Prof. J. B. Wiesner
20A-126

AN ELECTRONIC CORRELATOR

T. P. CHEATHAM, JR.

TECHNICAL REPORT NO. 122

MARCH 28, 1951

RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

The research reported in this document was made possible through support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the Army Signal Corps, the Navy Department (Office of Naval Research) and the Air Force (Air Materiel Command), under Signal Corps Contract No. W36-039-sc-32037, Project No. 102B; Department of the Army Project No. 3-99-10-022.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report No. 122 March 28, 1951

AN ELECTRONIC CORRELATOR

T. P. Cheatham, Jr.

This report is essentially the same as a doctoral thesis
in the Department of Electrical Engineering, M.I.T.

Abstract

The role and application of correlation functions in the statistical theory of communication is described in terms of the important basic properties of autocorrelation and crosscorrelation functions.

An electronic correlator constructed on the basis of (a) utilizing pulse-sampling techniques and (b) multiplication in time-amplitude coordinates is described in detail. General criticisms and suggestions for improvements, based on experience gained from the application of the correlator to several problems, are given to aid future developmental work. Some additional methods of computing correlation functions are outlined and discussed.

The experimental application of the electronic correlator is illustrated by several studies that are, in general terms, concerned with the detection, analysis, or filtering of a desired signal in noise.

AN ELECTRONIC CORRELATOR

I. Introduction

The statistical approach to communication theory is relatively new. However, the mathematical tools which inherently express its concepts are not relatively new. With the advantage of these familiar tools, the theoretical pioneering in the new field of communication at first far outstripped the experimental and practical investigation that modern science requires as an inseparable complement to theory.

One of the most basic and fundamental tools inherent in the new theoretical approach is the mathematical theory of probability. Because of its centrality of concept, a history of the development and applications of the theory of probability carries one through many of the thought processes which have led to the present broadening concept of a unifying communication theory. For a concise summary and evaluation of the history, philosophy, and principles of the theory of probability, the reader is referred to a recent essay on the subject by Ernest Nagel (1).

II. Statistical Theory of Communication

2.1 Measure of Information

One of the most basic and essential building blocks of the statistical approach to communication is the notion of a measure for information. The necessity for such a measure is intuitively apparent when we consider that no communication engineer today would object to the statement that his task is directly or indirectly concerned with the recording, preservation, transmission, and use of information. He would not object to general statements concerning his job in terms of improving the technique of these operations on information, but he would, in general, be at a loss to discuss quantitatively an optimum design of his equipment in terms of a maximum transmission of the information the equipment was to handle.

Wiener (3) and Shannon (4) and later, but independently, Tuller (5) have arrived at a simple definition and measure of information in terms of a basic binary decision made between two equally probable simple alternatives, one or the other of which is bound to happen. This definition has been given a weighted justification by Shannon (4) and because of its simple and unitary form is found to be capable, in its transformations and extensions, of consistent quantitative interpretation as well as orientation to fit those intuitive feelings generally held regarding the vague notion of information.

The fundamental nature and usefulness of the concept of a measure of information is best illustrated in terms of the following specific examples given by Wiener (3) and included here with clarifying details. As a first example, the amount of information contained in the perfectly precise measurement of a quantity known to lie between A and B, geometrically two end points on a line segment, is determined.

A———————————————B

If we have made the measurement sometime previously or if we have a strong feeling based on some previous knowledge as to the probable range of the measurement within the limits A, B, then it is consistent with our notions of information that our net gain in information is something less than if the measurement is made with a uniform a priori probability of being anywhere in the range A, B. There are two important notions implicit in our example to this point:

1. Information is a net or gained quantity with respect to an a priori level.
2. Information is a function of every link of the communication system including the final destination.

To obtain a quantitative measure of the information gained by a precise measurement in terms of the defined unit of information, let the a priori probability of measurement have a uniform distribution over the range A, B. " By setting A = 0 and B = 1, the measurement can be represented in the binary scale by an infinite binary number $.a_1a_2a_3 \ldots a_n \ldots$ where $a_1, a_2, \ldots$ each has the value 0 or 1.

$$.a_1a_2a_3 \ldots a_n \ldots = \frac{1}{2} a_1 + \frac{1}{2^2} a_2 + \ldots . \frac{1}{2^n} a_n + \ldots . \qquad (1)$$

The number of choices made and the consequent amount of information is infinite. "

It is helpful at this point to mention briefly the relationship that exists between the theorems of Lebesgue, the theory of probability, and the notion of a measure of information. Wiener (6) has pointed out that the notion of probability applies to a class of contingent situations and has the essential properties of a measure. For example, consider a sequence of independent tosses of a coin such as, " heads, tails, heads, heads, tails" . For such a finite sequence the probability on a basis of the classical " games-of-chance" origin of probability theory (1) is 2^{-n}, where n is the number of tosses. It is interesting to note, as pointed out by Wiener (6), that this probability of 2^{-n} is the same as the measure of the set of all points on the line segment 0, 1 with coordinates whose binary expansion begins 0.10110. The following is quoted directly from Wiener (6, p. 210):

> " This mapping immediately suggests a definition of probability for infinite sequences of tosses. The probability of any set of sequences of tosses is defined as the Lebesgue measure of the set of points whose binary representations correspond to sequences of tosses in the set, in such a manner that 1 corresponds to 'heads' and 0 to 'tails'. We can even represent sequences infinite in both directions by binary numbers in such a way as to define the probability of a set of sequences, by having recourse to some definite enumerative rearrangement of such a sequence.
>
> " If we have made 'probability' a mere translation of 'measure', 'average' becomes the equivalent of 'integral'. We are accordingly able to use the entire body of theorems concerning the Lebesgue integral in the service of the calculus of probabilities. "

The precise measurement represented by Eq. 1, therefore, corresponds to a Lebesgue measure of zero, a probability zero, and an infinite amount of information.

There exists, however, a limit to our ability to make an actual precise measurement. The recognition that noise is the fundamental limitation on the rate of transmission of information is a significant contribution of Wiener' s. (This basic point was also recognized by Shannon (4) and Tuller (5). For the moment ignoring the source of the error, we may evaluate its effect on the amount of information by assuming a uniformly distributed error having a measure in the binary scale of $.b_1 b_2 \ldots . b_n \ldots$ If b_k is defined as the first digit not equal to 0, then all decisions of our measurement from a_1 to a_{k-1}, and possibly to a_k, are significant, since

$$\sum_k \frac{1}{2^n} = \frac{1}{2^{k-1}}$$

while all later decisions are not. A detailed development and geometrical interpretation of this point of the theory has been given by R. M. Fano (7). Geometrically, n decisions subdivide the line segment 0, 1 into 2^n equal subsegments, where n is the number of significant decisions made in terms of the defined error.

Our original problem has now developed to a point of more specific designation. The information reference level is determined by the a priori knowledge that the variable to be measured lies between 0 and 1. A measurement is made as precisely as the assumed error will permit and a posteriori the variable is known to lie on the interval a, b inside 0, 1.

A priori zero decisions were made, and a posteriori n decisions have been made; the net gain in information is then n decisions. A general expression for n can be reached by many alternative but equivalent lines of reasoning. One method is to note that the line segment 0, 1 is divided geometrically into 2^n equal parts by the n binary decisions. Denoting the number of parts as M, then $M = 2^n$ and $n = \log_2 M$. Alternatively, each of the 2^n parts of the line is equal to the measure of the error (a, b) and hence

$$M = \frac{\text{measure of } (0, 1)}{\text{measure of } (a, b)} . \tag{2}$$

Therefore the gain in information resulting from our a posteriori knowledge is

$$\log_2 \frac{\text{measure of } (0, 1)}{\text{measure of } (a, b)} . \tag{3}$$

Since the Lebesgue notion of measure is directly translatable to the theory of probability, we may also write for the amount of information gained as a result of our a posteriori knowledge

$$\log_2 \frac{\text{measure of } (0, 1)}{\text{measure of } (a, b)} = \log_2 \frac{1}{p(a, b)} = - \log_2 p(a, b) \tag{4}$$

where

p(a, b) = probability of a measurement falling within the interval a, b.

The above expression for information gained a posteriori holds for a single measurement. A succession of measurements has an average gain per measurement which is given by the mean of the total measurements. If a message is conceived as a succession of independent measurements, which in the Wiener sense are stationary in time, i.e. represent a time series, then the mean is the expectation of our measure of information, familiarly defined in probability theory as

$$E(-\log_2 P_k) = -\sum_k p_k \log_2 p_k \ . \tag{5}$$

For the concept of a measure of information to be of any real value two important extensions are necessary. First, the definition must be extended to a more general continuous case, and, second, to the important but somewhat special case of determining the information gained by fixing one or more variables in the problem.

2.11 Continuous Case (Information associated with a knowledge of probability density distributions)

a. A priori knowledge is the probability that the variable should lie between x and $x + dx = p_1(x)dx$.

b. A posteriori knowledge is the probability that the variable is between x and x $+ dx = p_2(x)dx$.

We wish to determine the gain in information resulting from our knowledge of $p_2(x)$.

The answer to this question can be obtained systematically by various means (e.g. see Ref. 7). Physical reasoning on the basis of Wiener's translation of the Lebesgue measure of probability allows us, however, to go directly to the answer. Our problem here is closely related to our first, with the process reversed: i.e. our knowledge is stated in terms of statistical parameters which are the probability density distribution functions of the variable being measured. Each section of area p(x)dx of the functions $y = p_1(x)$ and $y = p_2(x)$ is directly translatable to an equivalent Lebesgue measure denoting a corresponding amount of information. The mean information associated with a specification of $p_1(x)$ and $p_2(x)$ can therefore be determined in terms of attaching a mean width to the measures corresponding to the respective $p_1(x)dx$ and $p_2(x)dx$ areas.

The mean or average amount of information associated with each curve is therefore

$$H_2(x) = -\int_{-\infty}^{\infty} p_2(x) \log_2 p_2(x)dx$$

and

$$H_1(x) = -\int_{-\infty}^{\infty} p_1(x) \log_2 p_1(x)dx \tag{6}$$

with the net gain given by the difference between these two measures.

2.12 Information Gained by the Process of Fixing One or More Variables

Specifically, the case illustrated is that of fixing the sum w of two variables u and v.

a. A priori knowledge is

1. Two variables, u and v, are independent.
2. Probability that u lies between x and x + dx is

$$\frac{1}{\sqrt{2\pi a}} e^{-\frac{x^2}{2a}} dx .$$

3. Probability that v lies between x and x + dx is

$$\frac{1}{\sqrt{2\pi b}} e^{-\frac{x^2}{2b}} dx .$$

b. A posteriori knowledge is

1. Measurements, $w = u + v$
2. and $p_w(u)$.

We desire to determine the average gain in information concerning the variable u, resulting from an a posteriori knowledge of the sum w of the desired variable u and the additional variable v. An evaluation of the a posteriori knowledge of the probability of u, knowing w, is obtained through the application of Laplace's generalized form of Baye's theorem.

$$p_w(u) = \frac{p(u)p_u(w)}{p(w)} \tag{7}$$

where

$$p(u) = \frac{1}{\sqrt{2\pi a}} e^{-\frac{x^2}{2a}}$$

$$p(w)^* = \frac{1}{\sqrt{2\pi(a+b)}} e^{-\frac{w^2}{2(a+b)}}$$

and $p_u(w)$ = probability of w, knowing that u has occurred. This is simply then the probability of v, where $v = w - u = w - x$

$$= p(v = w - x) = \frac{1}{\sqrt{2\pi b}} e^{-\frac{(w-x)^2}{2b}} .$$

*A well-known result given in many texts. See Fry (8).

By substitution in Eq. 7, the a posteriori distribution of u (knowing w and that w = u + v) is

$$p_w(u) = \frac{p(u)p(w - u)}{p(w)} = \frac{\frac{1}{2\pi\sqrt{ab}} e^{-\frac{x^2}{2a} - \frac{(w - x)^2}{2b}}}{\frac{1}{\sqrt{2\pi(a + b)}} e^{-\frac{w^2}{2(a + b)}}}$$

$$= \frac{1}{\sqrt{2\pi \frac{ab}{a + b}}} e^{-\frac{x^2}{2a} - \frac{(w - x)^2}{2b} + \frac{w^2}{2(a + b)}}$$

$$= \frac{1}{\sqrt{2\pi \frac{ab}{a + b}}} e^{-\frac{a + b}{2ab}\left[x - \frac{aw}{a + b}\right]^2} \quad (8)$$

The gain in information concerning u, resulting from a knowledge of w = u + v, is then given by the difference between the mean information associated with the functions p(u) and $p_w(u)$.

$$\frac{1}{\sqrt{2\pi \frac{ab}{a + b}}} \int_{-\infty}^{\infty} p(w)dw \int_{-\infty}^{\infty} e^{-\frac{a + b}{2ab}(x - \frac{aw}{a + b})^2} \left[-\frac{1}{2}\log_2 \frac{2\pi ab}{a + b} - (x - \frac{aw}{a + b})^2 (\frac{a + b}{2ab})\log_2 e\right] dx$$

$$- \frac{1}{\sqrt{2\pi a}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2a}} \left[-\frac{1}{2}\log_2 2\pi a - \frac{x^2}{2a}\log_2 e\right] dx \quad (9)$$

which after simplification is found to be equal to

$$\frac{1}{2}\log_2 \frac{a + b}{b}\ . \quad (10)$$

The above example has the following significant interpretation given it by Wiener. If the variables u and v are treated respectively as message and noise, then the information carried by a precise message in the absence of noise is infinite (since Eq. 10 goes to infinity as b tends to zero). In the presence of noise the amount of information is finite and approaches zero very rapidly with increasing noise. It is important to note that Eq. 10 is positive. It is also independent of w, as we should expect, since our mean information, defined by the conditional probability density $p_w(x)$, implicitly requires an equivalent integration over all of w, even though the differential of integration is dx. The parameters a and b represent the mean power of the assumed variables. The fact that Eq. 10 is expressed solely in terms of these parameters should not be generalized, since the result is unique for the type of distributions assumed.

2.13 Information as Negative Entropy

The most important and potentially far-reaching of Wiener's contributions has been his recognition that information is the negative of the quantity usually defined as entropy. We are concerned at this point with the fact that Wiener has been able to show that information is a quantity which on the average has the properties connected with the concept of entropy, and in particular the property described by the second law of thermodynamics.

If $P(A_1)$ and $P(A_2)$ are two probability densities, then $P(A_1 + A_2)$ can also be determined. A generalization of the theorem of total probability (10) leads to the important result

$$P(A_1 + A_2 + \dots A_n) \leqslant P(A_1) + P(A_2) + \dots P(A_n) \ . \qquad (11)$$

from this it follows that

$$- \int_{-\infty}^{\infty} P(A_1 + A_2) \log_2 P(A_1 + A_2) \leqslant - \left\{ \int_{-\infty}^{\infty} P(A_1) \log_2 P(A_1) + \int_{-\infty}^{\infty} P(A_2) \log_2 P(A_2) \right\} . \qquad (12)$$

The equal sign will hold only in the case where A_1 and A_2 are independent. In a two-dimensional interpretation of probability, Fig. 1, any overlap in the regions of $P(A_1)$ and $P(A_2)$ will reduce the maximum information belonging to $P(A_1 + A_2)$. The following relevant statement is quoted from Wiener (3, p. 79): "It will be seen that the processes which lose information are, as we should expect, closely analogous to the processes which gain entropy. They consist in the fusion of regions of probability which were originally distinct. For example, if we replace the distribution of a certain variable by the distribution of a function of that variable which takes the same value for different arguments, or if in a function of

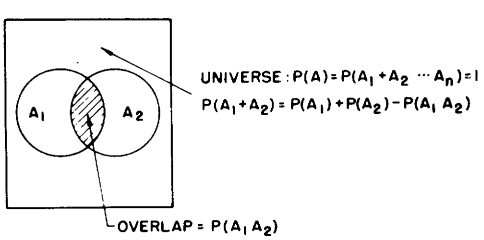


Fig. 1 Two-dimensional probability plot illustrating a gain of entropy due to overlap of regions.

several variables we allow some of them to range unimpeded over their natural range of variability, we lose information. No operation on a message can gain information on the average. Here we have a precise application of the second law of thermodynamics in communication engineering. Conversely, the greater specification of an ambiguous situation, on the average, will, as we have seen, generally gain information, and never lose it."

2.14 The Probability Density Distribution Corresponding to a Maximum Information

Equation 6 of Sec. 2.12 specifies the information associated with a knowledge of the probability density distribution $p_1(x)$ of the message variable. Here we ask the question what distribution function $p(x)$ will make the expectation (mean) of our measure of information $E(-\log p(x))$ a maximum. It can be shown that our measure is a maximum when the variable x is distributed at random, i.e. when $p(x) = \text{constant}$. This has been stated by Wiener (3, p. 79) as an inherent property of the function and demonstrated by Fry (8, p. 201) as a mathematical exercise. Fry's demonstration is included in more detail in Appendix 1.

2.2 Synthesis of Optimum Systems

Some of the basic notions connected with the idea of a measure of information have been covered briefly in Sec. 2.1. Many of the details which turn these basic notions into strong tools in the analysis, evaluation, and synthesis of communication systems have not been discussed, due to the number of cases involved. Although the details may have a large degree of variability, the basic notions will be found invariant. Among the general fundamental notions are those of order and measure, and loss due to a mixing or overlap of regions. The first notion is central to our definition of information and the second is central to the synthesis of optimum systems.

The evaluation of information in Sec. 2.1 was restricted in the sense that only probability density distributions, commonly known as w_1 functions, were considered. Such functions are independent of the frequency and phase of the time functions they statistically characterize. They do not allow for an evaluation of a priori knowledge of frequency and phase, and are therefore inadequate by themselves for the specification of an optimum filter which may be looked upon as a device to remove ambiguity in the desired time wave through an optimum operation performed in the frequency domain or its equivalent. It is suggestive of the process to be described to consider first the case of a message and corrupting noise occupying relatively different positions in the frequency spectrum, as shown by Fig. 2.

The information we shall have of the message by a measurement in an all-pass system of the message plus the noise will certainly be less than the information we shall have from a measurement made after passing the sum through a low-pass filter, as indicated in Fig. 2. In this case of negligible overlap, the type of filter required in the optimum sense of maximizing information is relatively simple and certainly not of

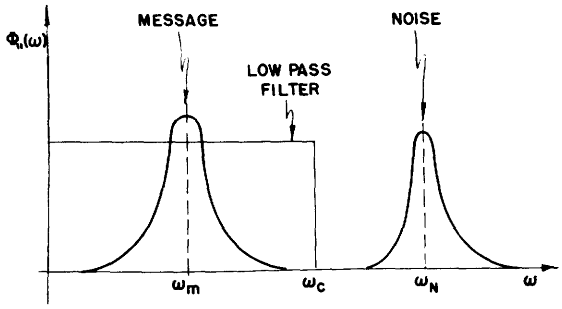


Fig. 2 Power density spectra of a message and corrupting noise occupying different portions of the spectrum.

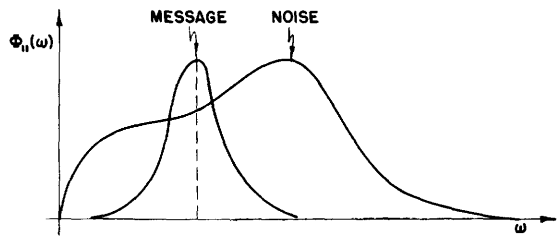


Fig. 3 Power density spectra of a message and corrupting noise occupying overlapping regions of the spectrum.

critical specification. However, if the spectrum of the message and the noise overlap, as indicated in Fig. 3, the specification of the optimum filter becomes more difficult and decidedly more subtle. The basic approach is clear, nevertheless, in that the optimum filter is determined through a compromise between rejection of the noise and a decrease in the negative entropy of the message itself.

Wiener (2) has outlined a general statistical approach to the synthesis of optimum systems. As a special case within the formulation of the problem, he has given a complete solution under the restrictions of linearity and an optimum separation or performance within the dimension of frequency.

