XIV. PROCESSING AND TRANSMISSION OF INFORMATION*

Prof. D. N. Arden
Prof. E. Arthurs
Prof. W. B. Davenport, Jr.
Prof. P. Elias
Prof. R. M. Fano
Prof. R. G. Gallager
Prof. F. C. Hennie III
Prof. E. M. Hofstetter
Prof. R. A. Howard
Prof. D. A. Huffman
Prof. I. M. Jacobs
Prof. E. F. Moore
Prof. W. F. Schreiber
Prof. C. E. Shannon
Prof. J. M. Wozencraft
J. H. Babcock
G. Cheadle
J. E. Cunningham
P. M. Ebert
G. Einarsson
J. Epstein
E. F. Ferretti
D. H. Friedman
T. J. Goblick, Jr.
J. P. Green, Jr.
U. F. Gronemann
T. S. Huang
F. Jelinek
R. S. Kennedy
L. Kleinrock
C. L. Liu
H. H. Loomis, Jr.
J. L. Massey
C. Moraga
R. C. Norris
E. L. O'Neill
J. W. Pan
R. R. Pfeiffer
L. G. Roberts
J. R. Sklar
I. G. Stiglitz
I. E. Sutherland
O. J. Tretiak
J. Ziv

A. PICTURE PROCESSING

1. LABORATORY EQUIPMENT

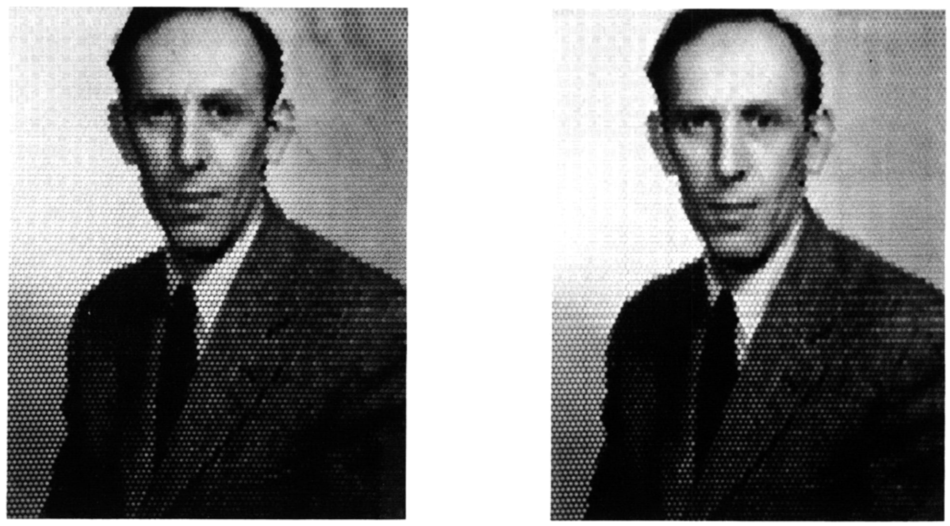
An important determinant of quality in digital image-transmission systems such as our laboratory picture-processing equipment is the optical impulse response of the pickup and display device.

In commerical television practice, quality usually improves monotonically with both electrical and optical focusing. This is so because optimum focus generally means a limiting resolution of not much more than the 525 lines of the standard television picture. Mertz and Grey recognized, in 1934, that for certain sharp pictures and with coarse scanning standards, spurious patterns may develop. At present, with the development of very high resolution scanners on the one hand, and with the renewed interest in lower-resolution TV systems for space applications on the other, this effect is no longer a mere curiosity.

When a continuous image is represented by its values at a discrete set of points in a regular array, the exactness with which the image can be reconstructed from the samples is determined by the relationship between the two-dimensional spatial frequency spectra of the image and the sampling pattern. The spectrum of the sampled image is the convolution of these two spectra. If overlapping of the subspectra occurs (a result of making the sample spacing greater than one-half the shortest wavelength of the spectral components of the image) moiré, or other spurious effects, will appear. To avoid such effects, the image must be bandlimited before sampling, or, that which amounts to the same thing, it must be sampled with something other than an impulse. For example,

*This research was supported in part by Purchase Order DDL B-00337 with Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the joint support of the U.S. Army, Navy, and Air Force under Air Force Contract AF19(604)-7400; and in part by the National Institutes of Health (Grant MH-04737-02).

the function $\frac{\sin \pi x/a}{\pi x/a} \cdot \frac{\sin \pi y/b}{\pi y/b}$, when used for sampling, produces results identical to those obtained by first bandlimiting the image to $\frac{1}{2a}$ and $\frac{1}{2b}$ cycles/unit length in the x- and y-directions, and then sampling with impulses.



(a) (b)

Fig. XIV-1. Experiments in Sampling and Filtering. In an effort to obtain a more isotropic reproduction, a hexagonal sampling pattern of 95×80 picture elements was used. The interpolation function in both pictures was a hexagonal spot, obtained by optical filtering. The sampling function was a very small spot in (a), and a small square spot in (b). The latter adheres to the requirements of the sampling theorem better and seems to give a little higher quality, especially near the eyes.

In the dispaly, a similar problem occurs. Here one must find an interpolation function with which to fill in the space between the samples. Signal theory tells us that simple lowpass filtering will produce an image that is free of sampling frequencies and nearly identical to the original (bandlimited) image. Complications arise, however, because of the nature of the observer. Sharp bandlimiting produces ringing that is highly visible and objectionable. Gradual bandlimiting loses much of the subjective sharpness before the sampling signal is completely suppressed.

We have been doing some preliminary experiments in the search for optimum sampling and interpolation functions. Some of the results are shown in the accompanying photographs.

J. E. Cunningham, U. F. Gronemann, T. S. Huang,
J. W. Pan, O. J. Tretiak, W. F. Schreiber

2. PROGRAMMING

Programs to process pictures on the IBM 7090 computer have been written in the form of pairs of subroutines called for by a simple main program. The pairs are: (a) input and output routines that read and write magnetic tape to and from memory locations specified by the main program (also, provision has been made to read picture data from cards and to print pictures on the off-line printer); (b) two-dimensional filtering and sampling routines that store the samples in memory and produce from these samples a smeared picture; (c) edge detection and synthesis routines that detect edges according to given parameters and synthesize the original edge from the stored information; and (d) some routines for housekeeping purposes, such as clearing the memory to accept a synthetic picture, storing a uniform gray in the memory to eliminate negative brightness, counting routines to record the total number of edges detected, initializing the indices to process a new picture, and various other simple operations.

J. W. Pan

B. DELAY IN SEQUENTIAL MACHINES

1. Introduction

I shall be concerned in this report with the problem of constructing a sequential machine from a set of synchronous elements that have delay inherent in their characteristics in addition to the Boolian function associated with each element. The basic problem will be presented in the form of a conjecture: If one is given a set of elements that operate synchronously and can accept information at the rate of R bits per second per input wire, then one can construct a sequential machine from these elements which will produce the desired output sequence at a rate of R bits per second per output wire.

A very complete discussion of this problem has been presented by D. N. Arden.[1] Much of the background information for my research is contained in Arden's report.

In addition to Arden, Arthurs[2] and Lewis[3] have demonstrated the truth of the conjecture stated above. Arden uses the technique of Regular Algebra,[1] while both Professor Arthurs and Dr. Lewis use the Boolian Algebra of Sequences. In this report, I shall use the latter technique for my presentation.

