of error. Moore and Shannon (6) have shown the same result for relay networks and have, moreover, found that this result is true even for arbitrarily poor relays. In both treatments the computational structure of the net, that is, the structure of the net in the error-free case, is maintained, and redundancy is applied on the components of the net.

Moore and Shannon (6) suggested that it may be more efficient to redesign the whole network to get redundancy than to replace each relay by an individual network. Especially with Shannon's block-coding method for arbitrarily reliable communication over a noisy channel (9) in mind, some methods of combined correction (3) appear promising for an efficient design of redundancy networks.

The purpose of this paper is to investigate this question for contact networks. As a measure of accuracy we shall take the largest weight, w, of the component errors for which the output will be corrected. A w-error correcting network corrects all errors of weights less than or equal to w. We shall assume that the probability of error in a branch (relay-contact) is independent of the state of the branch.

We find the following bounds on the minimum number of branches, n, (relay-contacts) in a w-error correcting redundancy network for a Boolian function B:

$$m(w+1)^2 \leq n \leq M(w+1)^2 \qquad (1)$$

where m is the number of nonredundant literals of B, and M is the number of contacts in a minimum branch network for B (without any specified error-correcting capability). If B has a nonredundant network (4, 5), m = M, and the two bounds coincide. However, when a nonredundant network for B does not exist, M is larger than m in order to meet the realizability conditions (see references 4 and 5). In the redundancy case the corresponding realizability conditions become less restrictive as w is increased, and so n will be close to the lower bound of Eq. 1.

Elias (1) has pointed out that, in a loose sense, time and number of components are interchangeable. If twice as many components are used in order to make a computer operate more reliably at the same rate, we could use the components to build two computers and compute twice as fast at the old level of reliability. In this sense, an increase of the number of components for obtaining greater reliability at the same rate reduces the rate of computation per component. Hence it is reasonable to consider the quantity

$$R = \frac{m}{n} \tag{2}$$

as a measure of the rate of computation per component. Evidently, R^{-1} is a relative measure of the redundancy. From Eq. 1, for the maximum rate of computation per component for w-error correction, we obtain

$$R_{max} \leqslant (w+1)^{-2} \tag{3}$$

Thus for arbitrarily high accuracy in the computation of any Boolian function, R_{max} is zero; that is, the capacity of computation is zero.

This may, at first, seem surprising when compared with the Shannon capacity theorem (9) for communication over a noisy channel. However, in order to communicate with an arbitrarily high reliability at nonzero rate, we must use an arbitrarily large encoder and decoder, which must operate, besides, under ideal (error-free) conditions. But, in the computation case we want to consider the computer (which eventually will contain separate parts as encoders and decoders) as an integrated system with a uniform application of noise to all components (at least to all components of the same kind). With this in mind, it is clear that we must pay with additional redundancy when we turn from the communication situation to the computation situation.

Elias (1) and Peterson (7) use a communication-like model with error-free encoders and decoders to prove negative results concerning computation.

In section 6 we shall see that if we do not restrict the relay network to be a contact network but allow a subnetwork of contact type to feed the relay coils of another contact subnetwork, then the lower bound of Eq. 1 is no longer valid. For instance, for single-error correction (w=1) we can, if B has a nonredundant network, approach a triplication of the number of components for very large m values instead of the quadrupling necessary in the case of contact networks.

2. Redundancy Possibilities in Contact Networks

Let us consider a Boolian function $B(a, a', b, \ldots)$ of the literals

$$(\gamma^* =)\ a, a', b, \ldots$$

We may think of the literals as a binary number γ^* with as many positions as there are literals of B. Next we want to introduce a redundant number of branches

$$(\gamma =)\ a_o, a_1, \ldots, a'_o, a'_1, \ldots, b_o, b_1, \ldots$$

In the error-free case these redundancy branches are related to the components of γ^* as follows: $a_i = a$, $a'_i = a'$, $b_i = b$, ..., for all i.