Wiener's approach to the problem starts from a demonstration (to be assumed here) that communication is a statistical process. Considered as such a process, a message or an ensemble of messages is described as a stationary time series, and in the general communication problem it is desired to convey the stationary time series (or its transformation: integral, derivative) to a specified destination in the presence of external interference in an optimum fashion, i.e. with minimum error. It is clear from our discussions that the stationary time series which the communication channel should handle is a function of the destination. To be efficient, the time series should contain only such information as can be comprehended by the destination. This portion of Wiener's theory is concerned in general terms with the elimination of correlated data from the message and its optimum coding or transformation. (The performance of such operations reduces the necessary statistical parameters to W_1 functions.) Many of the relevant details have been described independently by C. E. Shannon (4) and W. G. Tuller (5). A helpful and elucidating development has later been given by R. M. Fano (7). For further details the reader is referred to these three papers.

2.21 Error Criterion

If one assumes or neglects the optimum coding of the stationary time series as a function of the destination, and restricts the scope of the problem to conveying a given stationary time series, $f_1(t)$, (or some function of it, such as the derivative, integral, etc.) to the destination in the presence of interference but with a minimum of error, the transmission-link problem can be formulated.

In Fig. 4: $f_m(t)$ = actual ensemble of messages

$f_n(t)$ = actual interference

$f_d(t)$ = desired time function, which may be $f_m(t)$ [filter problem] or some function of $f_m(t)$ such as its derivative, integral, etc. or it may be $f_m(t)$ predicted in time α seconds [predictor problem]

$f_o(t)$ = actual output time function.

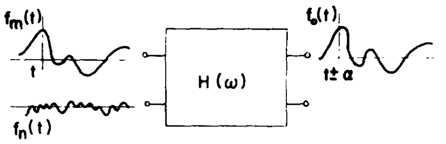


Fig. 4 The general filter problem. H(ω) is a linear system function.

For purposes of more specific discussion let us consider the filter problem only, where $f_d(t) = f_m(t)$. Then we shall want the instantaneous difference or error, $|f_o(t) - f_m(t - \alpha)|$, to be a minimum in a statistical sense. This means that we define a mean error

$$\mathcal{E} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \left| f_o(t) - f_m(t - \alpha) \right|^n dt \qquad (13)$$

where n is subject to our choice.

The next step in the general procedure is to realize the system function H(w) so as to minimize $\mathcal{E}$. This is accomplished principally through the use of the known branches of mathematics of (a) Functions of a Complex Variable and (b) Calculus of Variations.

We may summarize Wiener's general approach to the filter problem in the following three steps:

1. Recognition that communications is a statistical process.
2. Defining the error of the system in terms of the exact time series involved, and the requirements of the destination.
3. Optimum minimization of the error by means of a physically realizable system function.

2.22 Linear Case

As an example of the application of this approach to the problem, Wiener has solved the case where the error is defined as the least mean square error (n = 2, Eq. 13) and the synthesized system function is restricted to being linear.

The mathematical solution relevant to the synthesis of the optimum linear system function subject to a least mean square error criterion is given by Wiener (2). This material has been simplified and extended by Y. W. Lee (11) in a graduate course at M.I.T. The writer's understanding of the relevant theory and its practical engineering details is directly due to a close association with Dr. Lee both in the classroom and in the Research Laboratory of Electronics, M.I.T. The details of this material will be omitted here. It is sufficient to point out that Wiener has shown that the general system function can be expressed in terms of autocorrelation and crosscorrelation functions of the stationary time series at the input to the filter.

2.3 Comments on Nonlinear Case

The general determination of an optimum nonlinear system function requires a knowledge of an infinite number of statistical parameters as compared with the single parameter required above for the linear, least mean-square case. It is obvious that a practical engineering application on this basis is impossible, and the general problem must first be reduced to subclasses of functions which have a finite basis for optimization. It is fortunate that mathematically there exist special classes of time functions* for which a knowledge of a finite number of parameters is sufficient. It is not known whether the time functions generally dealt with in communications are of these types. It is clear that an experimental determination of the statistical characteristics of the time series generally used in communications may do much toward the establishment of a generalized nonlinear synthesis procedure. Certainly we have a basis for stating that experimental research work must precede any further practical theoretical work on the problem.

III. Correlation Functions: Properties and Interpretations

3.1 The Statistical Origin of Correlation Functions

The theory of correlation is a familiar tool of the statisticians, and is closely related to the theory of autocorrelation and crosscorrelation as applied to stationary time series. The use of the word correlation, whether used by the statistician to describe a correlation coefficient or by the communication engineer to describe autocorrelation or crosscorrelation functions, carries no special physical interpretation beyond the normal dictionary notion of denoting dependence and relation.

In order to describe and measure the degree of dependence between two sets of numbers, the statisticians have defined and formulated a quantitative measure known as the coefficient of correlation. It is reached in the following manner.

Consider two ordered sets of numbers X and Y

$$X = x_1, x_2, x_3, \dots x_n$$

$$Y = y_1, y_2, y_3, \dots y_n \ . \qquad (14)$$

If we plot corresponding pairs of points $x_j y_j$ on an orthogonal coordinate system, we will construct what is called a scatter diagram (Fig 5).

The center of gravity of the system defined by the points $x_j y_j$ can be made to lie at the origin by subtracting the means $\overline{x}$ and $\overline{y}$ of the two sets, X and Y, from each number

*From a personal communication of Walter Pitts, Mathematics Department, M.I.T.

in the respective sets, thus forming two new sets of deviations defined by

$$\xi = x_1 - \bar{x},\ x_2 - \bar{x},\ \ldots\ldots\ x_j - \bar{x},\ \ldots\ldots\ x_n - \bar{x}$$

$$\eta = y_1 - \bar{y},\ y_2 - \bar{y},\ \ldots\ldots\ y_j - \bar{y},\ \ldots\ldots\ y_n - \bar{y}\ . \quad (15)$$

Each of the sets ξ and η have a dispersion σ_ξ^2 and σ_η^2 defined by

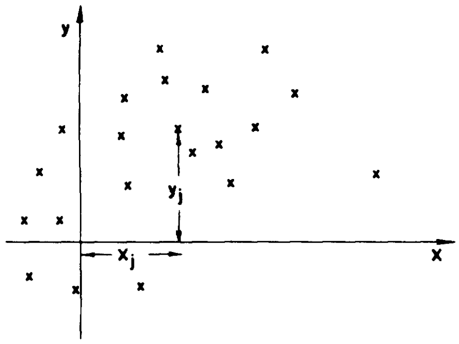


Fig. 5 Scatter diagram.

$$\sigma_\xi^2 = \frac{1}{n}\sum_1^n (x_j - \bar{x})^2$$

$$\sigma_\eta^2 = \frac{1}{n}\sum_1^n (y_j - \bar{y})^2\ . \quad (16)$$

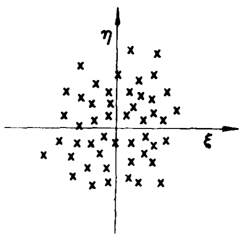


Fig. 6 Scatter diagram showing circular cluster having little correlation.

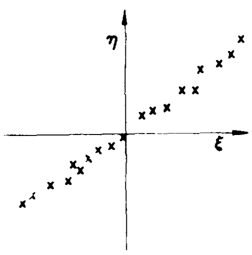


Fig. 7 Scatter diagram showing linear cluster having strong correlation.

If the deviations of Eq. 15 are now plotted on a scatter diagram, the statistician will describe the circular cluster of Fig. 6 as having little correlation and the linear cluster of Fig. 7 as having a strong correlation.

The degrees of correlation attached to Figs. 6 and 7 can be interpreted in terms of notions associated with probability. The points in Fig. 6 are distributed (with reference to direction) about the origin with equal probability, while the points of Fig. 7 have a preferred location with respect to the origin. A necessary condition for zero correlation in this sense is for

$$\frac{1}{n}\sum_1^n (x_j - \bar{x})(y_j - \bar{y}) = 0\ .$$

The quantity

$$\frac{1}{n}\sum_1^n (x_j - \bar{x})(y_j - \bar{y})$$

is usually defined by the statisticians as $\sigma_{\xi\eta}$, and is the parameter, which when normalized gives a quantitative measure of correlation between two ordered sets of numbers. This measure is called the correlation coefficient r, mathematically defined as

$$r = \frac{\sigma_{\zeta\eta}}{\sigma_\zeta \sigma_\eta} = \frac{\sum_1^n (x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_1^n (x_j - \bar{x})^2 \sum_1^n (y_j - \bar{y})^2}} \tag{17}$$

where the denominator $\sigma_\zeta \sigma_\eta$ normalizes r to have a maximum value of $\pm$ 1. The maximum value of $\pm$ 1 can be established by application of the Schwartz inequality, which shows that

$$\left[\sum_1^n (x_j - \bar{x})(y_j - \bar{y}) \right]^2 \leqq \sum_1^n (x_j - \bar{x})^2 \sum_1^n (y_j - \bar{y})^2 \tag{18}$$

where the equal sign holds only if

$$(x_j - \bar{x}) = \lambda(y_j - \bar{y}), \text{ with } \lambda = \text{a constant.}$$

For r = $\pm$ 1, linear dependence must hold. As a result, the correlation coefficient is often interpreted as a measure of degree of linearity and frequently called the coefficient of linearity.

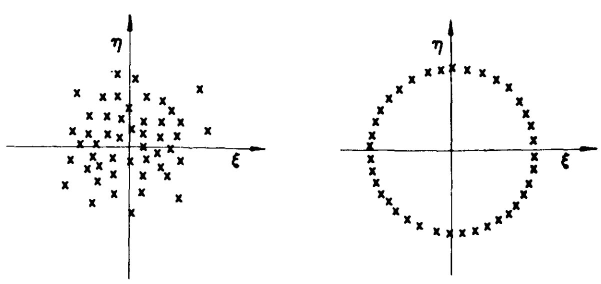


Fig. 8 Scatter diagrams having a correlation coefficient = 0.

The evaluation of r without plotting a scatter diagram can be misleading, as indicated in Fig. 8, where two scatter diagrams are drawn having a correlation coefficient r = 0, but which in our sense of correlation have widely differing degrees of measure. The left figure has little correlation, whereas the right figure has a definite preferred region of probability and hence strong correlation. The degree of linear dependence in both figures is zero and this interpretation of r is still consistent. It is interesting to note that the right-hand figure could be generated by two sets of numbers determined by either a random or periodic selection of values from a sinusoid where the corresponding pairs of the sets differ by 90°. Scatter diagrams determined in this fashion as a function of the phase angle between pairs of points will trace out the familiar Lissajous figures, and the values of r for each scatter diagram plotted as a function of the corresponding phase angle will determine a sinusoid.

It is, in fact, this process which determines the correlation functions used in statistical communication; that is, correlation functions are an extension of the statistical concept of a correlation coefficient in that they represent a plot of r versus a relative shift between two sets, ordering within the sets being maintained while new pairs are generated for each shift.

In communication, the time series dealt with are considered statistically, i.e. invariant with time, and of infinite extent in both directions. A mean, measured over all time, of the correlation measure is therefore defined. (A normalized coefficient is sometimes used for convenience.)

The autocorrelation of the sequence $X = \ldots x_1, x_2, x_3, \ldots x_n \ldots$ is therefore defined as

$$\psi_{XX}(j) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{-N}^{N} x_{k+j}\, x_k \tag{19}$$

and the crosscorrelation coefficient of the sequence X with the sequence $Y = \ldots y_1, y_2, \ldots y_n \ldots$ as

$$\psi_{XY}(j) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{-N}^{N} x_{k+j}\, y_k \tag{20}$$

where both the autocorrelation and crosscorrelation coefficients are functions of the lag j.

When dealing with continuous rather than discrete data, the sequences $X = \ldots x_1, x_2, \ldots x_n \ldots$ and $Y = \ldots y_1, y_2, \ldots y_n \ldots$ have as analogs the functions $f_1(t)$ and $f_2(t)$ respectively. Correspondingly, the autocorrelation function of $f_1(t)$ and the crosscorrelation of $f_1(t)$ and $f_2(t)$ are defined respectively as

$$\psi_{11}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f_1(t) f_1(t+\tau)\, dt \tag{21}$$

and

$$\psi_{12}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f_1(t) f_2(t+\tau)\, dt\ . \tag{22}$$

Correlation functions are simply a measure of the mean relationship existing between the product of all pairs of points of the involved time series, separated in time by a delay τ.

3.2 Some Important Properties of Correlation Functions in Communication

The autocorrelation (and similarly the crosscorrelation) function is equally well defined for both periodic and nonperiodic time series. From Sec. 3.1, it is clear that the autocorrelation of a periodic function will be again periodic, and in the case of a sinusoid, the autocorrelation function will be a cosine function of the same frequency. The determination of the autocorrelation of the periodic portion of a time function is seen to be equivalent to the early periodogram of Sir Arthur Schuster. Wiener (6) has extended the periodogram theory to show that the autocorrelation function not only determines the line portion of the spectrum of the time function, but also defines the random or continuous portion of the spectrum occupied by the time function. This important point is contained in his demonstration that the autocorrelation function, $\psi_{11}(\tau)$, and the exact power density spectrum, $\Phi_{11}(w)$, of the time function are uniquely related by their Fourier transforms, that is

$$\Phi_{11}(w) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi_{11}(\tau)e^{-jw\tau}d\tau \tag{23}$$

and conversely

$$\psi_{11}(\tau) = \int_{-\infty}^{\infty} \Phi_{11}(w)e^{jw\tau}dw \ . \tag{24}$$

These relationships equip the communication engineer with a powerful tool, not only in the synthesis problem mentioned in Sec. 2.22, but also in the problem of analysis and evaluation of existing modulation systems. Until the application of the concept of correlation to time series by Wiener (or more correctly, by G. I. Taylor (12), whom Wiener basically credits with its introduction in the study of irregular phenomena), analysis of the spectrum of existing systems was accomplished on an approximate basis, using in general a single frequency (periodic function) for modulation. In addition to the possibility of determining the exact spectral distribution of a message through transformation of the autocorrelation function, there exists the opportunity of measuring and interpreting results completely in the time domain, since the relation is one-one and no information is contained in one that is not contained in the other. In this respect, for instance, the power measurements of a time series are directly readable from the autocorrelation function. Wiener has shown (2) from the Schwartz inequality that the autocorrelation $\psi_{11}(\tau)$ is a maximum at $\tau = 0$, being equal to the mean of $f_1^2(t)$ (i.e. the 2nd moment of the amplitude density distribution of $f_1(t)$). As τ approaches infinity, $\psi_{11}(\tau)$ approaches the square of the mean of $f_1(t)$ (i.e. the square of the first moment of the amplitude density distribution of $f_1(t)$). These general properties allow us to sketch a rough picture of the bounds or envelope of autocorrelation functions for random (nonperiodic) time functions and to indicate directly measurements of d-c, random, and mean power. (See Fig. 9.)

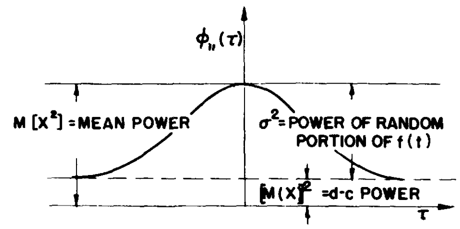


Fig. 9 Sketch of a possible autocorrelation function.

3.21 Preservation of Measure of Phase Difference (Periodic Time Series) or Time of Lead (Nonperiodic Time Series) by Crosscorrelation Functions

That this measure is not preserved in autocorrelation functions is apparent from our above discussion of the Schwartz inequality. A simple investigation of Eq. 22 will quickly establish the converse for crosscorrelation. It is a simple but important property which allows one to investigate complex communication systems such as the human body in terms of classifying the origins, the centers of distribution, and the regions of flow of various degrees of correlated data or information. The measurement of time delay or lead in a communication system having paralleled paths of flow is an important analog measurement in the time domain which determines the equivalent frequency characteristics of the system.

3.22 Decomposition of Correlation Functions of Composite Time Series into Linear Sum of Component Correlation Terms

As an example of this useful property, let us consider a time series $f_1(t)$ consisting of the sum of a signal $S_1(t)$ and noise $N_1(t)$. The autocorrelation of

$$f_1(t) = S_1(t) + N_1(t)$$

is then

$$\psi_{11}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left[S_1(t) + N_1(t)\right] \left[S_1(t+\tau) + N_1(t+\tau)\right] dt$$

$$= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \Big[S_1(t)S_1(t+\tau) + S_1(t)N_1(t+\tau)$$

$$+ N_1(t)S_1(t+\tau) + N_1(t)N_1(t+\tau)\Big] dt$$

$$= \psi_{S_1S_1}(\tau) + \psi_{S_1N_1}(\tau) + \psi_{N_1S_1}(\tau) + \psi_{N_1N_1}(\tau) \ . \qquad (25)$$

We see, therefore, that the autocorrelation $\psi_{11}(\tau)$ is decomposable into two component autocorrelation terms of the signal and noise respectively and two component crosscorrelation terms involving the signal and noise. If the signal and noise are incoherent (i.e. not correlated) then the cross terms will be zero and the difference between the autocorrelation of the signal and the noise can be used as a means of detecting the presence of the signal. Only for the case of a sinusoidal signal will the relationship between signal and autocorrelation be one-one in any sense and hence the information obtained from this detecting property is limited. However, if $S_1(t) = E \cos(wt + \phi)$ and $N_1(t)$ is normally distributed noise, then $\psi_{11}(\tau)$ will appear roughly as shown in Fig. 10.

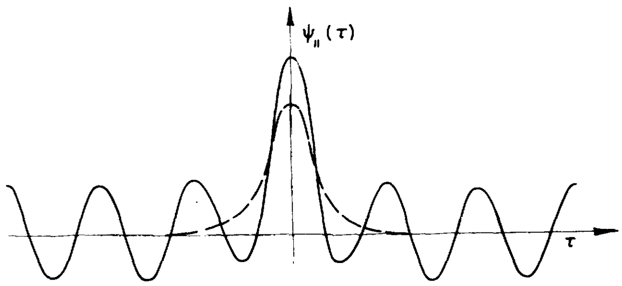


Fig. 10 Autocorrelation of $f_1(t)$ = sinusoid + normally distributed noise.

For large values of τ, a near infinite signal-to-noise ratio is possible. It is well to remember that an infinite time is also theoretically required. The information gained through a measurement of autocorrelations or crosscorrelations as a function of the time of correlation is covered in detail in Sec. 5.1.

It is pointed out that Eq. 25 allows one to compute the crosscorrelation between two time series by computation of autocorrelations only.

3.3 Fields of Application

There exists an extensive array of important problems within the broad scope of the statistical theory of communication in which the concept of correlation plays a vital if not a center role. A few of the more important and general applications of correlation functions can be listed in summary:

1. They are the most important single class of statistical parameters required in the synthesis of the optimum linear system.
2. They determine the bounds, or are at least a measure, of the compressibility of a time function in terms of its optimum coding.
3. They represent a powerful tool for the determination of actual power density spectra, and constitute an exact method for evaluating and comparing modulating systems on a spectra basis. Their important application in the study of complex communication channels, such as the human body, can, in principle, be extended to community and social groups.
4. They provide a possible strong initial method of attack on the general problem of nonlinear systems.

3.4 Illustrative Analytical Example

The process of determining correlation functions for discrete time series such as are found in the various pulse-modulation systems is greatly simplified by an assumption of independence between adjacent pulses. This assumption is made in the following simplified example of single edge pulse-width modulation*, illustrated in Fig. 11.

If we assume that the position of the trailing edge with respect to its unmodulated position has a flat distribution $p(x)$ = constant, we may use this simple example to investigate the change in autocorrelation as a function of the degree of modulation, defined for our purpose as

$$m = \frac{d_{max} - d_{min}}{2d_o} \qquad (26)$$

where

d_{max} = maximum width of pulses
d_{min} = minimum width of pulses
d_o = unmodulated width of pulses .

The assumption of independence between pulses allows us to compute the autocorrelation function from a statistical knowledge of $p(x)$ only. In this case

$$p_{d_k}(d_{k+1}) = p(x)$$

and the autocorrelation function is determined by (1) the shape of the unmodulated pulses, (2) the distribution of the trailing edge, $p(x)$, (3) duty factor = $(d_o)/T$, and (4) the degree of modulation, m.

In order to systematize the calculation of $\psi_{11}(\tau)$, the time function illustrated in Fig. 11 is conveniently broken up into the sum of two time series, $f_A(t) + f_B(t)$, such that each series is composed of alternate pulses (Fig. 12). This procedure eliminates the necessity of computing simultaneous overlap with two pulses, and allows for the separation of this effect, in addition to allowing one consistent method of calculation for all values of m and n.