In more precise language, the basis for my work is the following existence theorem advanced by Arden, Arthurs, and Lewis:

THEOREM 1: Given: A complete set of logic elements that operate synchronously, can accept information at rate R bits per second per wire, and may have inherent delay; and a description of a sequential machine, M, then: it is possible to construct from a set of copies of the complete set of elements a machine that produces the desired output sequence at rate R bits per second per wire. This output sequence, however, may be delayed by some finite amount.

The question that I am investigating is "How do we best apply Theorem 1 to the problem of sequential machine design in order to produce circuits that are good in the sense of minimum delay or minimum complexity?"

In the next section, I shall discuss some of the results of my research into this question, and some answers to questions suggested by this primary problem. In the interest of brevity, I shall not present the proofs of these results.

2. Results of Research

The results presented here deal with two types of sequential machines, the first type being the machine whose current operation depends only on a finite portion of the past of each of the input sequences, and the other type being the machine whose current operation depends on the entire past of the input sequences. I shall refer to the first as feedback-free machines, since any machine of this type can be produced by a network of logic elements with no feedback paths or circles. The second type, I shall refer to as machines with feedback.

The functional notation that I shall use has the following specific properties:

1. A sequence that is written as a function of other sequences shall not have any dependence on time other than that implied by the sequences that are arguments of the function.

2. The value of a sequence at time t, written in functional notation, shall depend only on the values of the sequences at time t which appear as arguments of the function.

A further notation[4] that is of considerable use in this study is the delay operator, D^{τ}. If x(t) represents a sequence, then D^{τ} is defined by the following equation:

$$D^{\tau}x(t) = x(t-\tau). \tag{1}$$

In this report, the dependence on time of a sequence will be understood and thus a sequence x(t) will normally be written as x alone.

Let us now consider some properties of networks that are in the class of feedback-free machines. Here we are concerned with the delay involved in the realization of any function of the type shown in Eq. 2 with logic elements that have inherent delay.

$$y = f(\overline{X}, D\overline{X}, \ldots, D^{\tau}\overline{X}), \tag{2}$$

where $\overline{X}$ is an n-bit vector of binary sequences, y is a binary sequence, and τ is a finite integer.

It was stated earlier that any function that depends on a finite portion of each of the input sequences can be produced by a machine without feedback. However, it is true that such a sequence can always be produced by machines with feedback. It is therefore of interest to know how the delay of a machine to produce such a sequence with feedback compares with the delay of a machine to produce the same sequence without feedback.

LEMMA 1: Given: a complete set of elements, each of which has unit delay and produces a function that depends at time t only on the values of its input sequences at $t - 1$, and a machine constructed from this set of elements that produces the sequence $u' = f(D^{\alpha}\overline{X}, D^{\alpha-1}\overline{X}, \ldots, D^{\beta}\overline{X})$ for time greater than some integer, μ, and that this machine contains feedback, then: there exists a machine constructed from the complete set which produces u', for time greater than μ, which contains no circles external to the elements.

Here we have the result that if one desires to build a machine and it is possible to build it without feedback, then the machine that has the minimum delay will also have no feedback.

This leads us to consider the delay through networks without feedback. Let us define L(M) as the maximum over all functions, of the minimum over all realizations of the delay in the network constructed to produce a function of M sequences.

The following theorems place bounds on the value of L(M) for certain complete sets of elements:

THEOREM 2: Given: The complete set of elements, including a k-input "and" element with unit delay, a k-input "or" element with unit delay, and the availability of all the input variables and their complements, then:

$$L(M) \leq \lceil (M-1)\log_k(2) \rceil + \lceil \log_k(M) \rceil = \alpha(M) \tag{3}$$

where $\lceil A \rceil$ is the smallest integer greater than or equal to A.

THEOREM 3: Given: The availability of any complete set of elements, all with unit delay and with k or fewer inputs, and the availability of all the input sequences and their complements, then:

$$L(M) > (M) \log_k (2) - \log_k (\log_2(2MQ)) = \beta(M), \tag{4}$$

where $Q = 2^{2^k}$.

Theorems 2 and 3 of course imply that:

$$\lim_{M\to\infty} \frac{L(M)}{M} = \log_k (2). \tag{5}$$

L(M) gives us a measure of the delay of the network to produce that function of M sequences which is, in the sense of delay required, the worst function. Theorem 4 gives some idea of the number of functions that obey the bounds on L(M).

THEOREM 4: Given: The availability of any complete set of elements, all with unit delay and with k or fewer inputs, and the availability of all the input sequences and their complements.

Let p(M) be the fraction of all functions of M sequences for which the least-delay realization of the i^{th} function of this set yields delay μ_i, such that:

$$\mu_i \leq \beta(M) - 2, \text{ for all } i.$$

Then

$$\lim_{M\to\infty} p(M) = 0. \tag{6}$$

This theorem means that as M gets very large, nearly all functions have delay that behaves as badly as L(M).

Let us now consider the class of sequential machines that must contain feedback in order to realize the desired function. Such a machine will produce an output sequence that, in general, must depend on the complete past of all input sequences. Such a machine can be specified by a set of equations of the following form:

$$\overline{Y} = \overline{F}(\overline{X}, D\overline{Y}) \tag{7}$$

$$\overline{Z} = \overline{G}(\overline{Y}) \tag{8}$$

where $\overline{X}$ is the m-bit input vector, $\overline{Y}$ is the n-bit state-variable vector, and $\overline{Z}$ is the p-bit output vector.

We shall call Eq. 7 the one-step recursive equation of the machine.

By application of the delay operator to Eq. 7 and substitution of the resulting expression in Eq. 7, the following two-step recursive equation results:

$$\overline{Y} = \overline{F}(\overline{X}, \overline{F}(D\overline{X}, D^2\overline{Y})). \tag{9}$$

Similarly, the τ-step recursive equation may be obtained:

$$\overline{Y} = \overline{F}(\overline{X}, \overline{F}(D\overline{X}, \ldots, \overline{F}(D^{\tau-1}\overline{X}, D^{\tau}\overline{Y})\ldots)). \tag{10}$$

This τ-step recursive equation defines a set of mappings from the set of all possible values of the state-variable vector at time $t - \tau$ to the set of all possible values of the state-variable vector at time t. Let us call this set of "τ-step mappings," Φ_τ.

There is a certain class of sequential machines which appears to have very interesting properties, properties that permit us to say a little more about their construction than we can for the general sequential machine with feedback.

DEFINITION: A sequential machine will be called "input reducible" if and only if there exists a function, $\overline{H}(\overline{X}, D\overline{X})$, such that the two-step recursive equation for the machine can be written in the following form:

$$\overline{Y} = \overline{F}(\overline{H}(\overline{X}, D\overline{X}), D^2\overline{Y}). \tag{11}$$

The following theorem implies a test for the input reducibility of a function and relates this property to the set of mappings on the set of all possible state-variable vectors:

THEOREM 5: A sequential machine is input reducible if and only if $\Phi_2 \subset \Phi_1$.

COROLLARY: A sequential machine is input reducible if and only if, for each ϕ_i, $\phi_j \in \Phi_1$, there exists a mapping $\phi_h \in \Phi_1$ such that $\phi_i \cdot \phi_j = \phi_h$.