As in coding theory, we shall represent an error with a binary number e with as

many positions as there are positions in γ. If there is an error in the branch that corresponds to the i^{th} redundancy literal, e has a 1 in the i^{th} position. Otherwise it has a zero. Hence

$$\gamma_e = \gamma + e \qquad (\text{mod } 2) \tag{4}$$

represents a state, corresponding to the error e of the redundancy network. In this state the network generates the value

$$\underline{B}(\gamma_e) \tag{5}$$

We are primarily interested in the synthesis of networks that are insensitive to all e-errors up to a certain weight, w. We denote by E_w the set of e-numbers of weight less than or equal to w. Thus we require that

$$\underline{B}(\gamma_e) = B \qquad e \in E_w \tag{6}$$

Because of the errors in γ_e we cannot require that a_i' shall always be the negation of a_i. Therefore we use the symbol prime to indicate the negation of the letters (with no index) of B. We shall use the notation $\tilde{X}_i$ for the negation of the literal X_i of the $\underline{B}$-function.

Concerning the minimum amount of redundancy that is necessary for the specification according to Eq. 6, we at once obtain a necessary condition. The set of numbers γ_e for which $\underline{B} = 1$ and the set for which $\underline{B} = 0$ must be disjoint for $e \in E_w$. This condition is related to the Hamming bounds (2) of coding theory. As we know, these bounds are in general not sufficient for the existence of an error-correcting code. There are additional construction conditions that must be fulfilled. This is true here, too, regarding the existence of a $\underline{B}$-function satisfying Eq. 6. For a Boolian function B of only one variable there is no further requirement. But for a function B of more than one variable there is a further restriction that depends upon the redundancy possibility that we have at our disposal (compare section 5). In the group codes (10) we have the composition addition mod 2 at our disposal when we are forming the check-position content. For networks we can only duplicate the a_o-value to a_1, a_2, ..., and we may also specify that we shall have a_i'-branches at our disposal. This is true because the network is built only of a_i, a_i', b_i, ... branches; and other compositions, such as addition mod 2, multiplication (intersection, conjunction), and addition (union, disjunction), must be reflected in the structure of the $\underline{B}$-network.

The condition of Eq. 6 is only a partial specification of a truth table for $\underline{B}$. It is not a complete specification, because for errors $e \notin E_w$ we can choose the values of $\underline{B}(\gamma_e)$ as we wish. We shall take advantage of this freedom in order to give $\underline{B}$ a form that contains a small number of short prime implicants (8). In this connection, $\underline{B}$ must

fulfill three further conditions. Each of these is necessary, and all of them together are sufficient for the existence of a network with only one branch for each of the redundancy literals. The first will be given in Theorem 1. The other two, the c-criterion and the subrearrangement theorem, have been derived in the theory of nonredundant Boolian branch networks (4, 5).

3. A Condition on the Truth Table of a $\underline{B}$-Function

Because of the appearance of at least single errors in the γ-numbers we must treat any pair of redundancy literals a_i, a_i' as two independent variables. When constructing a $\underline{B}$-function from the partial truth-table specification according to Eq. 6 we must therefore require that the $\underline{B}$-function shall not contain any negations $(\tilde{a}_i)$, $(\tilde{a}_i')$, $(\tilde{b}_i)$, ... of the literals a_i, a_i', b_i, ... For, otherwise, we would need more redundancy literals than we assumed at the beginning. In Theorem 1 we shall use this necessary condition for the formulation of a condition on the truth table of a $\underline{B}$-function.

For Theorem 1 we shall use the following notations. Instead of the notations $a_o, a_1, \ldots, a_o', a_1', \ldots, b_c, b_1, \ldots$ for the independent redundancy variables, we shall use $X_1, X_2, X_3, \ldots X_n$. We denote by γ_i a variable combination number. There are 2^n such numbers, of which some belong to E_w. The index i varies between 0 and $2^n - 1$, and here it is not related to the error e. Those γ_i-numbers for which $\underline{B}(X_1, X_2, \ldots) = 1$, or, for brevity, $\underline{B}(\gamma_i) = 1$, are called the implicant numbers. One number, γ_i, is said to cover another number, γ_j, if all of the 1-positions in γ_j are also 1-positions in γ_i. An implicant number that does not cover any other implicant number is called a prime implicant number. A minimal polynomial is an intersection of all n X_i-variables so that the X_i-letters that correspond to 1's in the corresponding implicant number are affirmed, and those that correspond to 0's are negated. An a-intersection is an intersection of only those (affirmed) X_i-letters that correspond to 1's in the prime implicant number.