Therefore the autocorrelation function of $f(t)$ is equal to

$$\begin{aligned}\psi_{11}(\tau) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left[f_A(t) + f_B(t)\right]\left[f_A(t+\tau) + f_B(t+\tau)\right]dt \\ &= \psi_{AA}(\tau) + \psi_{AB}(\tau) + \psi_{BA}(\tau) + \psi_{BB}(\tau)\end{aligned} \qquad (27)$$

The computation of each of the component correlation terms can be performed separately. The use of the terms "self-correlation and intercorrelation", introduced by

*For further details of correlation analysis of pulse-time modulation systems including a detailed physical interpretation of methods and results, the reader is referred to E. R. Kretzmer: Interference Characteristics of Pulse-Time Modulation, Technical Report No. 92, Research Laboratory of Electronics, M.I.T. May, 1949.

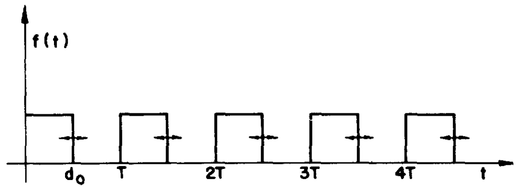


Fig. 11 Schematic representation of a pulse-width-modulation system.

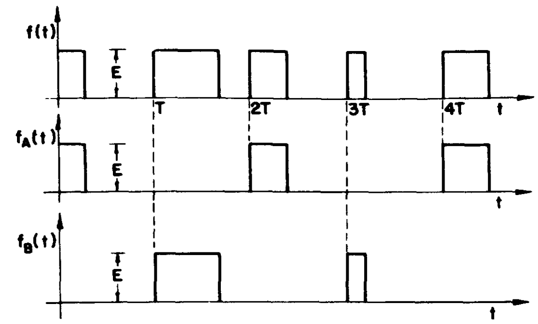


Fig. 12 Decomposition of f(t).

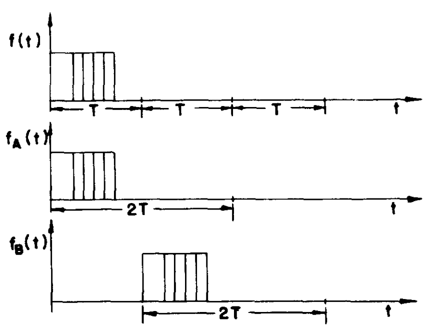


Fig. 13 Symbolic representation of Fig. 12.

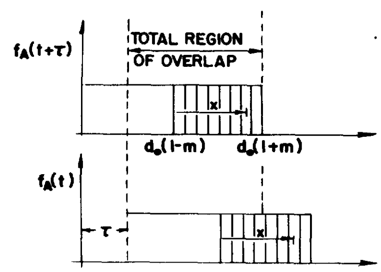


Fig. 14 Schematic representation of computation of self-correlation.

E. R. Kretzmer (13), to denote the autocorrelation and crosscorrelation of component terms of a composite time function will be adopted here.

It is helpful to the logic of the computation to reduce our picture of an infinite series of random-width pulses to one which both symbolically and physically characterizes our statistical knowledge of the time series. In Fig. 13, the trailing edges of the pulses are shaded to fit their distribution density, and periodicity and relative phasing are indicated on the time axis.

A further simplification in terms of organization of computation is to subdivide into regions of delay (τ), as follows

1. $\psi_{AA}(\tau)$, for $0 < \tau < T$

The computation of the self-correlation term $\psi_{AA}(\tau)$ is essentially the determination of the mean of the region of overlap (Fig. 14). The region of overlap of any pair of pulses can be expressed in three general subclasses of lengths

1. unmodulated vs. unmodulated, defined by

   $d_o(1 - m) - \tau$, for $\tau \leqslant d_o(1 - m)$

2. unmodulated vs. modulated, defined by

   τ, for $0 < \tau \leqslant d_o(1 - m)$

   $d_o(1 - m)$, for $d_o(1 - m) < \tau \leqslant 2md_o$ and

   $d_o(1 + m) - \tau$, for $2md_o < \tau \leqslant d_o(1 + m)$

3. modulated vs. modulated, defined by

   $2md_o - \tau$, for $0 \leqslant \tau \leqslant d_o(1 + m)$.

The mean of the contribution of each of these three subclasses of overlapping length will determine the self-correlation function we desire. The split into three regions is necessary, since the probability of any length of overlap of a pulse-pair, $d_o(1 - m) + x - \tau$, is equal to unity over part of the overlap and by a constant less than one over the remainder. A recourse to a Stieltjes integral form of expression of the problem is avoided by recognizing the three regions and performing the evaluation in terms of three component regions of integration.

It is one of the purposes of this illustrative problem, however, to point out that this degree of complexity can be avoided by recourse to a more physical interpretation of the process of correlation which reduces the computation to a physical picture and an operation on a single function.

We may construct a three-dimensional picture on the basis of the following reasoning. Consider a portion of $f_A(t)f_A(t + \tau)$ as indicated in Fig. 15. A pulse of length s_1 contributes to the product $f_A(t)f_A(t + \tau)$ an amount $E^2/(2T)\ (s_1 - \tau)$, and its average contribution to the self-correlation function is defined by $E^2/(2T)\ (s_1 - \tau)p(s_1)$. The sum of all contributions from pulses defined over the range $s = (1 - m)d_o$ to $s = (1 + m)d_o$ will determine the measure of $\psi_{AA}(\tau)$. It is well to note that although $p(s_1) = p(s_2) = p(s_3) = \ldots = 0$, their sum is finite, due to our notion of ordering and its equivalent Lebesgue measure. We

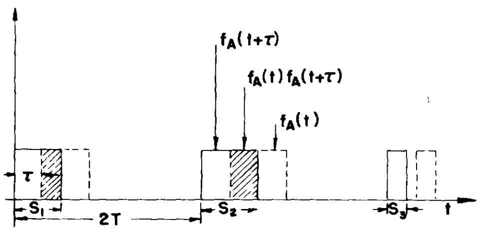


Fig. 15 Portion of $f_A(t)f_A(t + \tau)$ for determining $\psi_{AA}(\tau)$.

may therefore conceive the three dimensional plot of Fig. 16, where

$$\psi_{AA}(\tau)\Big|_{s_j}$$

the self-correlation contribution of a pulse of width s_j, is plotted as a function of τ on two coordinates and the probability density function $p(s)ds$ is plotted on the third. Only a few lines, ordered as a function of pulse length, have been indicated for

$$\psi_{AA}(\tau)\Big|_{s_j}$$

it being understood that these lines are everywhere dense within the region defined by $d_o(1 - m)$ and $d_o(1 + m)$. Pulses of length $s < \tau$ make no contribution to the measure of $\psi_{AA}(\tau)$ and this is indicated by the construction of p(s)ds on Fig. 16 for a value of $\tau > d_o(1 - m)$. Figure 16 therefore corresponds to our physical notion that the self-correlation should decrease linearly as a function of τ until $\tau = d_o(1 - m)$. It will then decrease less rapidly since the contribution of some pulse lengths will be reduced to zero and can decrease no further. At $\tau = d_o(1 + m)$, the contribution of all pulses has been reduced to zero and the measure will be correspondingly zero.

The value of $\psi_{AA}(\tau)$ for any τ is the mean of our ordered sum of contributions and by means of Wiener's translation of probability theory to the theorems of Lebesgue (Sec. II), we may write directly

$$\psi_{AA}(\tau_k) = \int_{d_o(1-m)}^{d_o(1+m)} \psi_{AA}(s)\Big|_{\tau_k} p(s)ds \quad . \tag{28}$$

We have simply to plot the contributions to $\psi_{AA}(\tau_k)$ as a function of s, multiply by p(s)ds and measure the area. Figure 17 illustrates this procedure for three different values of τ.

It is clear that this process corresponds to plotting

$$\psi_{AA}(s)\Big|_{\tau = 0}$$

normalized with respect to $\psi_{AA}(o)$ on its side and then scanning through it with a step function of amplitude $\psi_{AA}(o)\,u(t - \tau)$, as indicated in Fig. 18. Therefore

$$\psi_{AA}(\tau) = \begin{cases} \psi_{AA}(o)\left[\int_{\tau}^{d_o(1-m)} dt + \int_{d_o(1-m)}^{d_o(1+m)} \frac{d_o(1+m) - t}{2md_o}\,dt\right] & \text{for } 0 < \tau < d_o(1-m) \\ \psi_{AA}(o)\int_{\tau}^{d_o(1+m)} \frac{d_o(1+m) - t}{2md_o}\,dt, & \text{for } d_o(1-m) < \tau < d_o(1+m) \end{cases} \tag{29}$$

or

$$\psi_{AA}(\tau) = \begin{cases} \psi_{AA}(o)\left[d_o - \tau\right], & \text{for } 0 < \tau \leq d_o(1-m) \\ \psi_{AA}(o)\frac{\left[d_o(1+m) - \tau\right]^2}{4md_o}, & \text{for } d_o(1-m) < \tau < d_o(1+m) \end{cases} \tag{30}$$

2. $\psi_{BB}(\tau)$: A shift in the origin of t does not change an autocorrelation or self-correlation function. Therefore

$$\text{since } f_A(t) = f_B(t + 2T)$$
$$\text{then, } \psi_{AA}(\tau) = \psi_{BB}(\tau) \ . \tag{31}$$

3. $\psi_{AB}(\tau)$: An inspection of Fig. 13 will show that $\psi_{AB}(\tau) = \psi_{BA}(\tau)$. Due to the assumption of independence between adjacent pulses, it should also be clear that $\psi_{AB}(\tau)$ will be a periodic function and in addition that $\psi_{AA}(\tau)$ and $\psi_{BB}(\tau)$ will be of the same

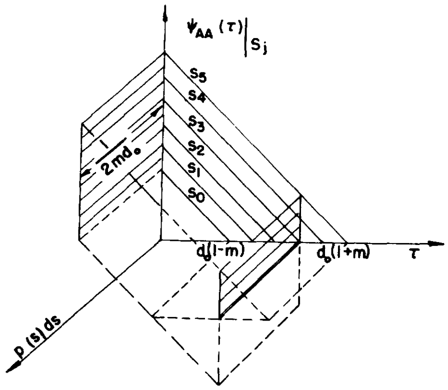


Fig. 16 Three dimensional plot for determination of $\psi_{AA}(\tau)$.

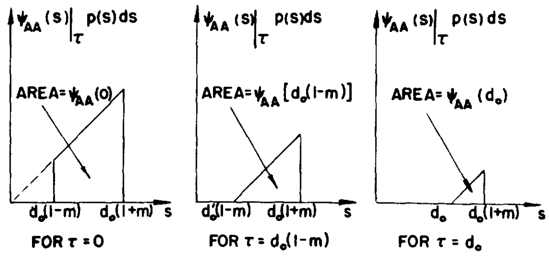


Fig. 17 Geometrical interpretation of Eq. 28 for three different values of τ.

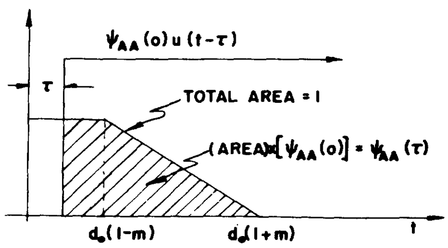


Fig. 18 Geometrically equivalent process for computing $\psi_{AA}(\tau)$.

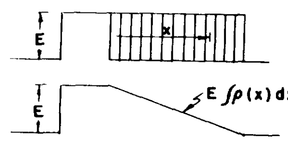


Fig. 19 Schematic representation of equivalent nonrandom pulse.

character for $\tau > T$. Since these portions of the autocorrelation curve are periodic in character (even though the time shift of the trailing edges is distributed at random), it is natural to expect that an equivalent nonrandom pulse shape can be determined from which these portions of the correlation function could be computed in a simple fashion.

That portion of the pulse which is unmodulated, i.e. its minimum width = $d_o(1 - m)$,

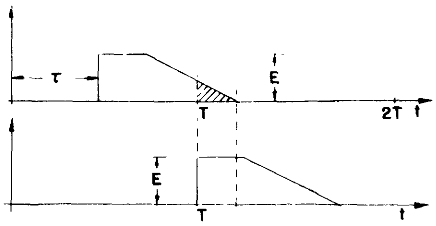


Fig 20 Schematic for computation of intercorrelation and self-correlation of independent pulses.

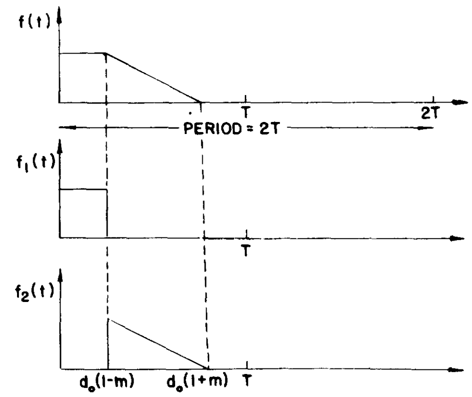


Fig. 21 Equivalent method for computing the self-correlation or intercorrelation of independent pulses.

will be unchanged. We have only to derive an equivalent shape for the shaded area, indicating random time shifts, of Fig. 13. The requirement placed on this portion of the equivalent nonrandom pulse is such that the area of pulse from $d_o(1 - m)$ to x is the same as the average area, measured over the same range of pulse lengths, of the time series. The region of the pulse from $d_o(1 - m)$ to $d_o(1 + m)$ is therefore defined by $E \int p(x)dx$, as indicated for our specific case in Fig. 19. Therefore, to determine $\psi_{AB}(\tau)$, we have only to analyze the situation of Fig. 20.

The computation of $\psi_{AB}(\tau)$ can be further simplified by recognizing that it is periodically symmetrical about the points $\tau = T, 2T, \ldots nT$, and hence the shape of the curve between T and 2T is the same as the self-correlation of a single equivalent nonrandom pulse. We may therefore analyze the equivalent situation of Fig. 22. Where

$$f(t) = f_1(t) + f_2(t)$$

and

$$\psi(\tau) = \overline{f_1(t)f_1(t + \tau)} + \overline{2f_1(t)f_2(t + \tau)} + \overline{f_2(t)f_2(t + \tau)} \tag{32}$$

$$\overline{f_1(t)f_1(t + \tau)} = \frac{E^2}{2T}\left[d_o(1 - m) - \tau\right], \text{ for } 0 < \tau < d_o(1 - m) \tag{33}$$

$$\overline{2f_1(t)f_2(t + \tau)} = \frac{E^2}{T}\int_{d_o(1 - m)}^{\tau + d_o(1 - m)} \frac{d_o(1 + m) - t}{2md_o}\, dt, \text{ for } 0 < \tau < d_o(1 - m) \tag{34}$$

$$= \frac{E^2}{T} \int_{\tau}^{\tau + d_o(1 - m)} \frac{d_o(1 + m) - t}{2md_o} dt, \text{ for } d_o(1 - m) < \tau < 2d_o m \tag{35}$$

$$= \frac{E^2}{T} \int_{\tau}^{d_o(1 + m)} \frac{d_o(1 + m) - t}{2md_o} dt, \text{ for } 2d_o m < \tau < d_o(1 + m) \tag{36}$$

$$\overline{f_2(t)f_2(t + \tau)} = \frac{E^2}{2T} \int_{d_o(1 - m) + \tau}^{d_o(1 + m)} \frac{(d_o(1 + m) - t)}{2md_o} \frac{(d_o(1 + m) + \tau - t)}{2md_o} dt \tag{37}$$

A plot of $2\psi_{AB}(\tau)$, based on the above method of computation is shown in Fig. 22 for $E = 1$, $T = 1$, $d_o = 1/2$ and $m = 1/2$. A composition of the various self-correlation and intercorrelation terms to form the desired autocorrelation function is shown in Fig. 23.

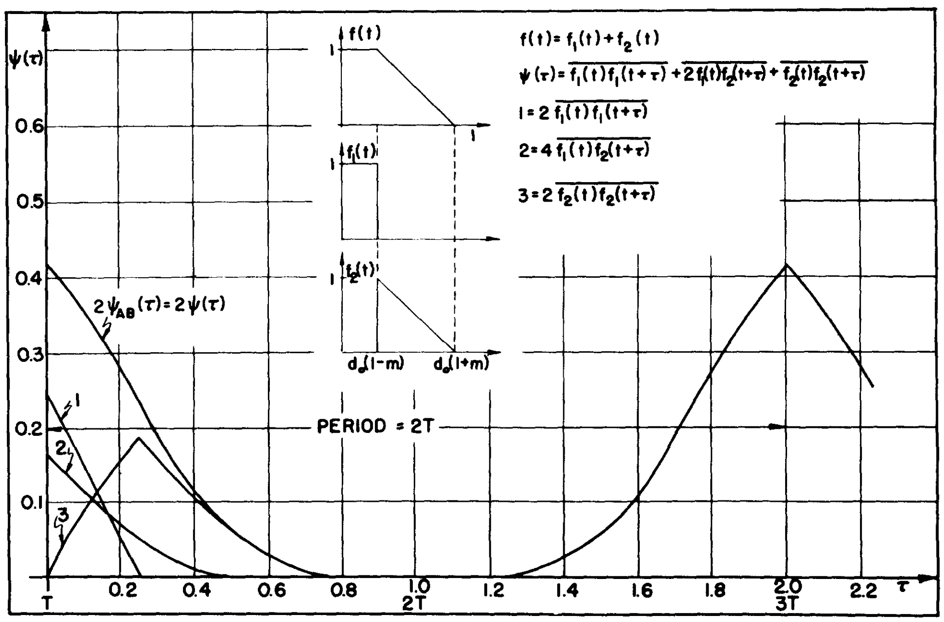


Fig. 22 Intercorrelation and self-correlation function of independent pulses.

IV. Experimental Measurement of Correlation Functions

4.1 Graphical Representation of Process

A graphical representation of the mathematical expression for the autocorrelation function

$$\psi_{11}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f_1(t)f_1(t + \tau)dt \tag{38}$$

is shown in Fig. 24, where the entire function f(t) is first delayed by an interval τ, giving $f_1(t + \tau)$. This time function is then multiplied point by point (continuously) by the undelayed function $f_1(t)$.

The mean of the area under the product function $f_1(t)f_1(t + \tau)$ is then determined as the measure of the correlation function for the specific value of τ_1. The value of τ is changed and the process continued.

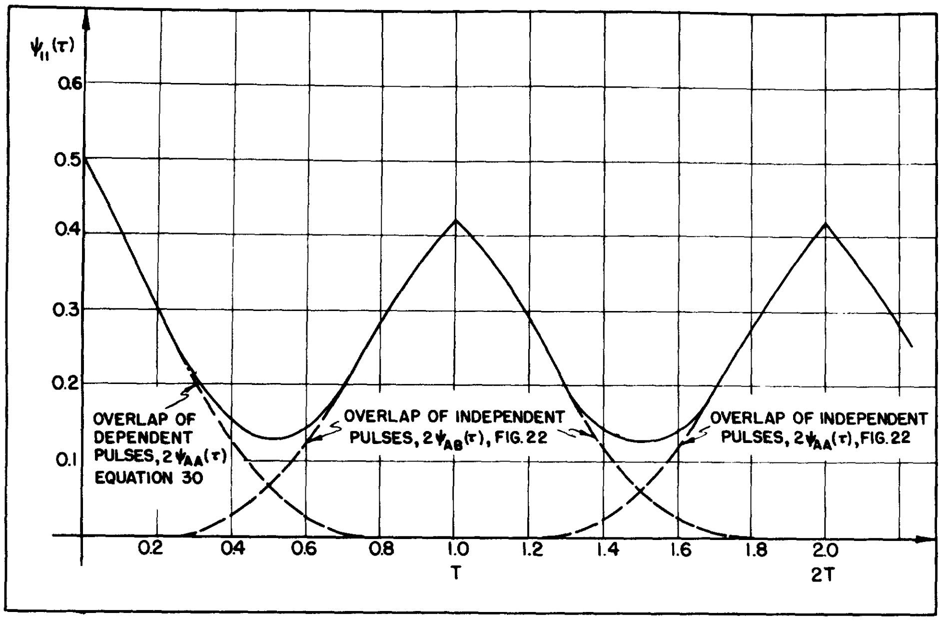


Fig. 23 Autocorrelation function for pulse-width-modulation system of Fig. 11.