A sequential machine that is input reducible has some interesting properties. Since there is no increase in the number of mappings required as we go from the τ-step recursive equation to the $(\tau+1)$-step recursive equation, the complexity of the feedback portion of the network can be fixed. Therefore any characteristics of the original one-step flow table which imply a good choice for the state variables for the machine designed in the ordinary way,[5] also apply to the τ-step flow table. Therefore the problem of selection of state variables must be solved only for the one-step flow table, not each time a higher-order flow table is considered. This fact suggests a construction procedure that may yield a fairly simple machine, with good delay. This procedure and further properties of the input reducible machine will be investigated during the next quarter.

3. Conclusions

Although some interesting results concerning the design of sequential machines using elements with inherent delay have been obtained, there is still much work to be done. The research reported on here has suggested the following questions:

(i) Are there any quantitative results available for machines with feedback which are analogous to the results for feedback-free networks which were presented in Theorems 2, 3, and 4?

(ii) What is a good synthesis procedure for machines that are input reducible?

(iii) Are there any other good design techniques that could perhaps be applied to more general classes of machines with feedback?

H. H. Loomis, Jr.

References

1. D. N. Arden, Delayed-logic and finite-state machines, Quarterly Progress Report No. 62, Research Laboratory of Electronics, M.I.T., July 15, 1961, pp. 163-189.

2. E. Arthurs, Private Communication, M.I.T., February 1961.

3. W. D. Lewis, Microwave Logic, Proc. International Symposium on the Theory of Switching, Part II (Harvard University Press, Cambridge, Mass., 1959), p. 334.

4. D. A. Huffman, The synthesis of linear sequential coding networks, Information Theory, edited by C. Cherry (Academic Press, New York, 1956) pp. 77-95.

5. D. A. Huffman, The synthesis of sequential switching circuits, J. Franklin Inst. 257, 161-190 and 275-303 (1954).

C. THRESHOLD DECODING OF GROUP CODES

1. General Background

Consider a group code consisting of n-tuples

$$\vec{a} = (a_1, a_2, \ldots, a_n) \tag{1}$$

where each $a_i \in GF(q)$ (i.e. each digit is an element in the finite field of q elements) and the symbols a_i, i = 1, 2, ... k, are taken as the information symbols and may be assigned arbitrarily. The remaining a_i, i = k+1, k+2, ... n, are determined from the information symbols by linear equations

$$a_i = \sum_{j=1}^{k} c_j^i a_j \qquad i = k+1, k+2, \ldots n \tag{2}$$

and the $c_j^i \in GF(q)$ are determined by the choice of the code. We assume that the received n-tuple,

$$\vec{r} = (r_1, r_2, \ldots r_n), \tag{3}$$

differs from the transmitted code word, $\vec{a}$, by an additive noise sequence

$$\vec{e} = (e_1, e_2, \ldots, e_n), \tag{4}$$

that is, we assume that

$$r_i = a_i + e_i \qquad i = 1, 2, \ldots, n \tag{5}$$

where the e_i have a probability distribution that depends only on the channel.

Each of the n - k equations in (2) defines a parity set for the code. We may rewrite these equations as

$$\sum_{j=1}^{k} c_j^i a_j - a_i = 0 \qquad i = k+1, k+2, \ldots, n. \tag{6}$$

We define the "parity check," S_i, to be the same sum formed at the receiver, that is,

$$S_i = \sum_{j=1}^{k} c_j^i r_j - r_i \qquad i = k+1, k+2, \ldots, n. \tag{7}$$

After substitution of Eqs. 5 and 6 in Eq. 7, we obtain

$$S_i = \sum_{j=1}^{k} c_j^i e_j - e_i \qquad i = k+1, k+2, \ldots n. \tag{8}$$

The $\{S_i\}$ thus constitutes a set of $n - k$ linear equations in the n unknowns e_i, $i = 1, 2, \ldots n$. The general solution can be written immediately from Eq. 8 as

$$e_i = \sum_{j=1}^{k} c_j^i e_j - S_i \qquad i = k+1, k+2, \ldots n. \tag{9}$$

The general solution has k arbitrary constants, namely, the values of the e_i for $i = 1, 2, \ldots, k$, each of which may have any of q values. Thus there are q^k solutions of Eq. 9. The decoding problem is to find the solution that is most probable from consideration of the channel. For example, with $q = 2$ and a binary symmetric channel, the most probable solution is the one with the fewest nonzero e_i. In practice, it is a very difficult task to find the most probable solution for large k because of the enormous number of solutions to be considered. We now give a simple, but not necessarily optimum, method for finding a probable solution of Eq. 9.

2. Orthogonal Parity Checks

We define a "composite parity check," E_i, to be a linear combination of the parity checks, S_j, that is,

$$E_i = \sum_{j=k+1}^{n} \alpha_j^i S_j \tag{10}$$

where the $\alpha_j^i \in GF(q)$. Substitution of Eq. 8 in Eq. 10 gives an expression for E_i of the form

$$E_i = \sum_{j=0}^{n} \beta_j^i S_j \tag{11}$$

where, again, the $\beta_j^i \in GF(q)$.

We define a set of N composite parity checks, E_i, $i = 1, 2, \ldots N$, to be <u>orthogonal on e_m</u> if

$$\beta_m^i = 1 \qquad i = 1, 2, \ldots, N \tag{12}$$

and

$$\beta_j^i \neq 0 \qquad \text{for, at most, one } j \neq m \text{ (i fixed).} \tag{13}$$

3. Majority Decoding

THEOREM 1: Provided that no more than $\frac{N}{2}$ of the e_j, $j = 1, 2, \ldots n$, which appear with nonzero coefficients in a set $\{E_i\}$ of N equations orthogonal on e_m, are nonzero, then e_m is given correctly by that value of GF(q) which is assumed by the greatest fraction of the set $\{E_i\}$. (Assume that $e_m = 0$ if there is no such value as stated in the theorem and zero is one of the values with most occurrences in the set $\{E_i\}$.)

PROOF 1: Let $e_m = V$ and suppose that all other e_j appearing with nonzero coefficients in the equations for the $\{E_i\}$ are zero. Then from Eqs. 11 and 12, it follows that all members of the $\{E_i\}$ have value V. If $V \neq 0$, then under the conditions of Theorem 1, fewer than $\frac{1}{2}N$ of the e_j $(j \neq m)$ are nonzero and hence, by Eq. 13, more than one-half of the members of $\{E_i\}$ still have value V. If $V = 0$, at most $\frac{1}{2}N$ of the e_j $(j \neq m)$ are nonzero and hence at least one-half of the members of $\{E_i\}$ have the value $V = 0$. This proves the assertion in Theorem 1.

We refer to the decoding rule of Theorem 1 as "majority decoding" of orthogonal parity checks.