THEOREM 1: A condition on a truth table (selected according to Eq. 6) that is necessary and sufficient for the truth table to specify a $\underline{B}$-function (which does not contain negated variables $\tilde{X}_i$) is that all variable combination numbers that cover a prime implicant number are implicant numbers. $\underline{B}$ may be specified as a union of a-intersections only.

4. Networks for a Function of a Single Variable

We want to investigate the least amount of redundancy that is required for a $\underline{B}$-network, insensitive to E_w-errors, for the identity operation B(a) = a. The $\underline{B}$-network contains the redundancy branches $a_o, a_1, \ldots, a_r$. Here there are two γ-numbers (compare section 2) which we denote α and β.

$$\alpha = 111\ldots11 \tag{7}$$

$$\beta = 000\ldots00 \tag{8}$$

We shall also use the set notations:

$$S_{w,\alpha} = \{\alpha_e | e \in E_w\} \tag{9}$$

$$S_{w,\beta} = \{\beta_e | e \in E_w\} \tag{10}$$

$$S_w = \{\gamma_i | \gamma_i \notin S_{w,\alpha},\ \gamma_i \notin S_{w,\beta}\} \tag{11}$$

where α_e and β_e are defined according to Eq. 4, and γ_i is a variable combination number. From Eq. 6 we obtain the requirements:

$$\underline{B}(\gamma_i) = 1 \qquad \gamma_i \in S_{w,\alpha} \tag{12}$$

$$\underline{B}(\gamma_i) = 0 \qquad \gamma_i \in S_{w,\beta} \tag{13}$$

If a prime implicant number $\gamma_i \in S_w$, we must specify, according to Theorem 1, that all covering γ_i-numbers are implicant numbers. The remaining condition on the implicant numbers is that the selected set of prime implicant numbers shall fulfill both the c-criterion and the subrearrangement theorem (4, 5). If there is a set of prime implicant numbers that fulfills these requirements, the amount of redundancy, r, is sufficient for the existence of an E_w-error correcting network. If there is no such set of prime implicant numbers, the amount of redundancy is too small for the existence of an E_w-error correcting network. In investigating the smallest amount of redundancy, we therefore start with the minimum amount for which the sets $S_{w,\alpha}$ and $S_{w,\beta}$ are disjoint. This minimum amount is obtained directly from the Hamming bounds (2) for an error-correcting code of one information position and r check positions. We have

$$w = 1 \qquad (r+1) \geq 3 \tag{14}$$

$$w = 2 \qquad (r+1) \geq 5 \tag{15}$$

$$\vdots$$

$$w \qquad (r+1) \geq 2w + 1 \tag{16}$$

In this case with only one information position there is only one possibility for the formation of the check position content. This is that each check digit equals the information digit, which is in agreement with the redundancy possibility in the network case. For the redundancy selected according to the Hamming bounds (Eq. 16) the set S_w is empty. Hence it is necessary that the set of γ_i-numbers of weight $n - w = r + 1 - w = w + 1$ be prime implicant numbers. This is the lowest bound

on the prime implicant weight:

$$w_{p.i.} \geq w + 1 \tag{17}$$

Let us consider the following set of γ_i-numbers of weight n - w:

T	a_o	a_1	...	a_{w-2}	a_{w-1}	a_w	a_{w+1}	a_{w+2}	...	a_r	
1	0	0		0	0	1	1	1		1	γ_1
1	0	0		0	1	0	1	1		1	γ_2
1	0	0		0	1	1	0	1		1	γ_3
1	0	0		0	0	0	0	1		1	γ_4

The corresponding three L_T-elements (compare refs. 4 and 5) generate the coset element corresponding to γ_4, that is, an element of weight n - w - 2. Hence it must belong to $S_{w,\beta}$. We then conclude that for the redundancy corresponding to the Hamming bounds there exist no redundancy networks. In any case, Eq. 17 is valid. It is equivalent with the statement that the shortest path between the terminal vertices of the redundancy network must contain w + 1 branches. This is obvious, for if there were a path of only w-branches the network could not be insensitive to all e_w-errors. If B = 0, that is, a = 0, and there are no errors, all branches are in state 0. Then the e_w-error that has all its components in the branches of the w-path will give $\underline{B} = 1$.