4.11 Disadvantages of Graphical Procedure and its Analogs

The obvious disadvantage to this method of computation is the tremendous amount of time-consuming labor involved. In terms of building an analog machine, the method carries with it an additional serious handicap in terms of the concept of having to delay the entire time function.* Closely coupled to this requirement is the consequent necessity of also storing the entire time function. If electronic techniques are used in order to attain a high speed of computation, an a priori knowledge of the bandwidth requirements of the time function is then necessary. This is a general disadvantage, since in most cases the correlation function is of interest to

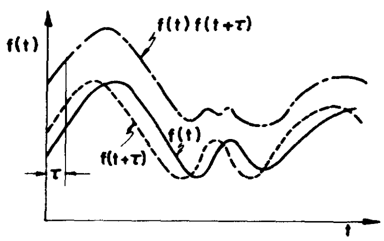


Fig. 24 Graphical method of obtaining an autocorrelation function.

*A practical computation of Eq. 38 must be within finite limits of integration - "entire time function" is used here in the sense of its being a finite portion (an ergodic subset) of the infinite time series.

determine the spectral distribution of the time function.

Earlier methods which were under consideration involved complicated storage tubes and scanning processes, magnetic tapes, rotating drums, etc. All these methods could be visualized as possible techniques for restricted and specific ranges of time functions. When the features of general flexibility and operation over a wide range with reasonable accuracy were added, the above methods rapidly appeared to become cumbersome or inaccurate, or they were considered overly complicated in mechanical and electronic structure for what was intuitively hoped and believed should be a simpler process.

4.12 An Alternate Approach

This simpler process was suggested through a reinterpretation of the mathematical expression of the correlation function, where emphasis is placed on the fact that it represents the mean relationship of the product of all pairs of points of the involved time series separated by an interval τ. That is, an arbitrary or random selection of a large number of pairs of points separated in time τ_1 seconds, multiplied together, summed, and their mean taken, determines one point on the correlation curve. The important point is that the correlation function can be represented by a summation and averaging of a large number of discrete multiplications, where sampled values of f(t), rather than the entire time function are delayed and stored. This means that the mathematical definition of Eq. 38 becomes approximated by

$$\psi_{11}(\tau_k) = \frac{1}{N+1} \sum_{0}^{N} a_n b_n \Big|_{\tau_k} . \tag{39}$$

Equation 39 is recognized as an approximated correlation coefficient, defined in Sec. 3.1, for a discrete time series.* The graphical process of Eq. 39 is represented in Fig. 25 by the arbitrary sets of points on the graph.

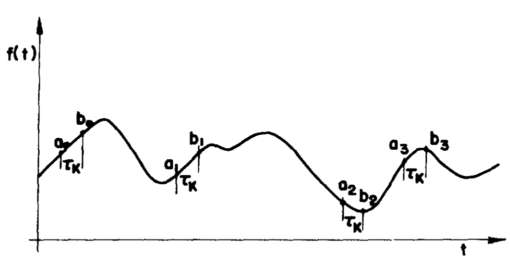


Fig. 25 Portion of a random time function, f(t), showing graphical representation of Eq. 39.

This simple procedure made it possible to design an electronic correlator in terms of well-known pulse-sampling techniques where the delay τ is placed on a single frequency.

4.2 Electronic Correlator

Having decided on a basically electronic structure utilizing the principle of pulse sampling, it was necessary to decide on a suitable coordinate system for multiplication.

*The return to a discrete approximation as a practical means of computation of the correlation function for a continuous time series was suggested independently to the writer by Dr. G. Duvall, formerly of Research Laboratory of Electronics, and by Dr. Y. W. Lee, M.I.T., who, also visualized a possible electronic solution to the problem.

Various known schemes were considered, but rapidly discarded in favor of what was originally thought to be a new technique of multiplying in time-amplitude coordinates. Such a notion was not completely new, however and had already been demonstrated for its adequacy and simplicity (14).

4.21 Early History (First Model)

On the basis of (a) utilizing pulse-sampling techniques and (b) multiplication in time-amplitude coordinates, a first model of the electronic correlator was designed jointly with Dr. E. R. Kretzmer, whose knowledge of pulse circuit techniques was invaluable. Unfortunately, the first correlator was found to be generally inadequate in the sense of performing a time invariant measurement, due to a basic mistake in naively hoping that the correlation functions in which we were interested could be measured over only one channel in approximately one second. This hope was based on a lack of experience with the statistical character of general time functions and was prompted by the desire to present a plot of correlation functions on an oscillograph.*

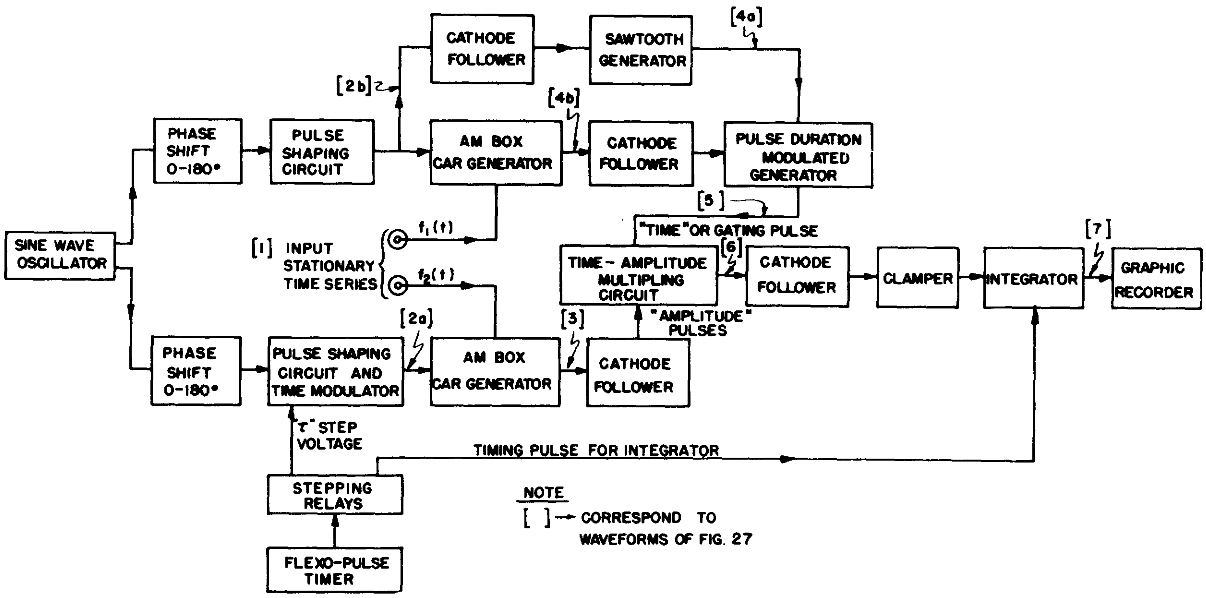


Fig. 26 Block diagram of electronic correlator.

The mistakes of the first correlator were easily corrected, and a second design was made incorporating various circuit changes but leaving the basic method of pulse-sampling and time-amplitude multiplication unchanged.

4.22 Second Model

The general scope of the second correlator and the techniques involved in its

*The use of this first model as a short time correlator (discussed in Sec. 5.23) however, would be appropriate and represents a practical technique.

construction are outlined in the block diagram and waveform descriptions of Figs. 26 and 27.

Waveform 1 of Fig. 28 shows a section of a random time function f(t). A sine-wave master oscillator is used to derive the timing and sampling pulses of waveforms 2a and 2b. In addition to wave-shaping networks, the sine wave is passed through an RC phase-shift network to give an initial coarse adjustment of τ. The timing pulses are

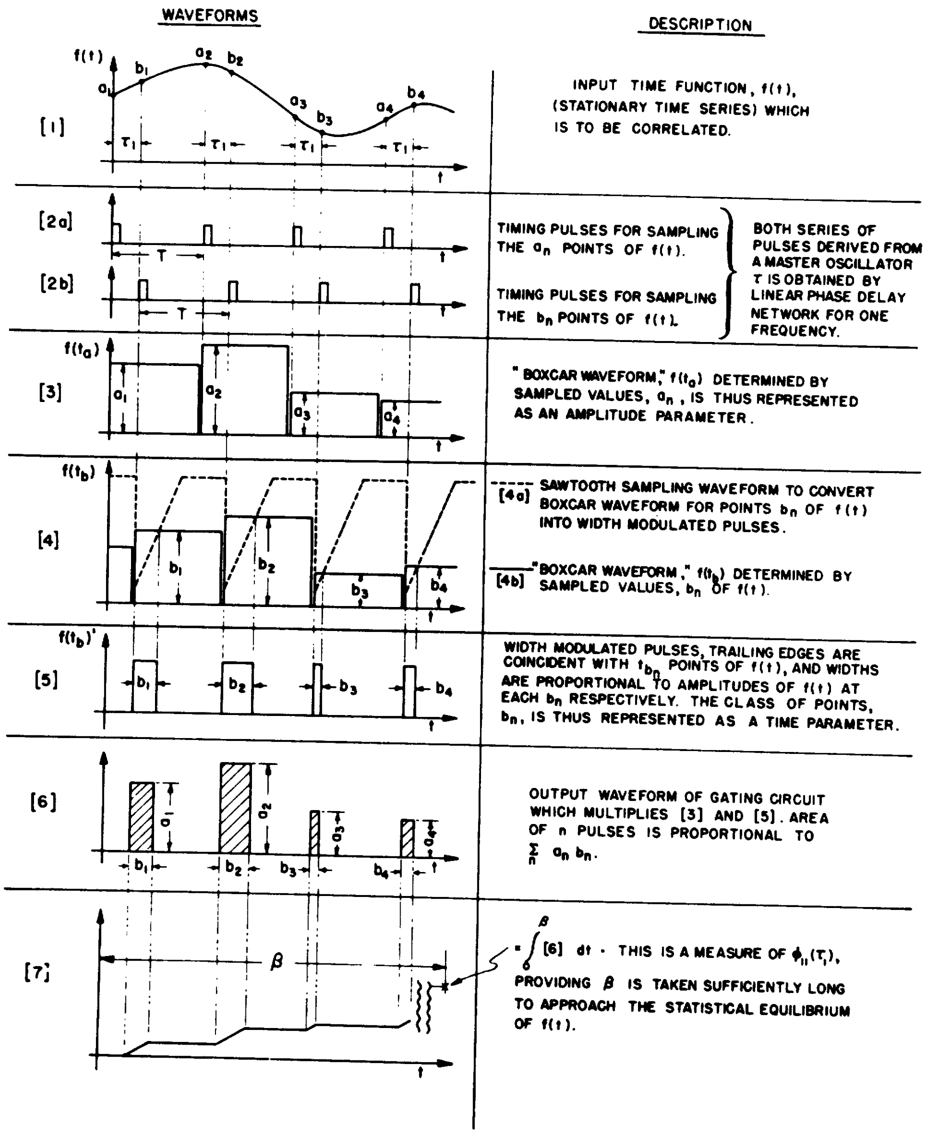


Fig. 27 Characteristic waveforms of electronic correlator.

then time-modulated in discrete steps, the time modulation being accomplished by means of a stepping relay mounted with precision resistors and connected to a well-regulated supply. There are 90 discrete voltage levels available from the relay. These voltage levels are linearly converted to corresponding time delays by voltage intersection with a linear sawtooth in a cathode-coupled slicing circuit. Schematically this process is illustrated in Fig. 28, and corresponding oscillograms for various voltage levels are shown in Fig. 29.

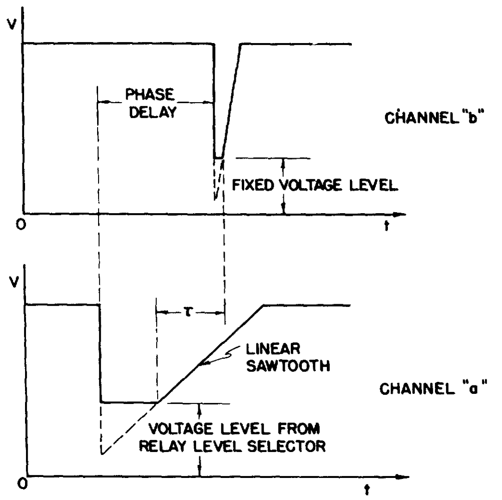


Fig. 28 Schematic of linear transformation of discrete voltage levels into time delay τ_k .

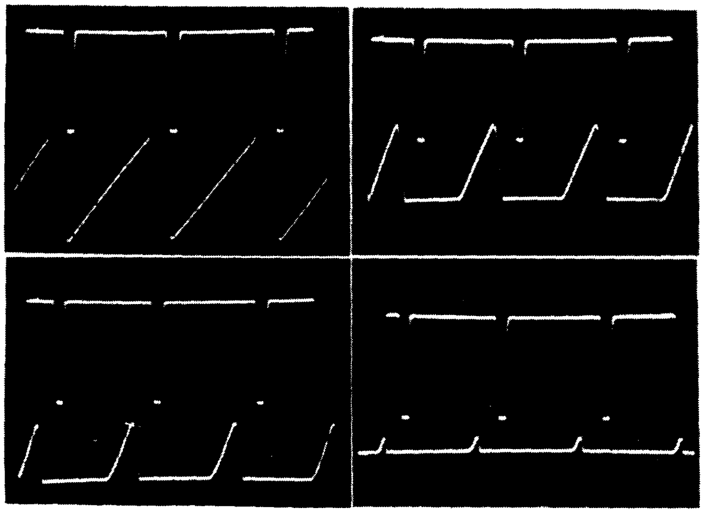


Fig. 29 Oscillograms of transformation of discrete voltage levels into time delays.

Since both the slope of the sawtooth and the voltage increment per step from the relay can be varied, a wide range of variation of τ_k can be selected to fit the time functions being investigated.

It would appear from Fig. 28 that either of the waveforms 2a or 2b could be time-modulated; however, it was found (for circuitry reasons) that time modulation of 2b results in a decrease in the allowable range of τ to something less than one half the sampling period T. By time modulation of 2a, the allowable range of τ is approximately given by T - (b_n) max, where (b_n) max is the maximum width of the derived gating pulses of waveform 5. Figure 30 shows oscillograms of waveforms 2a and 2b for two

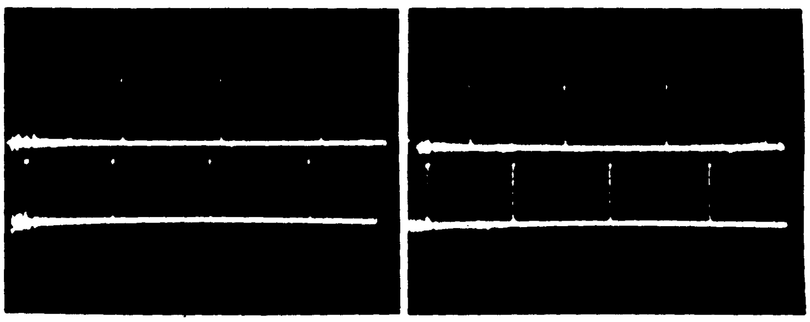


Fig. 30 Oscillograms of timing pulses (waveforms 2a and 2b, Fig. 27).

values of τ_k.

The sampling pulses are used to measure the amplitude of f(t) at the a and b points. These amplitudes are stored over a sampling period T, as indicated by waveforms 3 and 4b, Fig. 27 (shown also in oscillographic form in Fig. 31). In this operation the amplitude of the a_n and b_n points are stored on a capacitor in a so-called "boxcar" generating circuit.

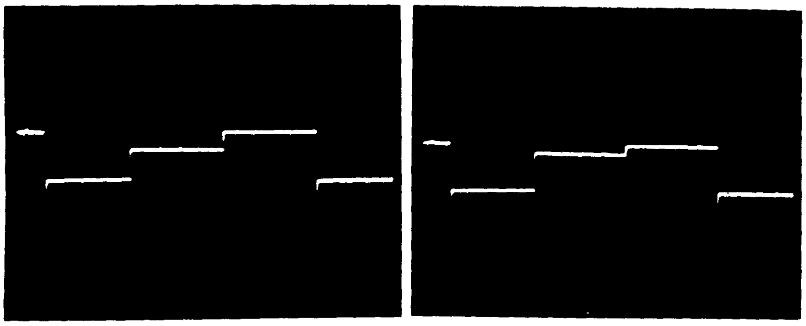


Fig. 31 Oscillograms of "boxcar" waveforms (3 and 4b, Fig. 27).

The amplitude of waveform 3, Fig. 27, is a function of the amplitude of f(t) at a_n, giving the desired amplitude coordinate. A linear sawtooth is used to transform the stored amplitudes of waveform 4b, Fig. 27, to an equivalent time parameter represented by the width-modulated pulses of waveform 5, Fig. 27. (Figure 32 is an oscillogram of these pulses.)

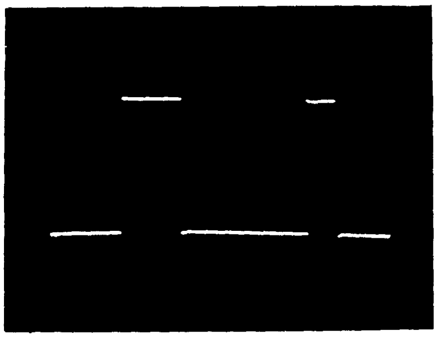


Fig. 32 Oscillogram of width-modulated pulses, waveform 5, Fig. 27.

Waveforms 3 and 5, Fig. 27, are then placed in coincidence in a gating circuit, the output being a series of pulses of varying amplitude and width, as indicated by waveform 6, Fig. 27, and by the corresponding oscillograms shown in Fig. 33. The area of each of these pulses is proportional to the product of a pair of points, a_n, b_n, separated in time τ_k seconds.

As the final step, the integral or summation of the output pulses of the multiplying circuit over a period β, which approximates the statistics of the random time function, then gives a measure of $\psi_{11}(\tau_k)$. The final value of the integral is recorded, τ shifted, and the process is repeated.

The integrator used in the correlator is of the basic Miller type shown in Fig. 34. It was designed and tested to have less than 2 percent error for integrating times up to seven minutes. Since an ordinary mica or oil capacitor will retain too much charge in the polarization of the dielectric to permit complete discharging in the time allowed between integrations, a special polystyrene capacitor had to be used for the integrating capacitor. The assistance and advice of Mr. Keith Boyer, of the Nuclear Research Laboratory, on the general problem of building an adequate integrator is gratefully acknowledged.

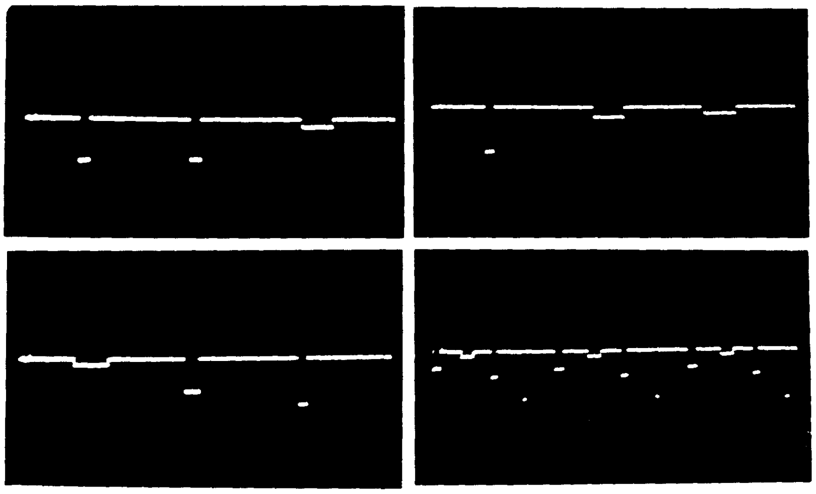


Fig. 33 Oscillograms of output of multiplying circuit (waveform 6, Fig. 27).

Because of the method of multiplication used, a product of $a_n b_n$ is always positive with respect to ground, and is positive and negative only with respect to the average direct current of the unmodulated pulse train from the multiplier. Initially it was found that the standard deviation of many random time functions under linear operation was only of the order of 10-20 percent of the d-c component of the unmodulated pulses. This, at first, put a seemingly heavy restriction on the accuracy with which measurements of the random portion of the correlation function could be measured. By providing means for an adjustable step function at the input to the integrator (accomplished by returning the clamper circuit to an adjustable positive voltage), as much of the d-c component of waveform 5, Fig. 27, could be cancelled as desired. This feature, together with an integrator sensitivity control, makes it possible to vary and control the final presentation of the correlation function not only in time but in amplitude as well.