4. A Posteriori Probability Decoding

Majority decoding is inefficient in the sense that it makes no use of the channel statistics, that is, of the probability distribution of the e_j. We now give a decoding algorithm that makes the best possible use of the information contained in the set $\{E_i\}$ of N equations that are orthogonal on e_m. We assume that the noise is independent from digit to digit, but not necessarily stationary. Using average error probability as the criterion of goodness, we seek that value of e_m, call it V_m, for which

$$\Pr\left(e_m = V_m \,|\, \{E_i\}\right) \geq \Pr\left(e_m = V \,|\, \{E_i\}\right), \tag{14}$$

for all $V \in GF(q)$. Using Baye's rule, we may rewrite Eq. 14 to obtain

$$\Pr\left(\{E_i\} \,|\, e_m = V_m\right) P_r\left(e_m = V_m\right) \geq \Pr\left(\{E_i\} \,|\, e_m = V\right) \Pr\left(e_m = V\right). \tag{15}$$

Because of the orthogonality on e_m of the $\{E_i\}$ and the digit-to-digit independence of the noise sequence, it follows that

$$\Pr\left(\{E_i\} \,|\, e_m = V\right) = \prod_{i=1}^{N} \Pr\left(E_i \,|\, e_m = V\right). \tag{16}$$

To simplify the notation, let

$$\Pr\left(E_i \,|\, e_m = V\right) = P_V(E_i), \qquad i = 1, 2, \ldots N, \tag{17}$$

and, for the sake of convenience, let

$$\Pr(e_m=V) = P_V(E_o). \tag{18}$$

Then substituting Eq. 16 in Eq. 15, taking logarithms, and using the reduced notation of Eqs. 17 and 18, we can reword the decoding rule of Eq. 14 as: Choose e_m to be that value $V \in GF(q)$ for which

$$\sum_{i=0}^{N} \log P_V(E_i) \tag{19}$$

is a maximum. We call this decoding algorithm the "a posteriori probability" decoding rule for orthogonal parity checks.

5. Threshold Decoding

For $q = 2$ and the binary symmetric channel, the majority and a posteriori probability decoding rules specialize to a very simple form. Let $p_o = 1 - q_o$ be the channel transition probability. Let $p_i = 1 - q_i$ be the probability that an odd number of the e_j, $(j \neq m)$ that have nonzero coefficients in the expression for E_i, have value one. If there are n_i such variables in E_i, then

$$p_i = \frac{1}{2}\left[1-(1-2p_o)^{n_i}\right]. \tag{20}$$

The a posteriori probability decoding rule then reduces to: Choose $e_m = 1$ if, and only if, the sum of the members of the $\{E_i\}$ (as real numbers) with each number weighted by a constant factor, $2 \log \frac{q_i}{p_i}$, exceeds the threshold value $\sum_{i=0}^{N} \log \frac{q_i}{p_i}$. Similarly, the majority decoding rule becomes: Choose $e_m = 1$ if, and only if, the sum of the members of the $\{E_i\}$ (as real numbers) exceeds the threshold value $\frac{1}{2}N$.

Because of the similarity between these decoding rules, we use the generic term "threshold decoding" to describe either majority or a posteriori probability decoding of orthogonal parity checks. (We shall use this term regardless of the choice of q and the channel.)

6. Implementation of Threshold Decoding

THEOREM 2: The parity checks, S_j, are formed at the receiver by subtracting the received sequence $(r_1, r_2, \ldots, r_n)$ from the code word produced by encoding $r_i, r_2, \ldots r_k$ as information symbols. The digits in positions $k + 1$ through n are, respectively, S_{k+1} through S_n.

PROOF 2: According to Eq. 2, the encoding process will assign the value

$$\sum_{i=1}^{k} c_i^j r_i$$

to position j for j = k+1, k+2, ... n. After subtraction of the received sequence, the value

$$\sum_{i=1}^{k} c_i^j r_i - r_j$$

is assigned to position j for j = k+1, k+2, ... n. But from Eq. 7, this sum is, by definition, S_j.

We can now describe the general implementation of threshold decoding. First, the parity checks, S_j, are produced in the manner indicated by Theorem 2. Next, that value of m, m = 1, 2, ... k, is selected for which the largest number of parity checks that are orthogonal on e_m can be formed by proper choice of the a_j^i in Eq. 10. The value of e_m is then computed by one of the threshold decoding algorithms. e_m can then be eliminated from all the parity checks, S_j, in which it appears. The same process is then repeated m - 1 times until all of the e_j j = 1, 2, ..., k have been determined. Finally, the information symbols a_j, j = 1, 2, ... k are computed from Eq. 5.

We have already indicated[1] how threshold decoding can be easily instrumented for convolutional codes. We are, at present, studying its application to block codes and attempting to find bounds on the ultimate capability of the method.

J. L. Massey

References

1. J. L. Massey, Majority decoding of convolutional codes, Quarterly Progress Report No. 64, Research Laboratory of Electronics, M.I.T., January 15, 1962, pp. 183-188.

D. INFORMATION FLOW IN LARGE COMMUNICATION NETS

In continuing our research[1] on the problems of information flow in large communication nets results have been obtained (for a single node) for two classes of queue disciplines: priority queueing, and time-shared servicing. A law of conservation has been proved which constrains the allowed variation in the average waiting times over the set of priority classes.

1. Priority Queueing

For priority queueing, the input traffic is broken up into P priority classes. Units from priority class p (p=1, 2, ..., P) arrive in a Poisson stream with an average rate

λ_p units per second; each unit from this priority class has a total required processing time selected independently from an exponential distribution, with mean $1/\mu_p$. We define

$$\rho_p = \lambda_p/\mu_p,$$

$$\rho = \sum_{p=1}^{P} \rho_p,$$

and

$$W_o = \sum_{p=1}^{P} \rho_p/\mu_p.$$

The priority structure is such that a unit from the p^{th} priority class entering the queue at time T is assigned a number b_p, where $0 \leq b_1 \leq b_2 \leq \ldots \leq b_P$. The priority $q_p(t)$, at time t, associated with such a unit is

$$q_p(t) = (t-T)\, b_p.$$

The effect of this priority assignment is to increase a unit's priority in proportion to the time that elapsed since that unit's arrival at the system (referred to as a delay-dependent priority system).

Let us define W_p to be the expected value of the time spent in the queue for a unit from the p^{th} priority class. We then state the following theorem.

THEOREM 1: For the delay-dependent priority system described above, and for $0 \leq \rho < 1$,

$$W_p = \frac{\dfrac{W_o}{1-\rho} - \displaystyle\sum_{i=1}^{p-1} \rho_i W_i \left(1 - \frac{b_i}{b_p}\right)}{1 - \displaystyle\sum_{i=p+1}^{P} \rho_i \left(1 - \frac{b_p}{b_i}\right)}.$$

From a designer's point of view, the introduction of the P independent quantities b_p is an asset. Consider the problem of a system designer who is faced with assigning some priority structure to a queueing system. Let us assume that he is given the quantities λ_p, μ_p, and P, that is, he is given the desired input traffic and partitioning. From these parameters, he can easily calculate ρ. With the free parameters b_p, he can then attain any value for W_p (for this value of ρ) within broad limits. Without these additional degrees of freedom, the set W_p would be fixed (as for a commonly used priority structure[2] for which $q_p(t) = a_p$ and a_p is independent of time).

2. A Conservation Law

As one might expect, there is a certain trade-off of waiting time among the various priority classes. In particular, let us define a class of queueing disciplines as follows:

(a) Arrival statistics are Poisson with an average arrival rate λ_p for the p^{th} priority class.

(b) Service-time statistics are arbitrary with mean $1/\mu_p$ for the p^{th} priority class.

(c) All units remain in the system until completely served.

(d) The service facility is never idle if there are any units in the system.

(e) Pre-emption (the replacement of a low-priority unit in service by a higher priority unit) is allowed only if the service-time distributions are exponential, and if upon re-entry into the service facility the low-priority unit continues from the point at which its service was interrupted.

THEOREM 2: For any queue discipline and any fixed-arrival and service-time distributions that are subject to the restrictions stated above

$$\sum_{p=1}^{P} \rho_p W_p = \text{constant} = \begin{cases} \dfrac{\rho}{1-\rho} V & \rho < 1 \\ \infty & \rho \geq 1 \end{cases}$$

where

$$V = \frac{1}{2} \sum_{p=1}^{P} \lambda_p E(t_p^2)$$

and

$$E(t_p^2) = \text{second moment of the service-time distribution for priority class } p.$$

This conservation law constrains the allowed variation in the average waiting time for any queue discipline that falls into this wide class.