If, on the other hand, we consider the error-free state corresponding to the implication number α, we must similarly require that the coset c (compare refs. 4 and 5) will contain at least w + 1 elements that are independent in such a way that in any position there is only one 1 in this set of w + 1 elements, except of course for the T-column. Having selected a single path, we must always find w additional paths in the network that are independent in this same way. For, otherwise, we can always find an e_w-error that breaks all paths; that is, $\underline{B} = 0$ instead of 1.

We now immediately obtain the lower bound on the number of branches in an e_w-error insensitive network:

$$n \geq (w+1)^2 \tag{18}$$

for the minimum weight of a prime implicant number is w + 1, and there must be at least w + 1 prime implicant numbers with all their 1's in different positions. For n's fulfilling Eq. 18 as an equality there is always at least one e_w-error insensitive network. The pure series-parallel network with w + 1 independent paths, each containing w + 1 branches, is one.

5. Networks for a Function of Several Variables

Let us consider a Boolian function $B(X_1, X_2, \ldots X_m)$, where X_i stands for nonredundant literals (not letters). We want to protect each literal with extra literal branches. There are 2^m γ-numbers, which we denote α, β, ... Let us consider two γ-numbers that differ only in the X_1-redundancy positions:

$$\alpha = \overbrace{1\ 1 \ldots 1}^{X_1}\ \overbrace{.....}^{X_2 \ldots X_m}$$

$$\beta = 0\ 0 \ldots 0\ .....$$

There must be such a pair for which $\underline{B}(\alpha) \neq \underline{B}(\beta)$. For, otherwise, $\underline{B}(\alpha) = \underline{B}(\beta)$ for any content in the redundancy positions in the common part of α and β. Then the literal X_1 would be redundant for the specification of $\underline{B}$, which is against our assumptions. This is true also if $X_1 = a$ and $X_2 = a^1$, for the two literals a and a^1 must be treated here as independent.

If by α_1 and β_1 we denote only the positions of α and β that correspond to the X_1-redundancy variables, we must accordingly require that the two sets S_{w, α_1} and S_{w, β_1}, defined according to Eqs. 9 and 10, shall be disjoint. The same requirement must evidently be made for each of the redundancy sets that corresponds to the other literals, X_i. It is evident that $\underline{B}(\alpha) = 1$, $\underline{B}(\beta) = 0$, for if $\underline{B}(\beta) = 1$, we must also require, according to Theorem 1, that $\underline{B}(\alpha) = 1$.

We conclude, as we did in section 4, that the minimum weight in the redundancy positions corresponding to X_i of an implicant number is w + 1. Also, there must be at least w + 1 paths containing w + 1 redundancy branches corresponding to each of the X_j $(i \neq j)$ literals that have a 1 in the α-number. There must be no common X_i-redundancy branch in these w + 1 paths. Hence there must be at least $(w+1)^2$ redundancy elements for each of the X_i-literals. In this case of a function of several variables we cannot, as we could in section 4, in general satisfy the realizability requirements (the c-criterion and the subrearrangement theorem of references 4 and 5) with this necessary amount of $m(w+1)^2$ branches. If, however, B has a nonredundant network, then we know that the realizability conditions on $\underline{B}$ must also be satisfied for this necessary amount of redundancy. An individual protection of each branch of the B-network with $(w+1)^2$ branches gives one $\underline{B}$-network. If B has no nonredundant network but a minimum branch network of M branches, evidently a $\underline{B}$-network of $M(w+1)^2$ branches can be constructed. This number can therefore be taken as an upper bound on the minimum number of redundancy branches, n. Hence we have proved Theorem 2.