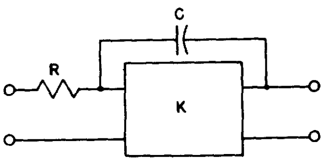


Fig. 34 Basic integrator circuit.

The operation of the electronic correlator is completely automatic. An adjustable timing-circuit (Flexo-pulser) is used to operate the τ-stepping relay and integrator relay of the circuits. Figure 37 is a wiring diagram of the timing circuits.

Figure 35 is a photograph of the correlator showing its size and general structure. Reading from top to bottom, the chassis in the eight-foot rack are (1) stepping relay and

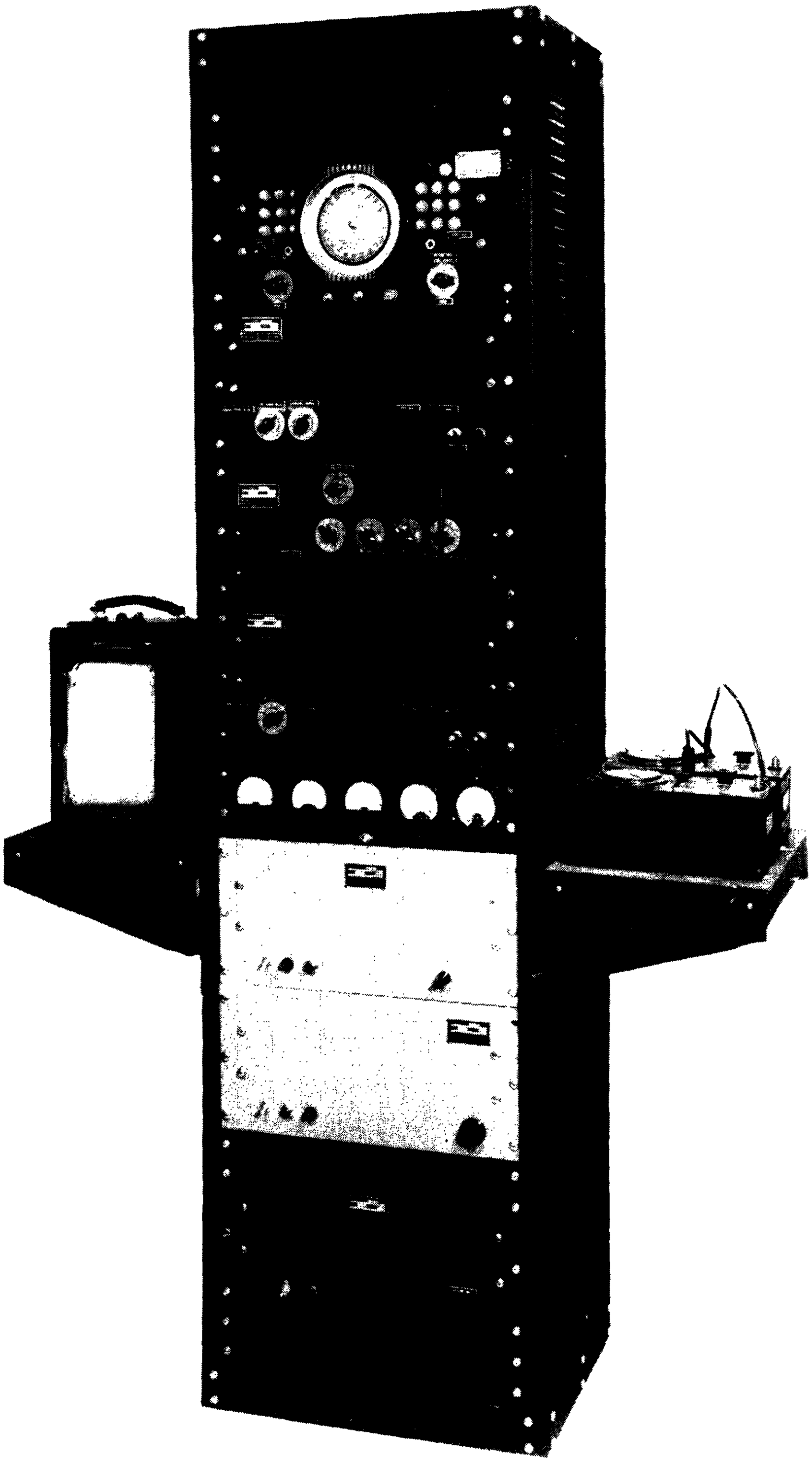


Fig. 35 Electronic correlator.

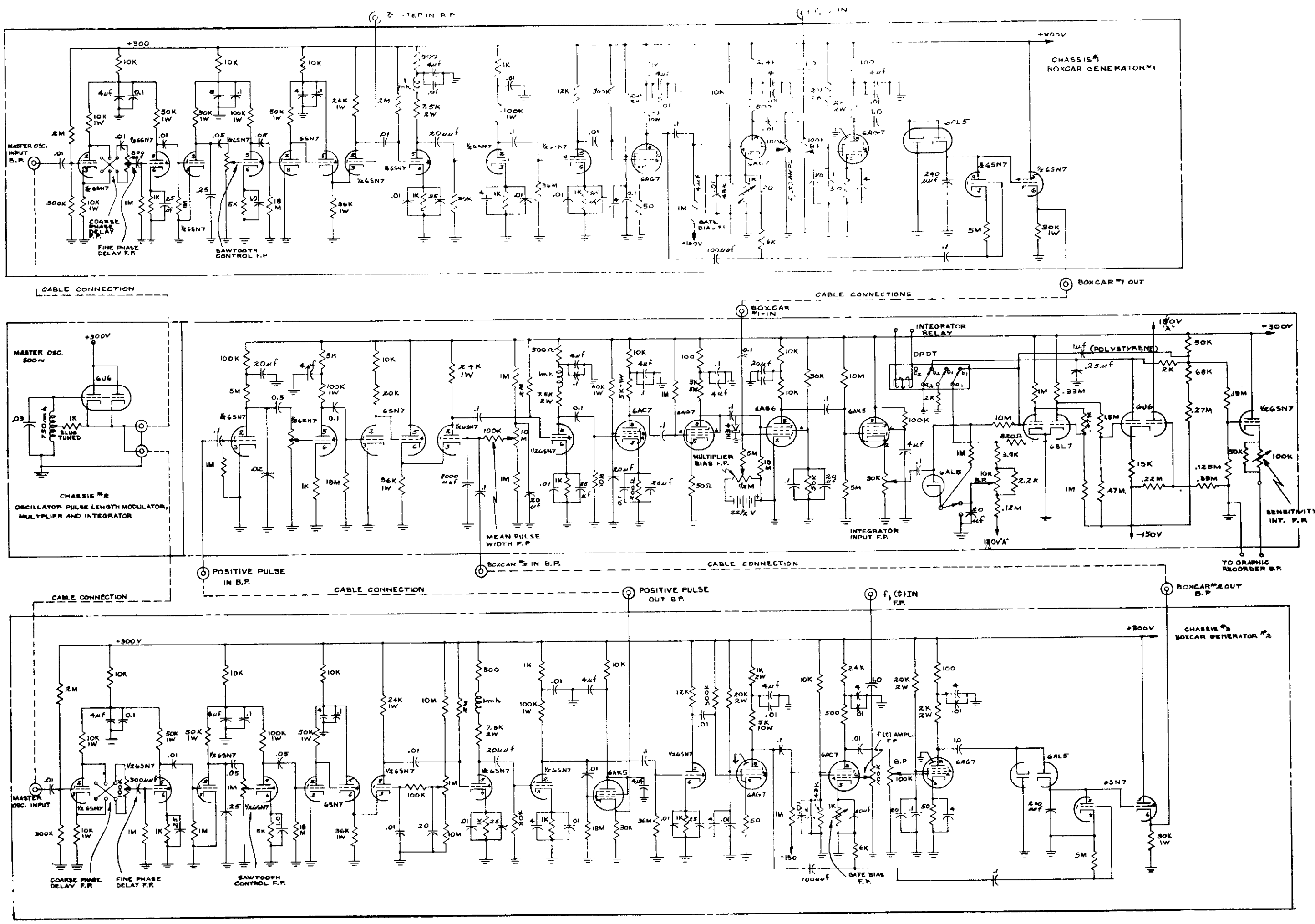


Fig. 36 Wiring diagram of electronic correlator.

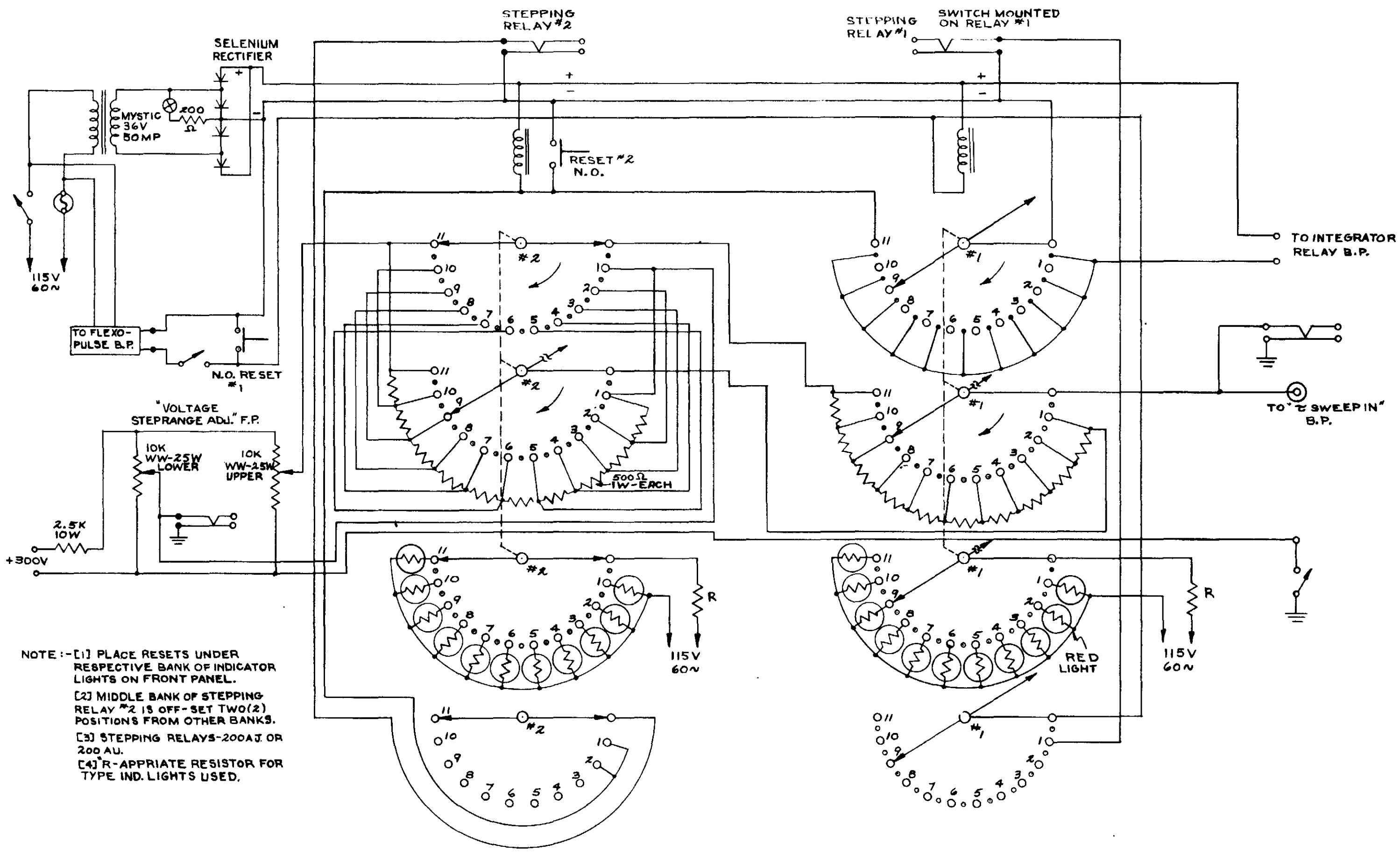


Fig. 37 Wiring Diagram of the timing circuits.

timing circuits, (2) channel a, (3) multiplying, control, and integrating networks, (4) channel b, (5) voltage and current meters, and (6) power supplies. The graphic recorder is on the left shelf and the input-time-function generator may be placed on the right shelf. A wiring diagram of the main electronic components is given in Fig. 36.

4.3 Illustrative Experimental Correlation Curves

A few experimentally measured correlation functions are included here as qualitative examples of the use and flexibility of the correlator. In Sec. V, quantitative experimental applications and measurements are discussed for a number of cases.

4.31 Periodic Functions

Since the behavior in time of a periodic function is known, its autocorrelation can be determined analytically by simple calculations. In general, the determination of correlation functions of periodic time waves is not sufficiently complicated to warrant electronic computation, but they do serve as an extremely useful method of checking the operation of the correlator. The correlator has been tested with several periodic waveforms of varying shape and harmonic frequency content. Figures 38 and 39 are examples

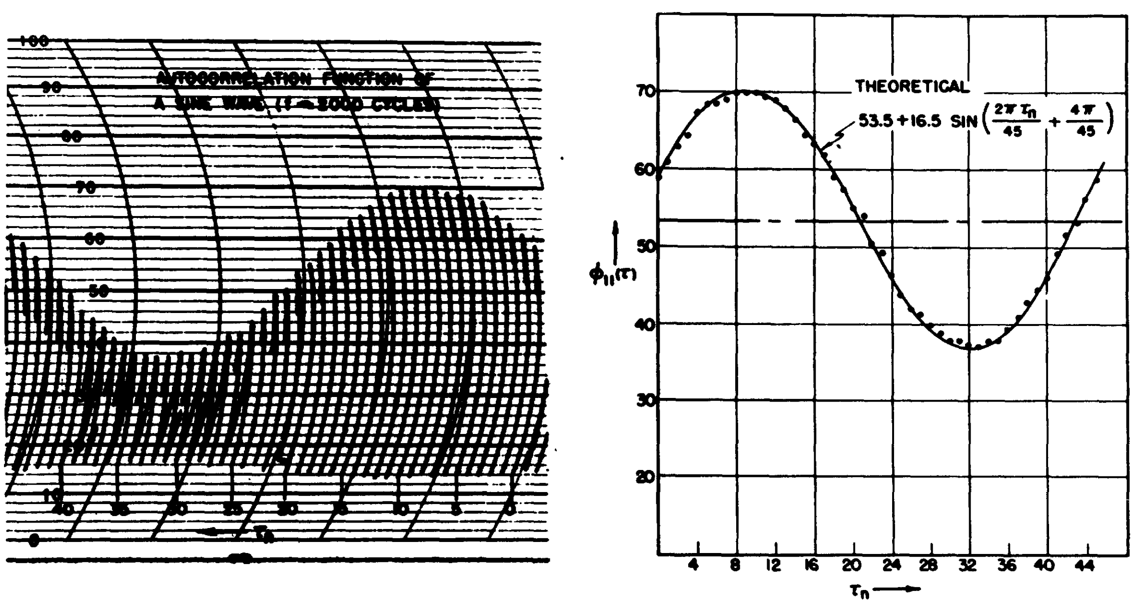


Fig. 38 Autocorrelation function of a sine wave (f = 2 kc) showing comparison with theoretical curve.

of results as recorded by the correlator for a sine wave and square wave respectively. The recorded data are shown replotted on linear rectangular coordinates for comparison of the experimentally determined points and the theoretical curves. Figure 40 shows the pulses at the output of the multiplier for three different values of τ for the sine-wave input of Fig. 38. It is interesting to note the Lissajous figure traced by the lower right-hand corner of the pulses and to recognize that our method of multiplication corresponds

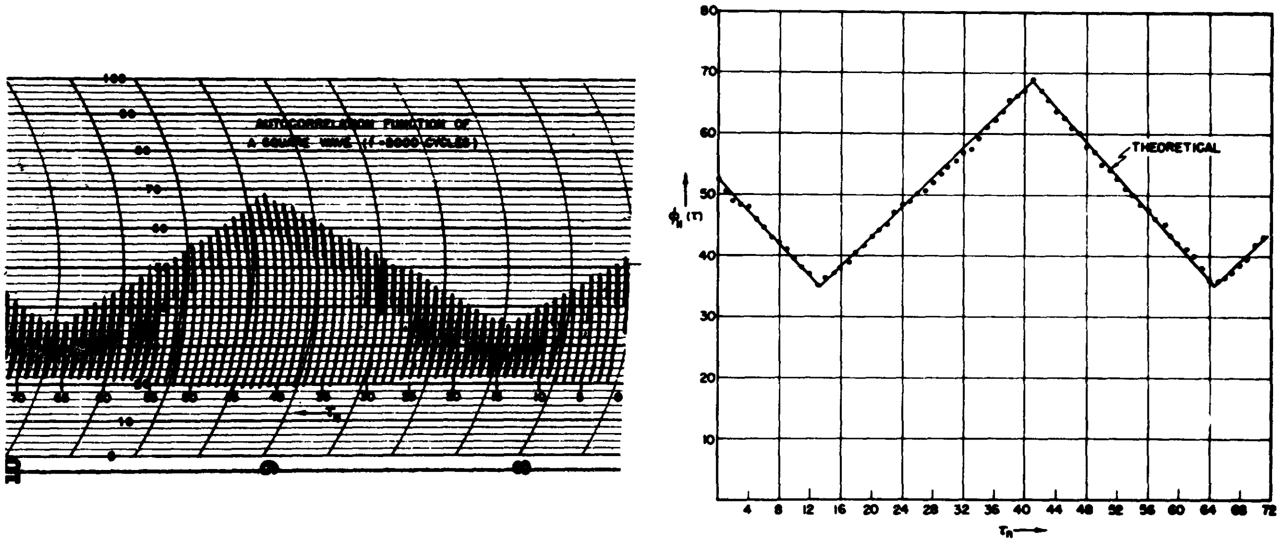


Fig. 39 Autocorrelation function of a square wave (f = 2 kc) showing comparison with theoretical curve.

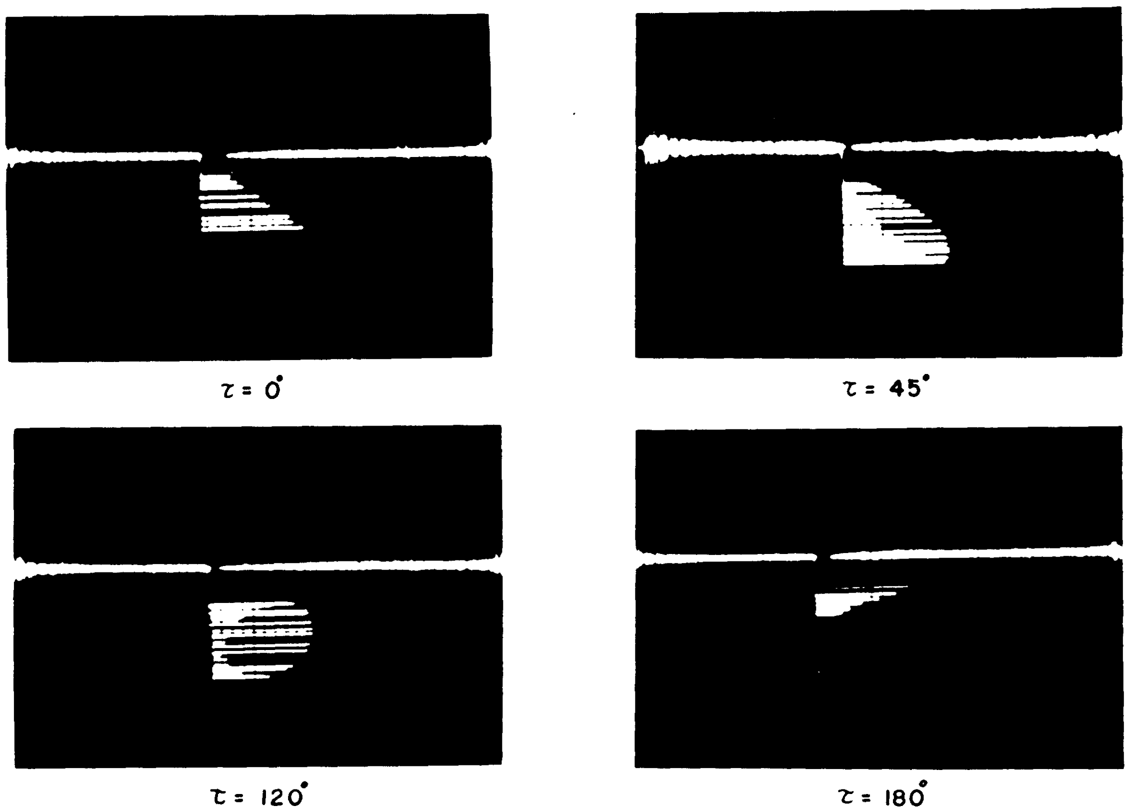


Fig. 40 Output pulses of multiplier for a sine-wave input showing plotting of scatter diagrams and generation of Lissajous figures as a function of τ expressed in degrees.

to an electronic plot of a scatter diagram, as discussed in Sec. 3.1.