3. Time-Shared Servicing

For a time-shared servicing facility, we consider time to be quantized into intervals, each of which is Q seconds in length. At the end of each time interval, a new unit arrives in the system with probability λQ (result of a Bernoulli trial); thus the average number of arrivals per second is λ. The service time of a newly arriving unit is chosen independently from a geometric distribution so that for $\sigma < 1$,

$$s_n = (1-\sigma)\,\sigma^{n-1} \qquad n = 1, 2, 3, \ldots$$

where s_n is the probability that a unit's service time is exactly n time intervals long.

The procedure for servicing is as follows: A unit upon arrival joins the end of the queue, and waits on line in a first come first served fashion until it finally arrives at the service facility. The server picks the next unit in the queue and performs a unit of service upon it. At the end of this time interval, the unit leaves the system if its service is finished; if not, it joins the end of the queue with its service partially completed. Obviously, a unit whose service time is n intervals long will be forced to join the queue a total of n times before its service is completed. Another assumption must now be made regarding the order in which events take place at the end of a time interval. We shall assume that the unit leaving the service facility is allowed to join the tail of the queue before the next unit arrives at the queue from outside the system (referred to as a late-arrival system). The case with reversed order has also been solved, but will not be reported on here, since the results are not essentially different.

Upon arrival, a unit finds some number of units, m, in the system. The expected value, E(m), of the number m is known[3] to be

$$E(m) = \frac{\rho}{1-\rho}\sigma$$

where

$$\rho = \frac{\lambda Q}{1-\sigma}.$$

We are now ready to state the following theorem.

THEOREM 3: The expected value, T_n, of the total time spent in the late-arrival system for a unit whose service time is nQ seconds, is

$$T_n = \frac{nQ}{1-\rho} - \frac{\lambda Q^2}{1-\rho}\left\{1 + \frac{(1-\sigma a)(1-a^{n-1})}{(1-\sigma)^2\,(1-\rho)}\right\}$$

where

$$a = \sigma + \lambda Q.$$

Now, instead of the round-robin type of structure just described, we shall consider a strict first come first served system in which each unit waits for service in order of arrival, and, once it is in service, each unit remains until it is completely serviced. Then for T_n defined as before, we state the following theorem.

THEOREM 4: The expected value, T_n, of the total time spent in the first come first served system for a unit whose service time is nQ seconds, is

$$T_n = \frac{1}{1-\sigma}QE(m) + nQ$$

where E(m) is as defined above.

Now if one wishes an approximate solution to the round-robin system, one might argue as follows: Each time a unit (the tagged unit, say) returns to the queue, it finds E(m) units in the system ahead of it (this is the approximation). Each of these units will spend Q seconds in the service facility before the tagged unit arrives at the service facility. Since the tagged unit must go through this process n times, the total time that it spends in the queue is nQE(m). Also, it spends exactly nQ seconds in the service facility itself. Thus, our approximate solution, T'_n, turns out to be

$$T'_n = nQE(m) + nQ.$$

Comparing this solution with the result for the first come first served case, we see that there is a critical value of n, say n_{crit}, at the point $n_{crit} = \frac{1}{1-\sigma}$. In fact, we observe that the quantity $\frac{1}{1-\sigma}$ is merely the mean value, $\bar{n}$, of the number of service intervals required by a unit. Thus, the approximate solution shows us that units whose service time is greater (or less) than the average time, $\bar{n}Q$, spend more (or less) time in the round-robin system than in a strict first come first served system, that is, units with short service-time requirements are given preferential treatment over units with longer requirements. The fact that the critical length is equal to the average length is a surprisingly simple result. It has also been shown that the approximation is excellent.

It is interesting to note that the round-robin and first come first served disciplines offer an example of the validity of the conservation law. That is, if we define $W_n = T_n - nQ$, which is the average waiting time in the queue, then it is a simple algebraic exercise to show that

$$\sum_{n=1}^{\infty} \rho_n W_n \text{ (first come first served)} = \sum_{n=1}^{\infty} \rho_n W_n \text{ (round-robin)} = \frac{Q\rho^2\sigma}{(1-\rho)(1-\sigma)}$$

where

$$\rho_n = \rho s_n = \rho(1-\sigma)\,\sigma^{n-1}.$$

L. Kleinrock

References

1. L. Kleinrock, Information flow in large communication networks, Quarterly Progress Report No. 62, Research Laboratory of Electronics, M.I.T., July 15, 1961, pp. 162-163.

2. A. Cobham, Priority assignments in waiting line problems, Operations Research, Vol. 2, pp. 70-76, 1954.

3. J. R. Jackson, Some problems in queueing with dynamic priorities, Naval Research Logistics Quarterly, Vol. 7, p. 235, 1960.

E. DETECTION OF SIGNALS WITH RANDOM PHASE AND AMPLITUDE

1. Introduction

The problem of detecting a known signal in additive white Gaussian noise, which has been called by Siebert[1] the "Canonic Detection Problem," is well known: The optimum reciever computes the likelihood ratio Λ and compares it with a threshold to decide whether or not a signal was transmitted. The likelihood ratio is defined as

$$\Lambda = \frac{p(z(t)/S)}{p(z(t)/N)}, \tag{1}$$

where $p(z(t)/S)$ and $p(z(t)/N)$ are the conditional probability densities of receiving a particular waveform $z(t)$, when a signal is or is not actually transmitted. For white Gaussian noise it is easy to show that

$$\Lambda = \exp\left(\frac{2}{N_o}\int_0^T z(t)\, s(t)\, dt - \frac{1}{N_o}\int_0^T s^2(t)\, dt\right), \tag{2}$$

where $s(t)$ denotes the transmitted signal (of duration T), and N_o is the power density spectrum of the noise.

Equation 2 yields the receiver structure

$$\int_0^T z(t)\, s(t)\, dt \quad \underset{\text{no signal}}{\overset{\text{signal}}{\gtrless}} \quad \text{constant}. \tag{3}$$

For a more complete treatment of this problem see Helstrom.[2]

The solution when the phase of the signal is completely random at the receiver, that is, when the phase of the signal carrier can take all values between 0 and 2π with equal probability, is also well known. The receiver simply takes the envelope of the correlation integral in (3) before comparing it with the threshold. A similar solution obtains when the amplitude is a random variable independent of the phase.

In this report we deal with the situation in which the receiver has statistical knowledge about the phase and amplitude of the received signal and knows the shape of the transmitted signal exactly. Thus two parameters, the phase ϕ of the carrier and the amplitude factor A, are random variables, but all other properties of the received signal are known, apart from additive noise.

2. Detection of Signals with Random Phase and Amplitude

The signals used in a practical communication situation often occupy only a narrow band around the carrier frequency. Our assumption of random phase and amplitude, for instance, can apply to communication through a slowly varying channel, for which the receiver has estimates of the channel conditions from previous transmissions.

a. Complex Representation of Real Waveforms

When dealing with bandpass functions it is convenient to represent the waveforms as the real part of complex-valued functions. (See, for instance, Dugundji.[3]) We write the transmitted signal s(t) as

$$s(t) = \mathrm{Re}\left[\xi(t)\, e^{j\omega_o t}\right] = s_c(t) \cos \omega_o t - s_s(t) \sin \omega_o t \tag{4}$$

where $\xi(t)$ is called the pre-envelope, and $s_c(t) = \mathrm{Re}\,[\xi(t)]$ and $s_s(t) = \mathrm{Im}\,[\xi(t)]$ the quadrature components, of the signal s(t).