THEOREM 2: Lower and upper bounds for the minimum number of branches, n, in a w-error correcting contact network for the generation of B are $m(w+1)^2 \leq n \leq M(w+1)^2$,

where m is the number of nonredundant literals of B, and M is the minimum number of branches in a B-generating contact network.

6. General Relay Circuits

If we do not maintain the restriction that the redundancy network shall be of contact type but, instead, allow one subnetwork of contact type to feed some of the coils in the relays of another subnetwork, then we can go beyond the lower bound of Theorem 2. This may be clear from the following example.

We want to design a single-error correcting network for a function B which has a nonredundant contact network of m (>14) branches. Let us make three separate copies of the nonredundant B-network. Let two of these each feed 5 parallel connected relay coils. The third copy shall feed 4 parallel connected coils. Out of these thirteen relay contacts we can realize a single-error correcting contact network for the majority function of three variables. Hence we can construct with 3m + 14 relays a single-error correcting network for B. If m is larger than 14 we have gone beyond the lower bound of Theorem 2. For very large values of m, the additional redundancy for single-error correction approaches a triplication (instead of a quadrupling as for contact networks).

For the realization of the single-error correcting contact network for the majority function of three variables, the two bounds according to Theorem 2 are 12 and 20, respectively. The actual minimum number 14 is close to the lower bound. The synthesis and proof of minimality for the 14-contact network is based on the methods of references 4 and 5.

Stimulating discussions on the subject with Professor P. Elias, of the Massachusetts Institute of Technology, are gratefully acknowledged. The author is indebted to Dr. E. Moore and Dr. D. Hagelbarger of the Bell Telephone Laboratories, Inc., for a generous supply of valuable counterexamples to conjectures. Part of this work was done at the Swedish Research Institute of National Defense.

L. Lofgren

References

1. P. Elias, Computation in the presence of noise, IBM Journal 2, 346 (Oct. 1958).

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

3. L. Lofgren, Automata of high complexity and methods of increasing their reliability by redundancy, Information and Control 1, 127 (1958).

4. L. Lofgren, Irredundant and redundant Boolean branch-networks, Transactions of the 1959 International Symposium on Circuit and Information Theory, Trans. IRE, vol. IT-5, p. 158 (May 1959).

5. L. Lofgren, Solution to the realizability problem for irredundant Boolean branch-networks, J. Franklin Inst. (in press).

(References continued on following page)

6. E. F. Moore and C. E. Shannon, Reliable circuits using less reliable relays. Part I, J. Franklin Inst. 262, 191-208 (1956); Part II, 262, 281-297 (1956).

7. W. W. Peterson and M. O. Rabin, On codes for checking logical operations, IBM Journal 3, 163 (1959).

8. W. V. Quine, A way to simplify truth functions, Am. Math. Month. 62, 627 (1955).

9. C. E. Shannon, A mathematical theory of communication, Bell System Tech. J. 27, 379 (1948).

10. D. Slepian, A class of binary signalling alphabets, Bell System Tech. J. 35, 203 (1956).

11. J. von Neumann, Probabilistic logics and the synthesis of reliable organisms from unreliable components, Automata Studies, edited by C. E. Shannon and J. McCarthy (Princeton University Press, Princeton, New Jersey, 1956).

E. DISTANCE BETWEEN BINARY SEQUENCES CONTAINING ERASURES

The distance between an input and an output sequence of a discrete memoryless channel has been defined for the binary symmetric channel (1) and the p-nary symmetric channel (2). We shall define a metric, or distance, which applies to the input and output sequences of a binary erasure channel that has a finite probability of crossovers.

The binary symmetric channel is a discrete memoryless channel that accepts binary digits at its input and reproduces them at its output, in such a way that the probability that a digit is reproduced incorrectly is p. The channel operates upon successive digits independently of its operation on past digits. The noise is of the same nature as the input and output; that is, a binary output sequence of length n may be defined as the term-by-term mod 2 sum of a binary input sequence and a binary noise sequence that has 1's in those places in which errors have occurred and 0's elsewhere.