4.32 Nonperiodic Functions

Figures 42, 43, and 44 show some illustrative examples of correlation functions for random-input time functions - specifically noise from an 884 gas tube at the output of linear filters.* The transfer characteristic of the noise-source amplifier is shown in Fig. 41.

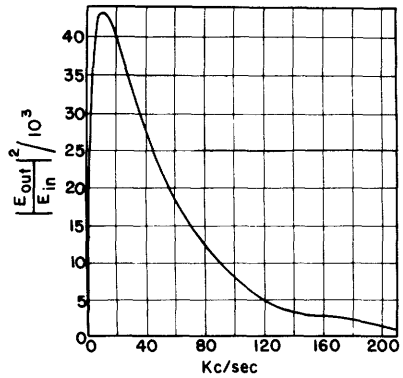


Fig. 41 Over-all transfer characteristic $|E_0/E_1|^2$ vs. kc/sec of noise-source panel measured with signal generator.

The autocorrelation function of the noise signal at output of the transfer characteristic of Fig. 41 is shown in Fig. 42.

The autocorrelation function of Fig. 43 is obtained by inserting a single tuned circuit ($Q \approx 14$, center frequency = 20 kc) within the transfer characteristic of Fig. 41.

Figure 44 illustrates the flexibility of the correlator by showing the first part of the autocorrelation function of Fig. 43 in greater detail, i.e. the delay per step of τ is rèduced to approximately one-tenth its value for Fig. 43, and the correlation function in the vicinity of $\tau = 0$ is shown with the greater detail of approximately ten times as many points.

4.4 Desirable Improvements in Electronic Correlator

A study of correlation functions and the construction of an electronic correlator to measure these functions experimentally was, as pointed out in Sec. I, part of a longer-range research program to investigate the statistical theory of communication. In agreement with the general philosophy of the program, the electronic correlator described here was not constructed with a goal of producing a near-perfect finished product, but rather from the viewpoint of building an adequate computer which could be quickly put to use to open up and suggest more branches of research in the general field under investigation. The acceleration of a program of research with emphasis on group effort from many specific fields and directed toward the exploitation of Wiener's basic theory was considered more desirable than a complete engineering development of a single problem or technique. It was decided that rather than continuously correct each small detail or mistake in order to obtain an optimum design, the correlator described

*The curves shown here are presented in a qualitative rather than quantitative sense. They were measured jointly with Mr. N. Knudtzon (15), who has since made a detailed quantitative statistical study of various noise sources. Some of his results on correlation functions measured with the electronic correlator are included in Sec. V.

above would be used in as many situations as possible in order to point out its limitations, so that concurrently with these studies, and lagging some months in construction time, a new correlator* could be developed which incorporated all the improvements indicated by the present model. The following features are listed as desirable improvements in the present correlator and should be taken into consideration

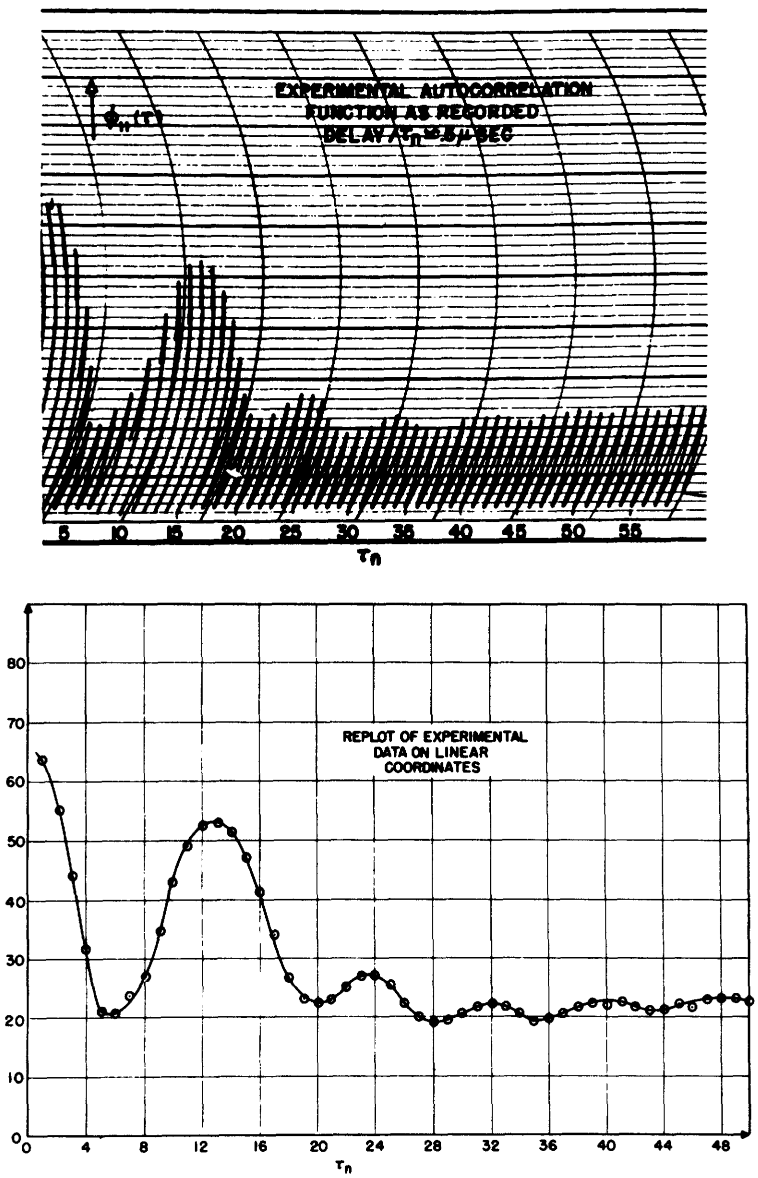


Fig. 42 Autocorrelation of noise from 884 gas tube after passing through transfer characteristic of Fig. 41.

*This correlator is of digital construction, and has been designed and engineered by H. E. Singleton, Research Laboratory of Electronics, M.I.T.

in future development of similar-type computers.

4.41 Jitter of Timing Pulses

Any hum or instability in the sawtooth or voltage levels indicated in Fig. 28 will cause a jitter or inaccuracy in the τ_k time delays. The amount of jitter introduced represents a high-frequency limitation, since it is somewhat analogous in its effect to transit time. For a given amount of hum, the jitter will be proportional to the slope of the sawtooth (Fig. 28); however, this effect is somewhat compensated by the fact that as measurements are made of higher- and higher-frequency time functions, the slope of the sawtooth is generally increased. When faced with the alternative of decreasing the range of voltage from the stepping relay or increasing the slope of the sawtooth in order to decrease τ_n per step, the slope of the sawtooth should always be

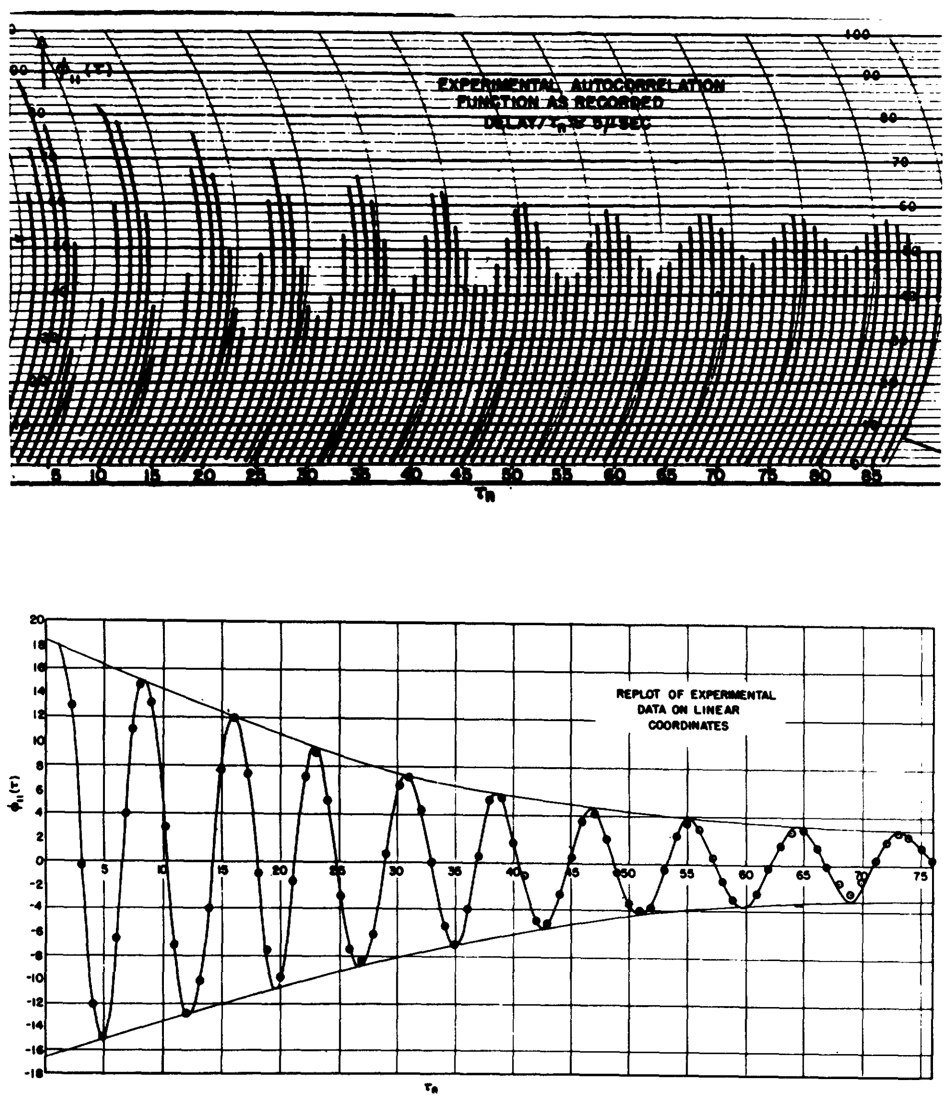


Fig. 43 Autocorrelation function of noise from 884 gas tube after passing through single tuned circuit (Q = 14).

increased. The method, indicated by Fig. 28, of obtaining time modulation of the sampling pulses is to be recommended for its simplicity and with care can certainly be used up to delays per step of 0.5 μsec as indicated by Figs. 42 and 44. However, many communication time functions such as found in television will require a greater accuracy; one method frequently used in radar circuits is to generate accurate timing pulses from a crystal-controlled oscillator and then to select the sampling pulses needed by means of a pedestal or sawtooth as indicated in Fig. 44. By this means a jitter equal to a period of the timing pulses is permissible before any error is introduced in the timing of the selected sampling pulses.

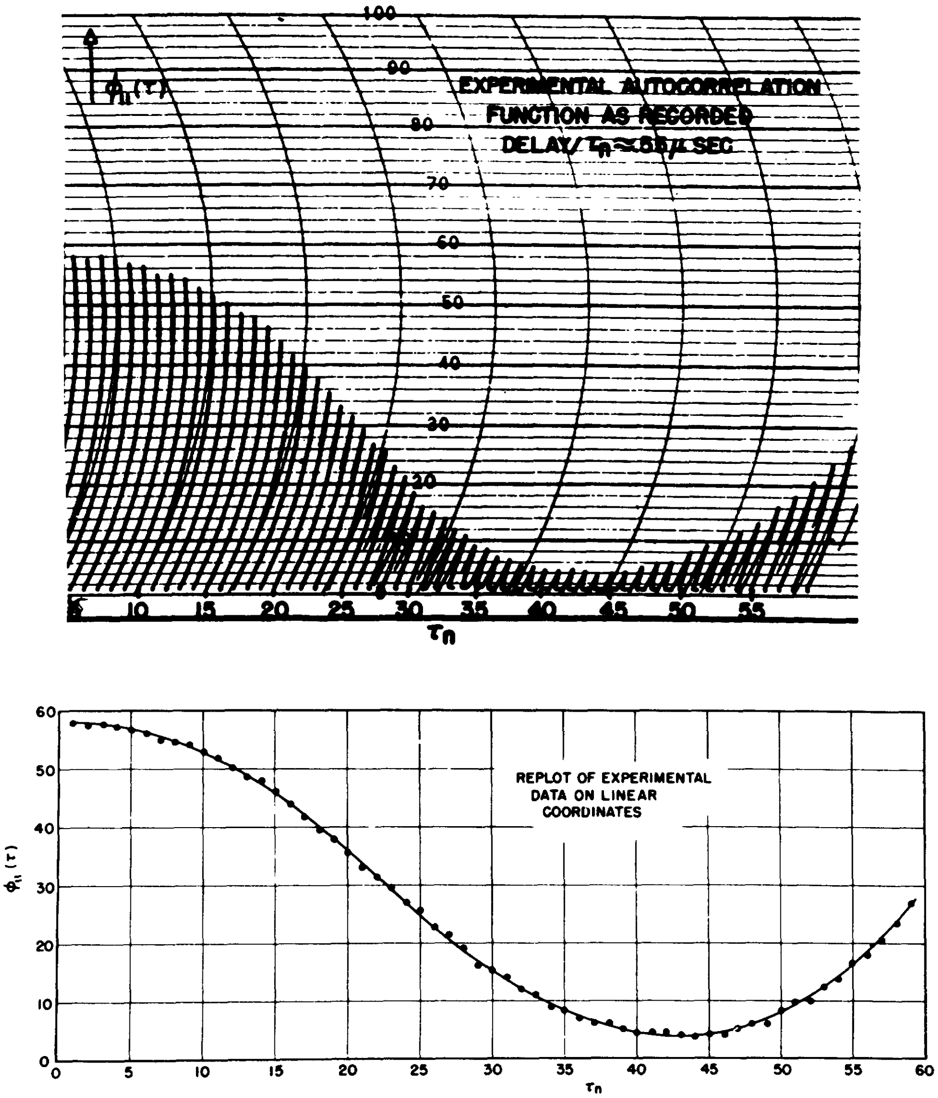


Fig. 44 Autocorrelation of noise from 884 gas tube after passing through single tuned circuit (Q = 14) showing in detail first portion of curve of Fig. 43.

4.42 Sampling Procedure

The construction of the electronic correlator was greatly simplified by using a periodic function for sampling. As long as the time functions being investigated are random in character no additional error is introduced by this technique. However, if a periodic component is present in the time series, a serious error may be introduced if there is any harmonic, or near-harmonic, relation between the periodic component and the sampling frequency. The effect introduced is a function of the correlating time β and may be evaluated in terms of the cross-correlation existing between the time series and the sampling voltage. This limitation interposed by the use of a periodic sampling frequency on the measurement of periodic components in a time series is important but not serious. A quick glance at the pulses at the output of the multiplier will generally suffice to determine the presence of a harmonic relationship, which can be easily corrected by a small change in the sampling frequency. In order to avoid all chances of error from this effect, theoretically one should use narrow-band noise as a sampling function. It would be a desirable feature, if this technique is used, to provide means for changing the center frequency and bandwidth of the noise.

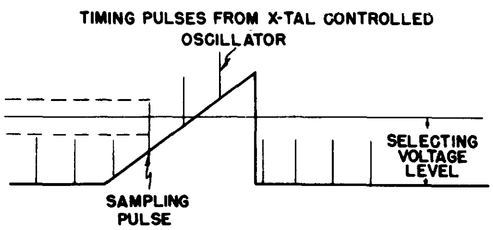


Fig. 45 A possible means of reducing jitter. Intersection of selection voltage level and first timing pulse determine timing of a sampling pulse.

4.43 Large Values of Delay τ

The correlator has been operated with a maximum delay of 2.5 msec. This value of τ_{max} can with only minor circuit changes be easily extended to 10 msec. It is pointed out that the storage requirement necessitated by large values of τ falls actually on only channel a. Although channel b stores the b_n points for an entire period, only that portion of the boxcar waveform defined by the maximum pulse width $(\dot{b}_n)$max is ever used. The boxcar waveform from channel a is used to modulate the amplitude of the multiplier pulses. Since it operates on the grid of the multiplier tube, it has the advantage of the gain in that stage. Consequently, the amplitude of the channel a boxcar need not be as great as the channel b. For these reasons a longer time constant is permissable in channel a, which should contribute to a substantial increase in allowable τ_{max}. It is the belief of the writer that an extension in τ_{max} beyond 10 msec by electronic means similar to those used in the correlator is an unnecessary struggle. Beyond this value of delay other means rapidly become more feasible. As a single suggestion in this direction, the time function under study could be stored on a twin-track magnetic tape, with the magnetic head of the recorder constructed with a multiple number of pickups spaced 10 msec apart in time as indicated in Fig. 46. The switching to appropriate pickups can be synchronized and controlled by the timing circuits of the correlator, so that a computation through all or any portion of the available range of τ can be selected and performed automatically. In this respect it is a desirable feature to modify the timing and control circuits of the correlator so that an appropriate number of τ steps can be selected within the range of 10 msec as desired. Since the number of steps must be synchronized to the distance set between pickups, not all numbers of τ steps are permissible.

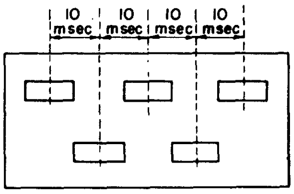


Fig. 46 Construction of a multiple pickup head to obtain larger values of τ.

Central to the suggestion presented here is the use of a magnetic tape to store the time function. As in the case of the correlator, it is not necessary for the entire time function to be stored. If the time function is continuous and available for long lengths of time, it is only necessary to use a sufficient length of tape in which to record, sample, and erase, as indicated in Fig. 47.

In summary, it is certain that the extension of τ_{max} in any general-purpose correlator is desirable. A preliminary investigation of the time functions important in the human body alone indicates the necessity for computation out to at least 100 msec and in some cases to one second. It is recommended that the present correlator be extended electronically to 10 msec delay and that larger delays be obtained with external delay systems controlled by appropriate timing and control circuits built in and available in the correlator. The use of storage tubes, when perfected, should not be overlooked as an external means of obtaining large delays.

4.44 Time Constant of the Integrator and Sensitivity Control

The time constant of the integrator should be made variable in discrete steps to correspond to various ranges of integration time. The time constants should be selected so that the average integration error in the various ranges is kept approximately constant. This will avoid the necessity of having to make large compensating changes in the sensitivity control. This modification will do much toward relieving the error introduced by high-sensitivity settings which reduce the cathode resistance of the cathode follower at the output of the integrator to a point of improper operation.

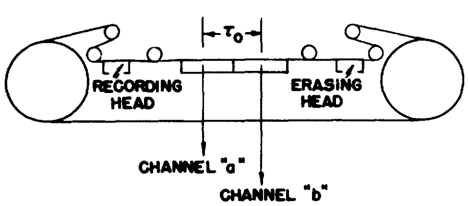


Fig. 47 Storage of a time function over a period T = time length between recording and erasing heads.

4.45 Discrete Selection

It is strongly recommended that most settings of the correlator be on a selector-switch basis. This will allow for the presetting of a given set of conditions and a return to these same conditions with a minimum of error. The present correlator modified for discrete selection of delays per step of 0.5, 1.0, 2.0, 5.0, 10.0, 20.0 and 30.0 μsec, would be a distinct improvement in ease and accuracy of measurement.

4.46 Measurement of Asymmetrical Time Functions

The boxcar waveforms of channel a and channel b are capacitor-coupled to the multiplier and pulse-width-modulator circuit respectively. Although long time constants are used, a small amount of error will be introduced by the averaging effect of the capacitor on asymmetrical waveforms. A clamper placed after the coupling capacitor and operating on the discharge pulse of the boxcar waveform would correct this error. This method of correction was suggested by L. G. Kraft, Research Laboratory of Electronics, M.I.T., and avoids the necessity of d-c coupling.