We shall use correlation integrals involving complex waveforms, and it can be shown that for narrow-band signals

$$\mathrm{Re}\left[\int_0^T \xi^*(t)\ \eta(t)\ dt\right] = \int_0^T (s_c(t)n_c(t)+s_s(t)n_s(t))\ dt \simeq 2 \int_0^T s(t)\ n(t)\ dt \tag{5}$$

$$\mathrm{Im}\left[\int_0^T \xi^*(t)\ \eta(t)\ dt\right] = \int_0^T (s_c(t)n_s(t)-s_s(t)n_c(t))\ dt \simeq 2 \int_0^T s(t)\ n(t)\ dt \tag{6}$$

$$\left|\int_0^T \xi^*(t)\ \eta(t)\ dt\right| \approx 2 \cdot \text{envelope of} \int_0^T s(t)\ n(t)\ dt \tag{7}$$

where

$$s(t) = \mathrm{Re}\left[\xi(t)\, e^{j\omega_o t}\right]$$

$$n(t) = \mathrm{Re}\left[\eta(t)\, e^{j\omega_o t}\right].$$

Here, an asterisk denotes complex conjugate, and the circumflex represents the Hilbert transform ($n(t) \approx n\left(t - \frac{\pi}{2\omega_o}\right)$ for narrow-band functions).

It is possible to compute these integrals by the matched filters represented in Fig. XIV-2.

b. Statement of the Problem

The transmitter sends a narrow-band signal s(t), which we represent by its pre-envelope $\xi(t)$ according to (4). White Gaussian noise is added at the receiver and, in addition, random channel fluctuation changes the phase and amplitude of the signal. The received signal z(t) is then

$$z(t) = \mathrm{Re}\left[\zeta(t)\ e^{j\omega_o t}\right] = \mathrm{Re}\left[A\, e^{-j\phi}\xi(t)\, e^{j\omega_o t} + \eta(t)\, e^{j\omega_o t}\right], \tag{8}$$

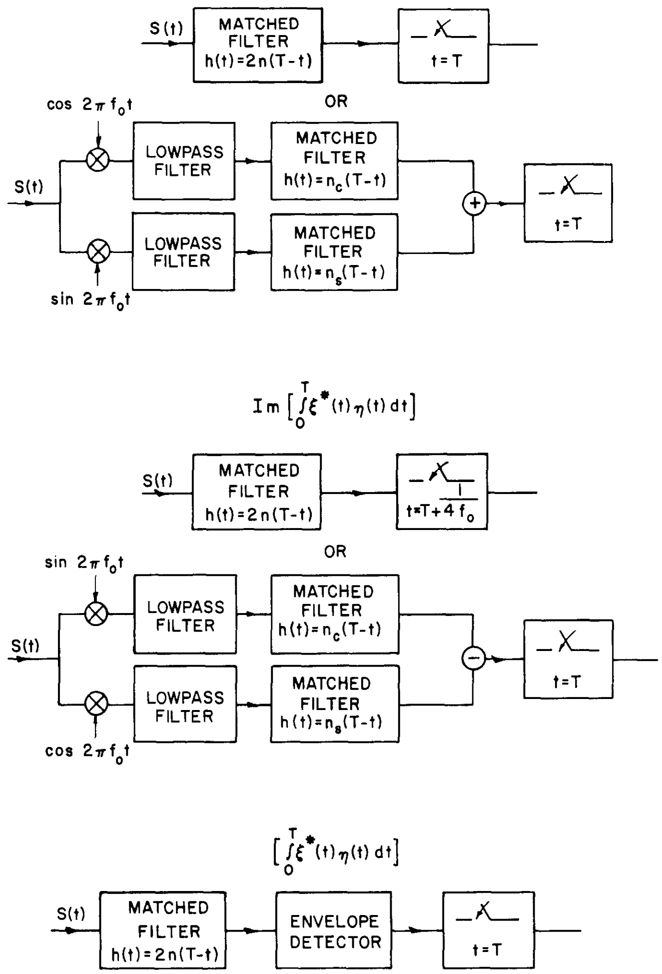


Fig. XIV-2. Computation of correlation integrals by matched filters.

where $\eta(t)$ represents the noise, and the amplitude A and the phase angle ϕ are random variables whose distribution is known to the receiver. We write $p(A, \phi)$ for the joint probability density of A and ϕ.

It is convenient to use Cartesian coordinates to represent the phase and amplitude, and we define the quadrature components of the channel as

$$\left.\begin{array}{l} x = A \sin \phi \\ y = A \cos \phi \end{array}\right\} . \tag{9}$$

Without loss of generality, we can assume that ϕ has zero mean, since a value different from zero is accounted for by changing the phase reference at the receiver. An arbitrary fixed gain or attenuation in the channel is immaterial, and we can assume therefore that the variance in, say, the y-direction is equal to unity. With these assumptions, if $p(A, \phi)$ is symmetric with respect to ϕ, we have

$$\left.\begin{aligned} E[x] &= 0 \quad ; \quad E[x^2] = c_\phi^2 \\ E[y] &= c_a \quad ; \quad E[(y-c_a)^2] = 1 \end{aligned}\right\} \tag{10}$$

where c_ϕ^2 and c_a are constants.

We call the corresponding probability density function of the quadrature components $p_o(x, y)$.

c. The Likelihood Ratio

To find the likelihood ratio, we first compute the conditional ratio, given that A and ϕ are known. Completely analogously with (1) and (2), we obtain

$$\Lambda_{A,\phi} = \frac{p(z(t)/S, A, \phi)}{p(z(t)/N, A, \phi)} = \exp\left(-\frac{1}{2N_o}\int_0^T |\zeta(t)-A\,e^{-j\phi}\xi(t)|^2\, dt + \frac{1}{2N_o}\int_0^T |\zeta(t)|^2\, dt\right) \tag{11}$$

According to (5) the energy in the transmitted waveform is

$$E = \int_0^T s^2(t)\, dt = \frac{1}{2}\int_0^T |\xi(t)|^2\, dt. \tag{12}$$

It is convenient to define two normalized correlation integrals

$$\left.\begin{aligned} U &= \operatorname{Im} \frac{1}{2E}\int_0^T \zeta^*(t)\, \xi(t)\, dt \\ V &= \operatorname{Re} \frac{1}{2E}\int_0^T \zeta^*(t)\, \xi(t)\, dt \end{aligned}\right\}. \tag{13}$$

It is then possible to write the conditional likelihood ratio

$$\Lambda_{A,\phi} = \exp\left\{\frac{E}{N_o}[2UA\sin\phi + 2VA\cos\phi - A^2]\right\} \tag{14}$$

To obtain Λ we average $\Lambda_{A,\phi}$ over the possible values of A and ϕ.

$$\Lambda = \iint \Lambda_{A,\phi}\, p(A, \phi)\, dA\, d\phi. \tag{15}$$

Thus the variables U and V are sufficient statistics in the sense that they contain all of the information about the received waveform needed for the decision. It is possible to obtain U and V by a filter matched to s(t), as shown in Fig. XIV-2. The integral (15) represents the way in which U and V are combined to get the likelihood ratio. If we define the weighting function

$$W(U,V) = \ln \Lambda = \ln \int_0^\infty \int_0^{2\pi} \exp\left\{\frac{E}{N_o}\left[2UA\sin\phi+2VA\cos\phi-A^2\right]\right\} p(A,\phi)\,dA\,d\phi \tag{16}$$

the receiver structure is

$$W(U,V) \underset{\text{no signal}}{\overset{\text{signal}}{\gtrless}} \text{constant}.$$