Consider a list decoding procedure for the binary symmetric channel. Let there be M message sequences of length n; then there are $N = 2^n$ possible received sequences. The optimum method of decoding consists of forming the term-by-term mod 2 sum of the received sequence and each of the message sequences and thereby generating a set of possible noise sequences. The 1's in the noise sequences represent those places in which the received sequence differs from the message sequences. The received sequence is decoded as that message sequence from which it differs in the least number of places. This is an optimum procedure in the sense that the noise sequence containing the least number of errors is the noise sequence most likely to have occurred as a result of the monotonic relation between the number of errors in a noise sequence and its probability of occurrence.

The number of places in which the message sequence and the output sequence differ (the number of 1's in the noise sequence) is the Hamming distance between them. Thus the optimum decoding procedure amounts to choosing the message sequence that is

"closest" to the received sequence.

The probability of a noise sequence containing x errors is

$$p^x q^{n-x} \qquad p < \frac{1}{2}$$

Let us consider an input sequence disturbed by two noise sequences: n_1 containing x_1 errors, and n_2 containing x_2 errors. It would seem logical to require that if $Pr(n_1) > Pr(n_2)$, then $D(n_1) < D(n_2)$, where D(n) is the distance associated with n. That is,

$$p^{x_1} q^{n-x_2} > p^{x_2} q^{n-x_2}$$

or

$$x_1 \log \frac{p}{q} > x_2 \log \frac{p}{q}$$

That is,

$$x_1 < x_2$$

Thus, we see that distance can be equated directly to the number of errors.

Let us now consider an analogous procedure for a binary symmetric erasure channel with crossovers. This is a channel that accepts binary digits and produces at its output an erasure with probability p_1, a crossover with probability p_2, or the correct digit with probability $p_3 = 1 - p_1 - p_2$. In this channel the noise sequences are of a different nature from the input sequences, but a set of noise sequences resulting from the comparison of a received sequence with each of the message sequences can be found by using the following algebraic procedure. Let the input sequences consist of +1's and -1's; and the output sequences consist of +1's, -1's, and 0's. Let the occurrence of a 0 in the noise sequence represent an erasure and the occurrence of a -1 represent a crossover. Then an output sequence y will be related to an input sequence x as $y = x \odot n$, where n is the noise sequence, and $\odot$ represents term-by-term algebraic multiplication. Since $x \odot x = 1 = (1111\ldots1)$; then $n = x \odot y$.

We now have the set of noise sequences, but it is not obvious which one represents the closest message sequence. Let us again require that if the probability of a noise sequence n_1 (containing x_1 erasures and y_1 crossovers) is greater than the probability of a noise sequence n_2 (containing x_2 erasures and y_2 crossovers) then the distance of n_1 (from the zero-error sequence) must be less than the distance of n_2. Then distance and, most likely, error sequence would be related as in the one-dimensional binary case.

Now, $Pr(n_1) > Pr(n_2)$; that is,

$$p_1^{x_1} p_2^{y_1} p_3^{n-x_1-y_1} > p_1^{x_2} p_2^{y_2} p_3^{n-x_2-y_2}$$

or

$$x_1 \log \frac{p_1}{p_3} + y_1 \log \frac{p_2}{p_3} > x_2 \log \frac{p_1}{p_3} + y_2 \log \frac{p_2}{p_3}$$

or

$$x_1 + ay_1 < x_2 + ay_2 \qquad \text{for } p_2 < p_1 < p_3$$

where

$$a = \frac{\log \frac{p_2}{p_3}}{\log \frac{p_1}{p_3}} > 1$$

Therefore distance may be defined as the weighted sum:

$$\begin{pmatrix}\text{number of erasures}\\ \text{in noise sequence}\end{pmatrix} + \begin{pmatrix}\text{number of crossovers}\\ \text{in noise sequence}\end{pmatrix} a$$

We may now associate with each message sequence in our list decoding the probability of its having produced the received sequence by computing the distance between the received sequence and the message sequences.

This concept of distance may, of course, be extended to any discrete channel having an arbitrary number of inputs and outputs.