4.5 Other Possible Methods and Techniques of Measuring Correlation Functions

4.51 Digital Correlator

A serious difficulty in the measurement of correlation functions by electronic means is instability and temperature drift. For instance, if a time function requires an integration time (β) of two minutes to reach a reasonable statistical equilibrium, then the measurement of 90 points to determine its autocorrelation function will require about three hours. For accuracy, it is necessary that the conditions of the measurement remain constant for this period. The ability of digital techniques to store, multiply and integrate with great stability and reliability is one of the chief advantages of a digital correlator. Its disadvantages are all in terms of its size, order of complexity, and cost, which serve to classify it as a technique for use in the larger research laboratories.

4.52 Shorted Transmission Line

A shorted transmission line can be used as a central part of a device for computing correlation functions. A probe placed in a shorted line will pick up the sum of the direct wave and the reflected wave as indicated in Fig. 48. The mean of the square of this sum is equal to two times the mean power of the time wave plus twice the autocorrelation function.

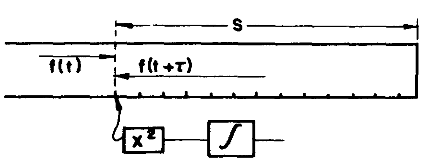


Fig. 48 Use of a shorted transmission line to measure autocorrelation functions.

Let $x = f_1(t) + f_1(t+\tau)$

$$x^2 = f_1^2(t) + 2f_1(t)f_1(t+\tau) + f_1^2(t+\tau)$$

$$\overline{x^2} = \psi_{11}(0) + 2\psi_{11}(\tau) + \psi_{11}(0)$$

$$= 2\psi_{11}(0) + 2\psi_{11}(\tau) \ . \qquad (40)$$

Therefore, a succession of probes placed in the shorted line (or waveguide) and connected to square-law devices and integrators will measure correlation functions of time waves within the frequency characteristics of the line. Means should be provided for removing the d-c component, $2\psi_{11}(0)$. It is clear that the only function of the short is to decrease the length of line required and an unshorted line used as a pure delay would theoretically serve as well. The principal advantage of the above method is in the measurement of correlation functions at extremely high frequencies where very small time delays are necessary and difficult to obtain by usual techniques. An application of this technique at lower frequencies has recently been made by P. E. A. Cowley and R. M. Fano, Research Laboratory of Electronics, M.I.T., in the construction of a short-time correlator for speech.

4.53 Magnetic Tapes, Rotating Drums, etc.

Magnetic tapes, rotating drums and similar devices for storing time functions may be modified into correlation computers by the addition of adjustable pickup heads (for determining) and suitable multipliers and integrators. All these devices and techniques have their value for specific applications but have a serious disadvantage in their limited range and inflexibility.

V. Experimental Applications of Correlation Functions

5.1 Evaluation of A Priori Information in the Detection of a Sinusoidal Signal in Normally Distributed Noise*

A sinusoidal signal E cos (wt + ϕ) is completely described by a knowledge of the three constants:

amplitude - E
angular frequency - w
phase angle - ϕ .

One knows intuitively from the concepts of information theory that if any a priori knowledge, either exact or weighted, is available of these three constants of the signal, the knowledge should be of some advantage in the detection of the sinusoid. Any detection system which does not take full advantage of all a priori knowledge of the signal will inherently operate below its maximum rate of transmission of information. In this sense we may define an ideal detector as one which takes complete advantage of all a priori information, i.e. it is a device which does not waste time remeasuring and computing what is already known.

The relative and practical advantage of an a priori knowledge of the above three

*The material presented here was suggested as a result of early conversations with Prof. J. B. Wiesner and Prof. Y. W. Lee, M.I.T., concerning possible applications of correlation functions in radar detection problems.

constants can be evaluated in terms of correlation and amplitude-density functions of the signal and the noise. In passing it should be noted that the practical ideal detector as defined above will inherently be nonlinear since

$$f(\text{signal}) = f(E,\ w,\ \phi) = \text{implicit nonlinear function} \ .$$

An evaluation of an a priori knowledge of each of the three constants is considered separately, and, for convenience of organization, the angular frequency constant w is taken first, in order that the properties of autocorrelation and its transition to crosscorrelation may be used as an introduction to the notions involved.

5.11 Autocorrelation

In order to arrive at a clearer picture of the basic concepts involved in the evaluation of an a priori knowledge of frequency w, consider first the general case of two time functions $f_1(t)$ and $f_2(t)$, both consisting of a signal plus random noise as indicated by Eq. 41.

$$\begin{aligned} f_1(t) &= S_1(t) + N_1(t) \\ f_2(t) &= S_2(t) + N_1(t) \ . \end{aligned} \tag{41}$$

The crosscorrelation of these two time functions is made up of four component crosscorrelation terms

$$\begin{aligned} \psi_{12}(\tau) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f_1(t) f_2(t+\tau)\,dt \\ \psi_{12}(\tau) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left[S_1(t) + N_1(t)\right]\left[S_2(t+\tau) + N_2(t+\tau)\right] dt \\ \psi_{12}(\tau) &= \psi_{S_1S_2}(\tau) + \psi_{S_1N_2}(\tau) + \psi_{N_1S_2}(\tau) + \psi_{N_1N_2}(\tau) \ . \end{aligned} \tag{42}$$

If the two time functions $f_1(t)$ and $f_2(t)$ are independent, then each of the four component terms of Eq. 42 will be identically zero as the period of correlation approaches infinity. For a finite time of correlation, however, the standard deviation of the points of the correlation curve will approach zero as a function of the number of discrete independent samples taken by the correlator.*

Consider the properties of autocorrelation as applied to a practical problem of detecting a periodic signal in random noise where the only a priori knowledge of the signal is, perhaps, its probable location within a general broad bandwidth. In terms of

*It was early recognized by Prof. J. B. Wiesner that the inherent dispersion of points, resulting from a finite correlation time, imposed the basic limitation on correlation techniques for detection purposes.

the previous general case, $f_2(t)$ is identically $f_1(t)$, both channels of the correlator operating in parallel from the same input. Here again the correlation function will be made up of four terms

$$\psi_{11}(\tau) = \psi_{SS}(\tau) + \psi_{SN}(\tau) + \psi_{NS}(\tau) + \psi_{NN}(\tau) \ . \tag{43}$$

The first term is the autocorrelation of the signal component: it will be nonzero. The next two terms are the crosscorrelation of the signal and noise components; these, theoretically, in the limit as T approaches infinity, will be identically zero. The fourth term is the autocorrelation of the noise component; theoretically, this term will be zero for large τ and can be made zero for all τ by compensation. The first term is a measure of our desired signal, and the last three are essentially noise terms whose dispersion approaches zero as a function of the number of samples computed by the correlator. It is important to note that there are at least three effective noise terms and actually a fourth if one includes an inherent amount of dispersion in the measurement of the signal term in a finite length of time.

It is convenient to evaluate the use of autocorrelation in the detection of a signal in noise in a finite time interval in terms of the rms signal-to-noise ratio improvement at the output of the correlator. This improvement can be determined as a function of the rms signal-to-noise ratio at the input and the length of correlation time.

It should be noted that the crosscorrelation terms of Eq. 43, as well as the autocorrelation term for the noise component (for large τ), are computed from the average product of two independent variables. The contribution of the effective noise terms at the output of the correlator can, therefore, be evaluated in terms of the dispersion of the product of two independent variables, which is shown below to be equal to the product of the individual dispersions, providing first moments are zero.*

Let x and y be two independent variables and define $z = xy$. Then

$$E(x) = \text{expectation of } x = m_x$$

$$E(y) = \text{expectation of } y = m_y \ .$$

Since x and y are independent

$$E(xy) = E(x)E(y) = m_x m_y \ .$$

By definition

$$\text{dispersion of } x = \sigma_x^2 = E(x^2) - [E(x)]^2 = E(x^2) - m_x^2$$

$$\text{dispersion of } y = \sigma_y^2 = E(y^2) - [E(y)]^2 = E(y^2) - m_y^2$$

*A method of proof indicated to the writer by Dr. M. Loewenthal, formerly of Research Laboratory of Electronics, M.I.T.

and

$$\text{dispersion of } z = \sigma_z^2 = E\left[(xy)^2\right] - \left[E(x+y)\right]^2 = E(x^2y^2) - \left[E(x)E(y)\right]^2$$
$$= E(x^2)E(y^2) - m_x^2 m_y^2 .$$

Then from above

$$E(x^2) = \sigma_x^2 + m_x^2$$
$$E(y^2) = \sigma_y^2 + m_y^2 .$$

Substituting

$$\sigma_z^2 = (\sigma_x^2 + m_x^2)(\sigma_y^2 + m_y^2) - m_x^2 m_y^2$$
$$\sigma_z^2 = \sigma_x^2 \quad \sigma_y^2 + m_x^2 \quad \sigma_y^2 + m_y^2 \sigma_x^2 . \tag{44}$$

For our case, of a sinusoid signal and gaussian noise, both $m_x = m_y = 0$, and therefore

$$\sigma_z^2 = \sigma_x^2 \sigma_y^2 . \tag{45}$$

Equation 45 makes it possible to evaluate the dispersion of each component term of Eq. 43 and hence the aggregate rms noise term at the output of the correlator for a finite averaging time.

At input to the correlator: if the signal is E $\cos(wt + \phi)$, and the noise is normally distributed with a standard deviation, then the input rms signal-to-noise ratio is given by

$$\left.\frac{\text{Signal}}{\text{Noise}}\right|_{\text{rms}} = \frac{E}{\sqrt{2}\,\sigma} .$$

At output of the correlator: the measure of the desired signal is the self-correlation signal component term of Eq. 43. Analytically, it is given by

$$\psi_{SS}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left[E \cos(wt + \phi)\right]\left[E \cos(w(t + \tau) + \phi)\right]dt$$

$$\psi_{SS}(\tau) = \frac{E^2}{2} \cos w\tau \quad . \tag{46}$$

The above expression for $\psi_{SS}(\tau)$ is its theoretical value. The electronic correlator practically measures a value which approaches close to

$$\psi_{SS}(\tau) = \frac{NE^2}{2K} \cos w\tau \tag{47}$$

where

N = number of samples computed

K = a variable normalizing factor .

The desired output rms signal component as measured by the correlator is therefore

$$\frac{NE^2}{2\sqrt{2}\,K} \;. \tag{48}$$

We need now to evaluate the rms contributions of the effective noise component terms at the output of the correlator. Let us set minimum τ sufficiently large so that the correlation of the noise at the discrete sampling points is effectively zero.* It has already been pointed out that this is an unnecessary restriction since the correlation of the noise, once measured and known a priori, can be compensated. However, in either case, we can assume our a_n and b_n sampled points of the correlator to be two discrete sets of independent variables X and Y.

$$X = x_1, x_2, x_3, x_4, \ldots x_N \} \quad \text{standard deviation } \sigma_x$$

$$Y = y_1, y_2, y_3, y_4, \ldots y_N \} \quad \text{standard deviation } \sigma_y \;.$$

Define Z = XY and assume, as appropriate to our case, that $m_x = m_y = 0$. Then

$$\left.\begin{aligned} Z &= x_1y_1, x_2y_2, x_3y_3, x_4y_4, \ldots x_Ny_N \\ &= z_1, \quad z_2, \quad z_3, \quad z_4, \ldots z_N \end{aligned}\right\} \quad \text{standard deviation } \sigma_x\sigma_y \text{ (see Eq. 45)}$$

The repeated summation of N discrete products of our new variable Z, as performed by the correlator, will give us a new set of points having a standard deviation $\sqrt{N}\,\sigma_x\sigma_y$

$$\sum_1^N Z = z_1 + z_2 + z_3 + z_4 \ldots + z_N \} \quad \text{standard deviation} \sqrt{N}\,\sigma_x\sigma_y \;. \tag{49}$$

This can be simply shown as follows: the standard deviation of the defined $\sum_1^N Z$ set of points is by definition

$$\sigma_{\Sigma Z} = \sqrt{E(\overline{\Sigma Z}^2) - E(\Sigma Z)^2}$$

but

$$E(\Sigma Z) = E(z_1) + E(z_2) + \ldots E(z_N) = 0$$

*The derivation of the probability density function for the product of two normally distributed and independent variables is given in Appendix 2, together with equations for the evaluation of all moments. It has not been necessary here to deal with the exact distribution function due to our simplified rms signal-to-noise criterion. The results of Appendix 2 and, in addition, the distribution density function of the product of a sine wave and noise would be necessary if a more searching criterion were adopted.

and

$$E(\overline{\Sigma Z}^2) = E(z_1^2 + z_2^2 + z_3^2 + \dots z_N^2 + 2z_1z_2 + 2z_1z_3$$

$$+ \dots 2z_2z_3 + 2z_3z_4 + \dots)$$

$$= E(z_1^2) + E(z_2^2) + E(z_3^2) + \dots E(z_N^2)$$

$$= N\sigma_x^2\sigma_y^2 \ . \qquad (50)$$

Therefore

$$\sigma_{\Sigma Z} = \sqrt{N\sigma_x^2\sigma_y^2} = \sqrt{N}\ \sigma_x\sigma_y \ . \qquad (51)$$

In the correlator, a normalizing factor, K, defined in Eq. 47, is also present in Eq. 51, so that experimentally the correlator measures

$$\sigma_{\Sigma Z} = \frac{\sqrt{N}\ \sigma_x\sigma_y}{K} \ .$$

Considering now each of the component correlation terms of Eq. 43, we have

1. $\psi_{SS}(\tau)$ = desired signal = $\frac{E^2}{2}\cos w\tau$
   dispersion in this measurement will have a standard deviation = $\frac{\sqrt{N}\,E^2}{2K}$
2. $\psi_{SN}(\tau) = 0$
   standard deviation from this value = $\sqrt{\frac{N}{2}}\ \frac{E\sigma}{K}$
3. $\psi_{NS}(\tau)$ same as $\psi_{SN}(\tau)$
4. $\psi_{NN}(\tau) = 0$ (for τ large, or after compensation)
   standard deviation from this value = $\frac{\sqrt{N}\,\sigma^2}{K}$.

Since the four effective noise terms are independent, their aggregate contribution can be evaluated as the square root of the sum of the squares. Therefore, for autocorrelation

$$\text{output } \frac{\text{signal}}{\text{noise}}\Bigg|_{\text{rms}} = \frac{\dfrac{NE^2}{2\sqrt{2}\,K}}{\sqrt{\dfrac{N\sigma^4}{K^2} + \dfrac{2NE^2\sigma^2}{2K^2} + \dfrac{NE^4}{4K^2}}} \qquad (52)$$

$$= \sqrt{\frac{N}{2}}\ \frac{E}{\sqrt{2}\,\sigma}\ \frac{1}{\sqrt{2 + \dfrac{2\sigma^2}{E^2} + \dfrac{E^2}{2\sigma^2}}}$$

If we define

$$Z = \text{input } \frac{\text{signal}}{\text{noise}}\bigg|_{\text{rms}} = \frac{E}{\sqrt{2}\,\sigma}$$

$$W = \text{output } \frac{\text{signal}}{\text{noise}}\bigg|_{\text{rms}}$$

Eq. 52 can be simplified symbolically to the following relation

$$W = \sqrt{\frac{N}{2}}\ \frac{Z^2}{Z^2 + 1} \quad . \tag{53}$$

The derivation of Eq. 53 serves to illustrate the fundamental notions and equivalent experimental techniques involved in the use of autocorrelation functions in the detection of a small signal in noise. While the relation of Eq. 53 is explicit and useful, its greatest importance lies in its role as an introduction to and part of the more general concept already discussed concerning the evaluation and use of a priori knowledge in the detection of a desired signal.

Figure 49 shows some examples of experimentally measured correlation curves that check Eq. 53 fairly closely. The data from these curves have been plotted on Fig. 50, discussed in the next section.

5.12 A Priori Knowledge of Angular Frequency W

With an a priori knowledge of the exact frequency of the signal we wish to detect we can locally generate as an input to the second channel of the correlator a sinusoid having the same frequency, thus measuring crosscorrelation. In this case, we shall find that the crosscorrelation $\psi_{12}(\tau)$ is made up of the desired signal correlation component and only one noise term.

$$f_1(t) = S_1(t) + N_1(t)$$

$$f_2(t) = S_2(t)$$

$$\psi_{12}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \Big[S_1(t) + N_1(t)\Big]\Big[S_2(t+\tau)\Big]dt$$

$$\psi_{12}(\tau) = \psi_{S_1S_2}(\tau) + \psi_{N_1S_2}(\tau) \quad . \tag{54}$$

We see, therefore, that crosscorrelation has reduced the effective number of noise terms from three to one or, if we include the effect of a certain dispersion in the measurement of the signal component in a finite time interval, from four to two. The quantitative advantage of this over autocorrelation is derived below in a manner similar to Eq. 53, but with a parameter a introduced to show the transition from autocorrelation to crosscorrelation, thus allowing for a possible corresponding physical method of

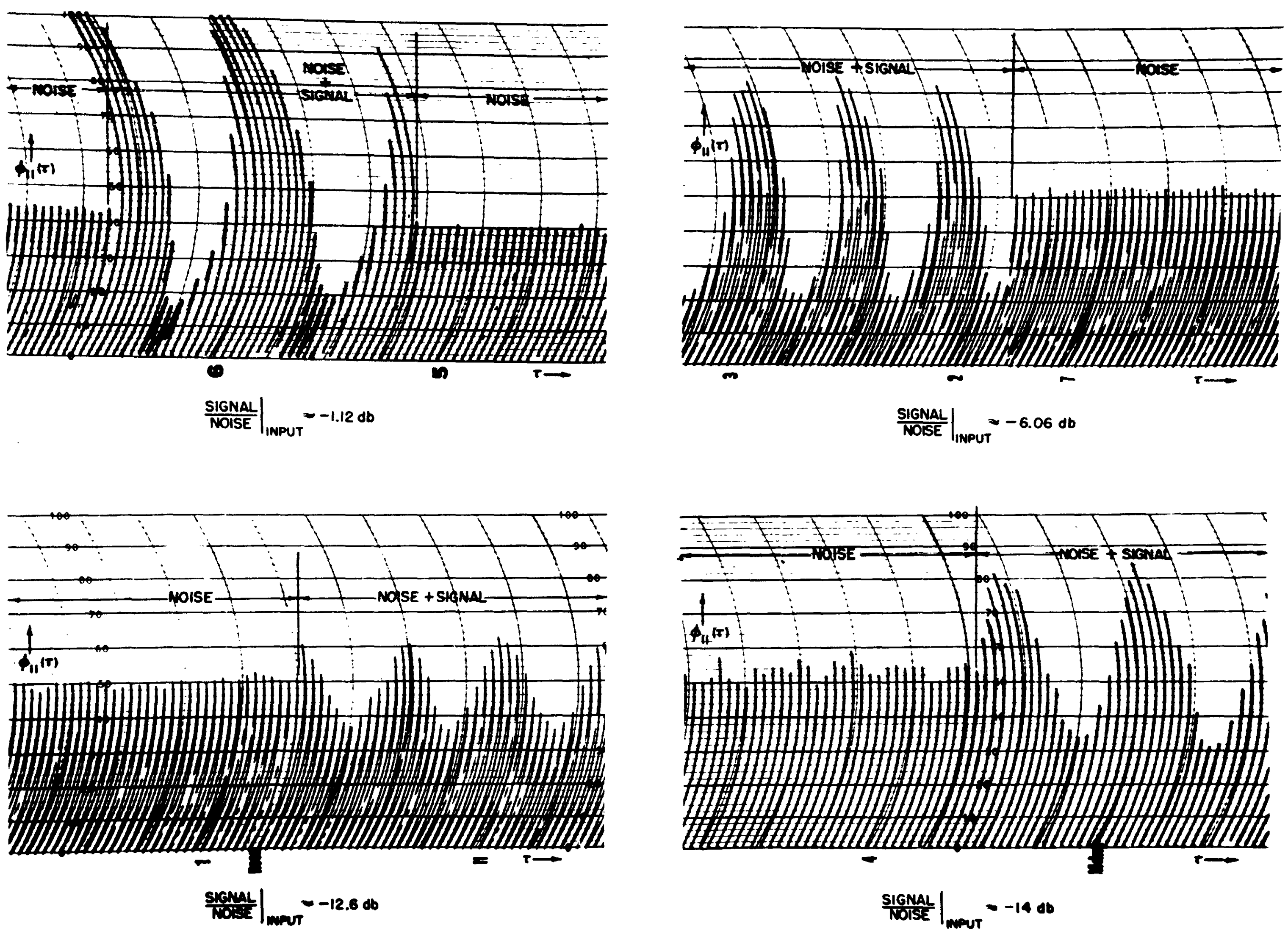


Fig. 49 Experimental results showing use of autocorrelation in the detection of a pure tone (8kc sine wave) in random noise.

utilizing any a priori knowledge of the probable signal location in frequency.