We can make the following geometrical interpretation of this decision rule: Determine the point (U, V), and decide that the signal is present if this point falls outside the contour determined by

$$W(U,V) = \text{constant}. \tag{17}$$

Otherwise, decide that no signal is present (see Fig. XIV-4).

d. The Weighting Function W(U, V)

Using the probability density in rectangular coordinates (9), we can write the weighting function

$$W(U,V) = \ln \int_{-\infty}^\infty \int_{-\infty}^\infty \exp\left\{\frac{E}{N_o}\left[2Ux+2Vy-x^2-y^2\right]\right\} p_o(x,y)\,dx\,dy. \tag{18}$$

The form of W(U, V) depends on $p_o(x,y)$. It is not possible, in general, to solve the integral in (17), but nevertheless it is possible to determine certain properties of W(U, V). A proof has been given elsewhere[4] that W(U, V) has a minimum at some finite point in the U, V plane, unless the whole probability mass of $p_o(x,y)$ is located in a half-plane. It can also be shown that the function is monotonically increasing from the minimum point in all directions. The contour line given by (17), in general, is closed, and it surrounds a simple region. If $p_o(x,y)$ is symmetric around a line through the origin in the x, y plane, W(U, V) is symmetric around the same line in the U, V plane. The particular form of W(U, V) when the variables x and y have Gaussian distributions is derived below.

e. The Sufficient Statistics U and V

The random variables U and V contain all of the information that the receiver needs to make its decision. The receiver makes an error if the point (U, V) falls outside the contour given by the weighting function when only noise was received, or if it falls inside the contour when the signal is present.

When only noise is received we have

$$\left.\begin{aligned} U_N &= \mathrm{Im} \\ V_N &= \mathrm{Re} \end{aligned}\right\} \frac{1}{2E} \int_0^T \eta^*(t)\, \xi(t)\, dt. \tag{19}$$

We see that U_N and V_N are obtained by a linear operation on a Gaussian process, and therefore are Gaussian variables. It is easy to show that

$$\left.\begin{aligned} & E[U_N] = E[V_N] = E[U_N V_N] = 0 \\ & E[U_N^2] = E[V_N^2] = \frac{N_o}{2E} \end{aligned}\right\}. \tag{20}$$

When the signal is present we have

$$\begin{aligned} U_S &= \mathrm{Im} \\ V_S &= \mathrm{Re} \end{aligned} \left\{ \frac{1}{2E} \int_0^T \left[A\, e^{-j\phi} \xi(t) + \eta(t)\right]^* \xi(t)\, dt \right. \tag{21}$$

which we can write as

$$\left.\begin{aligned} U_S &= x + U_N \\ V_S &= y + V_N \end{aligned}\right\}. \tag{22}$$

In this case the decision variables are the sums of two random variables and, if the noise is independent of the phase and amplitude, the distribution for U_S and V_S is obtained by convolving $p_o(x, y)$ and the Gaussian distribution for U_N and V_N.

3. Examples

a. Completely Known Signals

When A and ϕ are known, the normalized density $p_o(x, y)$ is a unit impulse located on the point (0, 1), and from (18) we have

$$W(U, V) = \frac{E}{N_o} [2V-1]. \tag{23}$$

The decision boundary in the U, V plane is a line parallel to the U-axis, and from (17) it is easy to see that the resulting decision rule is the same as that given by (3). See Fig. XIV-3.

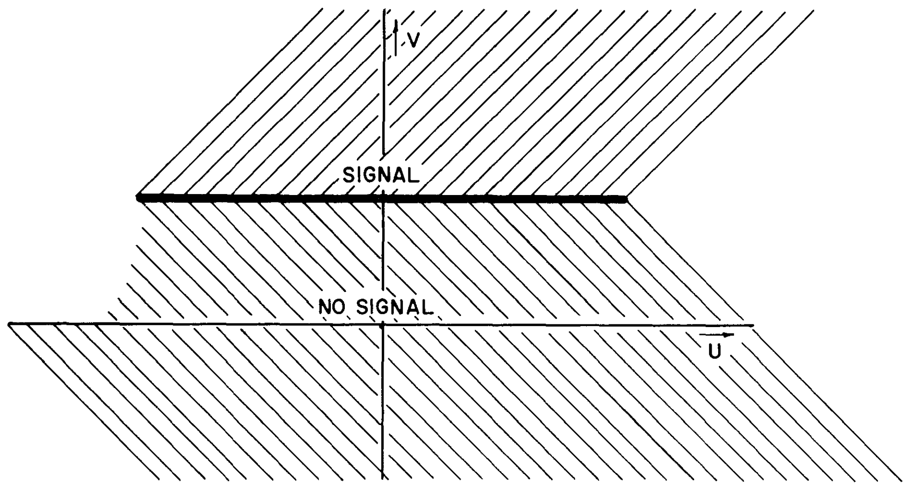


Fig. XIV-3. Completely known signal.

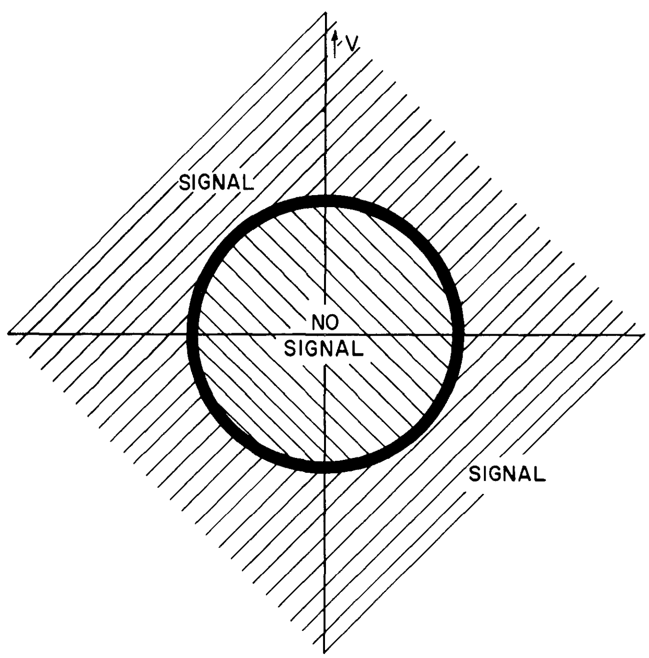


Fig. XIV-4. Completely random phase.

b. Completely Random Phase

If the phase is independent of the amplitude and has a uniform distribution, we have

$$W(U,V) = \ln \int_0^\infty \exp\left(-\frac{EA^2}{N_o}\right) p(A) \left[\int_0^{2\pi} \exp\left\{\frac{2EA}{N_o}(U \sin\phi + V\cos\phi)\right\} \frac{d\phi}{2\pi}\right] dA \tag{24}$$