D. C. Coll

F. GATING CIRCUITRY FOR ANALOG-DIGITAL CONVERSION

In this Laboratory a DATRAC (Epsco, Inc., trade name) reversible analog-digital converter is employed in a digital facsimile system. The DATRAC, which produces a ten-bit digitalization (exclusive of the sign bit) is used in conjunction with the IBM 704 digital computer, whose magnetic tape equipment has a seven-bit (lateral) format. Previous data-handling equipment permitted the use of only six of the ten bits available from the DATRAC.

We constructed a sequential gate selector that sequentially gates the ten bits onto magnetic tape in the IBM 704 format and thus allows full use of the DATRAC's capabilities. Compatible circuitry, known as the resynthesis gate, was also constructed to permit the DATRAC to recover the original analog variable in sampled form. Although the sequential gate selector and the resynthesis gate are both functioning as desired, the over-all facsimile system is not yet operational. The system's tape equipment or DATRAC seems to be at fault, but further testing will be required to locate the actual source of difficulty.

The author's Master of Science thesis, "Gating Circuitry for Analog-Digital Conversion" (Department of Electrical Engineering, M.I.T., September 1959), describes the sequential gate selector and the resynthesis gate and explains the operation of the facsimile system and the DATRAC.

N. K. H. Tam

G. DESIGN AND EVALUATION OF A LITERATURE RETRIEVAL SCHEME

[This report is based on an abstract of the author's thesis, "Design and Evaluation of a Literature Retrieval Scheme," which was submitted to the Department of Electrical Engineering, M.I.T., in partial fulfillment of the requirements for the degree of Master of Science, June 1959.]

Electrical engineers have become increasingly aware of the problem of information retrieval because the conception of appropriate machines and coding schemes is part of their field. One of the foremost problems is the library search problem. Today's libraries are so large that is has become more and more difficult to find information efficiently and successfully. A literature retrieval scheme remedies the situation by presenting a small set of documents, rather than the whole collection, from which to choose.

In this study we show that all such information retrieval systems have to make some use of human intelligence. Many schemes have been developed for smaller libraries – libraries with 10^4-10^5 documents. They use little, but expensive, information. Cost prohibits the use of similar systems for large libraries. We need a scheme that will make use of machine processes and that is adapted to them. This system should use existing information that can be furnished by human intelligence, in view of the fact that the collection of such information just for retrieval purposes is too expensive. Use should be made of inexpensive "surplus" information.

The scheme that we propose works along the lines proposed by V. Bush. A machine presents us with documents, and our choice of the ones that seem to be pertinent to our problem directs the machine to point out other pertinent information. The contribution of human intelligence comes in through the use of the reference lists contained in the documents. These references define a relation network among the documents, and it is along the branches of this network that the machine is guided in its selections. In addition, the documents are grouped according to their importance – their importance being evaluated by the number of references branching out from a document. The machine proceeds to lead us from level to level, beginning at the top, always determining the next set of possibly pertinent documents by topological procedures in the relation network. After each such step of the machine the human operator has to decide which ones are important to him, and then the machine starts again from this point.

Two hand-imitations of such machine processes were made to evaluate the system and determine the main difficulties. The difficulties are the need for a reading machine for the automatic input of the mass of information contained in the reference lists, and a program that brings all of the different forms of references into a standard form.

The proof of a retrieval scheme is its practical behavior. In the evaluation of any system two figures are of interest: the percentage of relevant documents that we were

not made aware of during the retrieval operations (retrieval loss) and the percentage of relevant documents among all the documents we considered during the operation (efficiency). The results of an experiment that used Stumper's Bibliography on Information Theory, Part I, as the model library, are the following: The two evaluation figures are not independent. If one wants to get all of the pertinent information contained in the collection by means of this procedure, he has to accept a lower efficiency than if he tolerates some loss. The actual figures are the following: less than 5 per cent retrieval loss with an efficiency of 5-15 per cent; less than 30 per cent retrieval loss with an efficiency of 20-30 per cent.

We found no procedures for improving these figures when no more than the information contained in the relation network was used. It seems that our retrieval grammar is not precise enough.

However, for retrieval applications in which these results are sufficient, this very simple scheme, adapted to the use of machine methods as much as possible might be practical, once the two mentioned difficulties are solved.

H. A. Ernst