The integration over the phase angle gives Bessel functions, and we obtain

$$W(U,V) = \ln \int_0^\infty \exp\left(-\frac{EA^2}{N_o}\right) p(A)\, I_o\left(\frac{2EA}{N_o}\sqrt{U^2+V^2}\right) dA, \tag{25}$$

where $I_o(\,)$ is the Bessel function of the first kind, zero-order, with imaginary argument. The decision boundary, given by

$$\sqrt{U^2+V^2} = \text{constant}, \tag{26}$$

is a circle centered around the origin. See Fig. XIV-4. We can write

$$\sqrt{U^2+V^2} = \frac{1}{4E^2}\left|\int_0^T \zeta^*(t)\,\xi(t)\,dt\right| \tag{27}$$

and, from Fig. XIV-2, see that an easy way to compute the right-hand side is by sampling the envelope of the output from a matched filter. This is a well-known result from statistical decision theory.

c. Modified Rician Distribution

Another case for which it is possible to solve the integral in (18) and obtain W(U, V) in closed form exists when x and y are independent Gaussian variables. When both have zero mean and equal variance, the amplitude A has a Rayleigh distribution. Rice,[5] among others, has studied the case for which the mean is not zero but the variances still are equal. We make the extension that the variances need not be equal, and have

$$p_o(x,y) = \frac{1}{\sqrt{2\pi}\,c_\phi} \exp\left(-\frac{x^2}{2c_\phi^2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-c_a)^2}{2}\right). \tag{28}$$

By completing the squares in the exponent in (18) and using (28), we get

$$W(U,V) = \frac{U^2}{a^2} + \frac{\left(V+\frac{c_a}{d}\right)^2}{b^2} - \frac{c_a^2}{2} - \frac{1}{2}\ln\left[(dc_\phi^2+1)(d+1)\right], \tag{29}$$

where

$$d = \frac{2E}{N_o}$$

$$a = \frac{\sqrt{2(c_\phi^2 d+1)}}{dc_\phi}$$

$$b = \frac{\sqrt{2(d+1)}}{d}$$

The decision boundary is an ellipse centered around the point c_a/d. This situation is pictured in Fig. XIV-5, in which the joint probability distributions for U and V under both hypotheses are also drawn. The situation shown here corresponds to a Baye's solution for equal costs and signal probability 1/2.

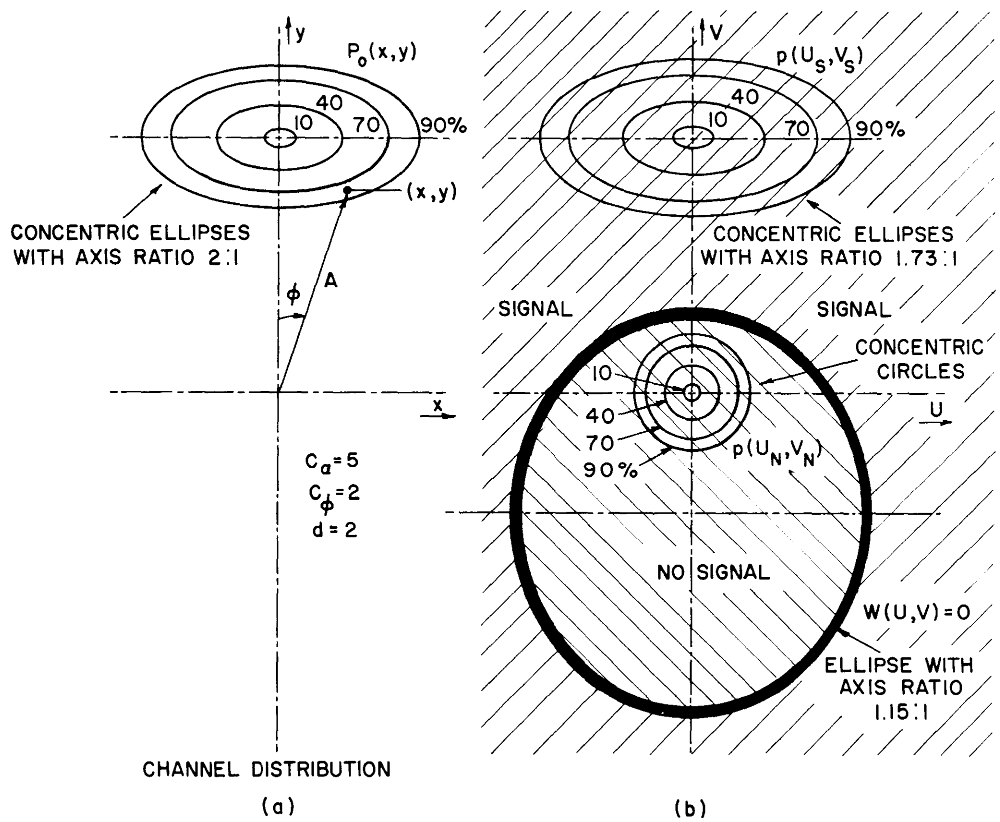


Fig. XIV-5. (a) Probability densities and decision boundary for the modified Rician case. (b) Decision boundary and distributions for decision variables. (Numbers on contour lines for densities represent the probability mass inside the contour.)

When $c_\phi = 1$ (the Rician distribution for A) the decision boundary is a circle just as for the case of completely unknown phase, but it is now centered around a point on the negative V-axis.

4. Conclusions

For the case of a completely known signal, the optimum receiver samples the output of a matched filter and compares it with a threshold to decide whether or not a signal is present. Our results show that when the phase and amplitude is only known in a statistical sense, the receiver samples the matched filter twice, and determines whether the point having these two coordinates (U, V) falls within or outside a boundary. The boundary line is given in terms of the weighting function W(U, V), which is determined by the probability distribution for the phase and amplitude. We have remarked upon some of the general properties of W(U, V), and derived it in closed form in some special cases. The form of W(U, V) seems to be relatively insensitive to the detailed form of the phase and amplitude distribution; we conjecture that the function given in (29) should work well even in cases for which it is not optimum.

The receiver makes an error when the point (U, V) falls on the wrong side of the boundary. For the Rician case ($c_\phi=1$), the error probability has been computed by Turin.[6]

We have assumed that the additive noise was white. For non-white noise, the only necessary modification is to match the filter to the signal given by an integral equation involving the autocorrelation function of the noise. See Helstrom[2] for further details.

It is easy to extend the simple detection problem considered here to the case in which one of a set of possible waveforms is transmitted. In this case the decision variables U and V are computed for each possible waveform and substituted in the weighting function W(U, V). The receiver decides which waveform was transmitted by comparing the different values of W(U, V). We have dealt with this situation for communication over a random multipath channel.[4]

I wish to thank Professor William M. Siebert for helpful criticism and discussions.

G. Einarsson

References

1. W. M. Siebert, Statistical decision theory and communication: The simple binary decision problem, Lectures on Communication System Theory, edited by E. J. Baghdady (McGraw-Hill Book Company, New York, 1961), Chapter 8.

2. C. W. Helstrom, Statistical Theory of Signal Detection (Pergamon Press, London, 1960).

3. J. Dugundji, Envelopes and pre-envelopes of real waveforms, Trans. IRE, Vol. IT-4, March 1958.

4. G. Einarsson, A Communication Analysis of High-Frequency Ionospheric Scattering, S.M. Thesis, Department of Electrical Engineering, M.I.T., January 1962. (To be published as Technical Report 400, Research Laboratory of Electronics, M.I.T.)

5. S. O. Rice, Mathematical analysis of random noise, Bell System Tech. J. 23, 282 (1944); 24, 46 (1945).

6. G. Turin, Error probabilities for binary symmetric ideal reception through nonselective slow fading and noise, Proc. IRE 46, 1603 (1958).