The generalized problem of the detection of a signal in noise through autocorrelation and crosscorrelation as outlined above can be presented in the following manner. Let

$$f_1(t) = S_1(t) + N_1(t)$$

$$f_2(t) = S_2(t) + N_2(t)$$

where $N_1(t)$ is the noise from a bandwidth B, and $N_2(t)$ is the noise from a reduced bandwidth aB, where $a < 1$. If the rms value of $N_1 = \sigma$, then $N_2\big|_{rms} = \sqrt{a}\,\sigma$. Also, if $S_1(t) = E \cos wt$, let $S_2(t) = mE \cos(wt + \phi)$ where $m = |S_2|/|S_1|$, and ϕ = phase angle between $S_1(t)$ and $S_2(t)$. By substitution in Eq. 2, the crosscorrelation of $f_1(t)$ and $f_2(t)$ is found to be

$$\psi_{12}(\tau) = \psi_{S_1S_2}(\tau) + \psi_{S_1N_2}(\tau) + \psi_{N_1S_2}(\tau) + \psi_{N_1N_2}(\tau) \quad . \tag{55}$$

In a fashion analogous in its reasoning to the specific case of autocorrelation considered above, the effective signal and noise component terms can be evaluated as follows

1. $\psi_{S_1S_2}(\tau)$ = desired signal = $\frac{mNE^2}{2K} \cos(wt + \phi)$

   dispersion in this measurement will have a standard deviation = $\frac{m\sqrt{N}\,E^2}{2K}$

2. $\psi_{S_1N_2}(\tau) = 0$

   standard deviation from this value = $\sqrt{\frac{aN}{2}}\ \frac{E\sigma}{K}$

3. $\psi_{N_1S_2}(\tau) = 0$

   standard deviation from this value = $m\sqrt{\frac{N}{2}}\ \frac{E\sigma}{K}$

4. $\psi_{N_1N_2}(\tau) = 0$ (for τ large, or after compensation)

   standard deviation from this value = $\sqrt{Na}\ \frac{\sigma^2}{K}$

Therefore

$$\text{output } \frac{\text{signal}}{\text{noise}}\bigg|_{rms} = \frac{\dfrac{mNE^2}{2\sqrt{2}K}}{\sqrt{\dfrac{m^2NE^4}{4K^2} + \dfrac{aNE^2\sigma^2}{2K^2} + \dfrac{m^2NE^2\sigma^2}{2K^2} + \dfrac{Na\sigma^4}{K^2}}}$$

$$= \sqrt{\frac{N}{2}}\ \frac{E}{\sqrt{2}\,\sigma}\ \frac{1}{\sqrt{1 + am^2 + \dfrac{E^2}{2\sigma^2} + \dfrac{2am^2\sigma^2}{E^2}}} \quad . \tag{56}$$

If we define as before

$$Z = \text{input}\left.\frac{\text{signal}}{\text{noise}}\right|_{\text{rms}} = \frac{E}{\sqrt{2}\,\sigma}$$

and

$$W = \text{output}\left.\frac{\text{signal}}{\text{noise}}\right|_{\text{rms}}$$

we may write Eq. 56 symbolically as

$$W = \sqrt{\frac{N}{2}}\,\frac{Z^2}{\sqrt{Z^4 + Z^2(1 + am^2) + am^2}} \; . \tag{57}$$

Equation 57 has several interesting interpretations. If we let a = m = 1, we see that it reduces both physically and analytically to Eq. 53 for the case of autocorrelation. For the case of crosscorrelation, a is equal to zero, and it is interesting to observe that W is then independent of m. That is the advantage of knowing, a priori, that the frequency of our desired signal is exact and complete and independent of the amplitude parameter of the crosscorrelated signal; our advantage is solely determined by the extent of our a priori knowledge. A physical interpretation can be given for values of a other than 0 (crosscorrelation) and 1 (autocorrelation). For intermediate values of a, the process is essentially autocorrelation, but a strong a priori probability of the signal's location in a restricted bandwidth makes it possible to shape the bandwidth characteristic of one channel of the correlator to fit this a priori knowledge and hence take advantage of it by reducing a, and increasing the input rms signal-to-noise ratio. It should be noted that physical reasoning restricts m to be equal to unity except when a is equal to zero (crosscorrelation). This physical requirement is consistent with Eq. 57, since a and m are wedded together as a single parameter.

Equation 57 expressed in decibels has been plotted in Fig. 50 with $am^2 = 1$ (autocorrelation) and $am^2 = 0$ (crosscorrelation) as a function of the input signal-to-noise ratio in decibels for N (the number of independent samples) equal to 60, 000. From Fig. 50, we see that if the input signal-to-noise ratio is -15 db, then the output signal-to-noise ratio from autocorrelation is +14.5 db. If the frequency of the signal under detection is known, however, then crosscorrelation will increase the output signal-to-noise ratio, for the same correlation time, to +30 db.

The difference between the autocorrelation and crosscorrelation curves of Fig. 50 is the advantage to be gained as a result of an a priori knowledge of frequency. In Fig. 51, this advantage is shown to be an approximately logarithmic function of the input signal-to-noise ratio. It is significant to note that this curve is independent of N, the number of samples, and hence the averaging time. For example, if a certain output signal-to-noise ratio in decibels could have been obtained through autocorrelation for an input signal-to-noise ratio of, say, -40 db, then a possibility of having switched to crosscorrelation through a knowledge of the frequency constant (w) will give a +40 db improvement.

As a point of considerable importance, it should be noted that an attempt to use crosscorrelation in the detection of a signal of unknown frequency, and hence with a flat a priori probability of location within a broad bandwidth, by hunting will give no advantage over autocorrelation.*

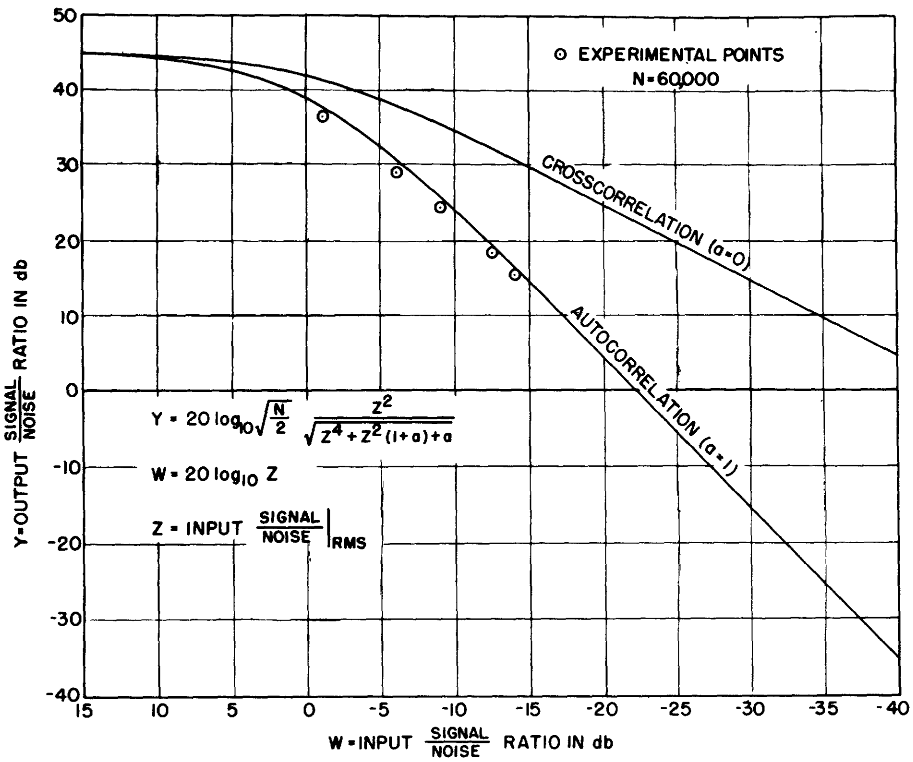


Fig. 50 Improvement in signal-to-noise ratio through autocorrelation and crosscorrelation.

5.13 Advantage of A Priori Knowledge of Phase Angle Constant, ϕ

In the case of crosscorrelation, it should be noted that since no knowledge is assumed as to the phase angle constant ϕ, of the desired signal, the maximum of the signal crosscorrelation component must be determined. This will require a measurement of the correlation curve at a minimum of two different values of τ. For example, it would be convenient to take values of τ such that $w\tau_2 - w\tau_1$ equals 90°, in which case the amplitude of the signal at the output of the correlator is measured by the square root of the sum of the squares of the two measurements about the mean. This is illustrated in Fig. 52.

*This fact was pointed out to the writer by Prof. J. B. Wiesner, Research Laboratory of Electronics, M.I.T.

However, if it is assumed that both the frequency and phase angle constants of the desired signal are known, there will be no ambiguity as to the position of the maximum since, as already pointed out in Sec. 3.2, it must occur periodically at $\tau = 0$, T, 2T, etc. This gives us an advantage of a factor of 2 in time, representing a constant 3 db improvement in output signal-to-noise ratio. This improvement is independent of the input signal-to-noise ratio.

5.14 Advantage of A Precise A Priori Knowledge of Amplitude E

If we know the amplitude constant E of a sinusoid (x), then statistically we know its amplitude distribution density function p(x) precisely. It should be noted in this respect that p(x) is independent of the frequency or phase angle constant of the sinusoid, and is uniquely determined by E only.

If we define the signal as a variable X and the normally distributed noise as a variable Y, we may, after the method and definitions of Wiener, outlined in Sec. II, ask what is the gain in information made through a measurement W = X + Y. In terms of our a priori distribution p(x) we ask what is $p_w(x)$ - the probability of x, knowing w -,

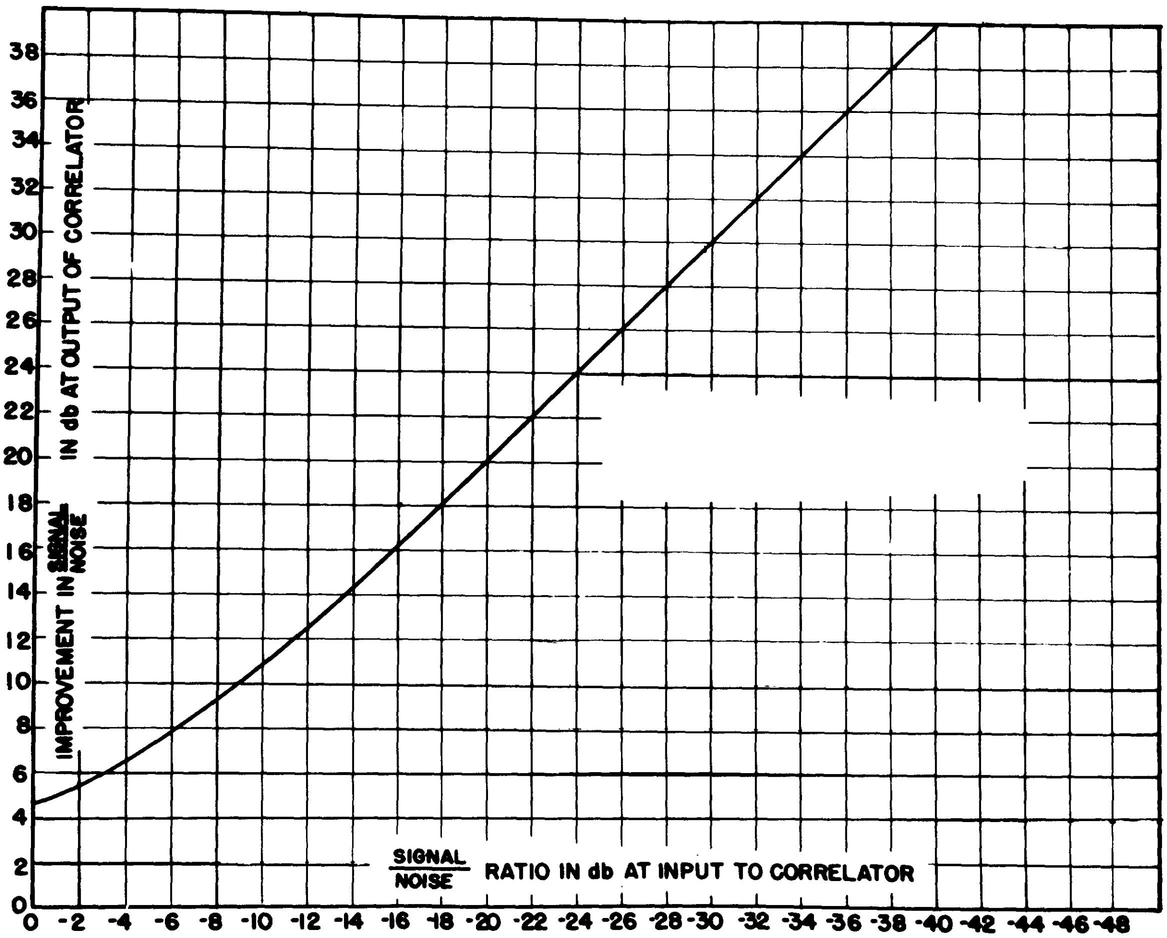


Fig. 51 Improvement in signal-to-noise ratio through crosscorrelation for a pure tone located in white noise.

and what is the amount of information associated with each of the curves p(x) and $p_w(x)$?

The evaluation of this amount of information (or negative entropy) is given below with the slight modification that the result is used as a measure of the improvement in signal-to-noise ratio that should be possible through the use of our previously defined ideal detector.

Let X = sinusoidal signal

Y = normally distributed noise.

Then

$$p(x)dx = \frac{dx}{\sqrt{E^2 - x^2}}, \text{ where } E = \text{amplitude of the sinusoid} \tag{58}$$

$$p(y)dy = \frac{1}{\sqrt{2\pi\psi_o}} e^{-\frac{y^2}{2\psi_o}} dy, \text{ where } \psi_o = \text{mean square value of the noise} \tag{59}$$

We desire to know the probability of x, knowing w, $\left[p_w(x)\right]$, where W = X + Y. This can be evaluated through the use of Laplace's extension of Baye's theorem.

$$p_w(x) = \frac{p(x)p_x(w)}{p(w)} . \tag{60}$$

In the above

$$p_x(w) = \left\{p(y), \text{ where } y = w - x\right\} = p(y = w - x)$$

$$p_x(w) = \frac{1}{\sqrt{2\pi\psi_o}} e^{-\frac{(w-x)^2}{2\psi_o}} . \tag{61}$$

In order to evaluate p(w), we note that since x and y are independent, we may write

$$p(x, y)dx\, dy = \frac{1}{\pi\sqrt{E^2 - x^2}\sqrt{2\pi\psi_o}} e^{-\frac{y^2}{2\psi_o}} dx\, dy . \tag{62}$$

To transform the above distribution function into a function of the defined variable W, an additional variable of integration, u, may be defined as follows

$$W = X + Y$$

$$u = X . \tag{63}$$

The relation between the old and new variables of the distribution function is

$$p(w, u) = \frac{p(x, y)}{\frac{\partial(w, u)}{\partial(x, y)}} \tag{64}$$

with the Jacobian of the transformation

$$\frac{\partial(w,u)}{\partial(x,y)} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1 \; . \tag{65}$$

By substitution in Eq. 62

$$p(w,u) = \frac{1}{\pi\sqrt{2\pi\psi_o}\sqrt{E^2 - u^2}}\, e^{-\frac{(w-u)^2}{2\psi_o}} \; . \tag{66}$$

p(w) is obtained by integrating over all values of u.

$$p(w) = \frac{1}{\pi\sqrt{2\pi\psi_o}} \int_{-E}^{E} \frac{e^{-\frac{(w-u)^2}{2\psi_o}}}{\sqrt{E^2 - u^2}}\, du \; . \tag{67}$$

Then by substitution in Eq. 60

$$p_w(x) = \frac{e^{-\frac{(w-x)^2}{2\psi_o}}}{\sqrt{E^2 - x^2}\displaystyle\int_{-E}^{E} \frac{e^{-\frac{(w-u)^2}{2\psi_o}}}{\sqrt{E^2 - u^2}}\, du} \; . \tag{68}$$

The gain in information as measured by $p_w(x)$ and $p(x)$ is then

$$\int_{-\infty}^{\infty} dw \int_{-E}^{E} \frac{e^{-\frac{(w-x)^2}{2\psi_o}}}{\sqrt{E^2 - x^2}\displaystyle\int_{-E}^{E} \frac{e^{-\frac{(w-u)^2}{2\psi_o}}}{\sqrt{E^2 - u^2}}\, du} \log_2 \frac{e^{-\frac{(w-x)^2}{2\psi_o}}}{\sqrt{E^2 - x^2}\displaystyle\int_{-E}^{E} \frac{e^{-\frac{(w-u)^2}{2\psi_o}}}{\sqrt{E^2 - u^2}}\, du}\, dx$$

$$- \int_{-E}^{E} \frac{1}{\pi\sqrt{E^2 - x^2}} \log_2 \frac{1}{\pi\sqrt{E^2 - x^2}}\, dx \; . \tag{69}$$

The evaluation of Eq. 69 is extremely difficult, due to the form of Eq. 67, which, in an alternate form, has been evaluated by S. O. Rice (16) in a series expansion over only a restricted range of the variable; for the major range of values he has resorted to a numerical integration.

The writer, after several fruitless analytical attempts, has resorted to a numerical

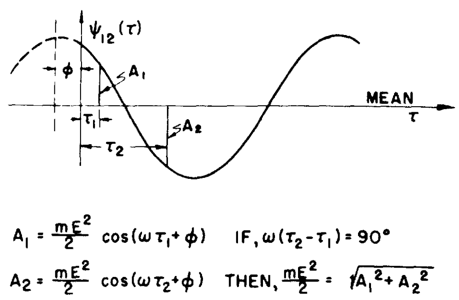


Fig 52 Two-point determination of maximum value of the crosscorrelation function for a sinusoid of known frequency.

evaluation, based on replacing the signal and noise distribution functions by discrete functions of sufficient partitioning to give a reasonable qualitative estimate of the behavior of p(w) over an interesting range of rms signal-to-noise ratios. Figure 53 shows the discrete models that were used to approximate the continuous functions. The amplitude of the sinusoid has been arbitrarily taken as equal to one and variations in signal-to-noise ratios are obtained by operation on Y.

Graphically the computation of p(w) for the continuous variables is indicated in Fig. 54. The graphical picture serves as a guide for the efficient programing of the computations and also to indicate the range over which our model is a reasonable approximation. By holding E = 1 = constant, it is clear that evaluations will be more valid for small signal-to-noise ratios than they will be for large ratios; this fact has influenced the choice and method of computation, since we are in general more interested in the case where the signal is buried in the noise.

Table 1 gives a sample page of computation and serves to indicate the procedure used. Column 1 indicates the midpoint of our discrete levels. Column 2, p(x) x, is equivalent to the discrete model of the signal, Fig. 53. It is evaluated from the relation

$$\int_x^{x+\Delta x} \frac{1}{\pi\sqrt{1-x^2}} = \frac{1}{\pi}\sin^{-1}x\Big|_x^{x+\Delta x} .$$

Column 3 is taken from our discrete model* for the noise Y according to the relationship of Eq. 61. Column 4 is the product of Columns 2 and 3, and p(w) is then the sum of Column 4, since we have an exhaustive set of values of x.

$$p(w) = \sum_x p(x)p_x(w) . \qquad (70)$$

Column 5 is evaluated by dividing Column 4 by $p(w_1)$, making direct use of Eq. 60 (Baye's theorem). Columns 6 and 7 facilitate the computation of Column 8.

Computations similar to Table 1 were carried out for values of W from 0 to $\approx$5 at 0.2 intervals for rms signal-to-noise ratios of 0.500, 0.707 and 1.00. Figure 55 shows graphically the variations and trend in the function $p_W(x)$ for changes in W over a range 0 to 1.00 and for the three selected signal-to-noise ratios. Qualitatively, it is to be observed that our most probable location of x as a result of a measurement W sharpens with increasing W. That is, our knowledge of x as a result of a measurement W (and

*W.P.A. Tables.

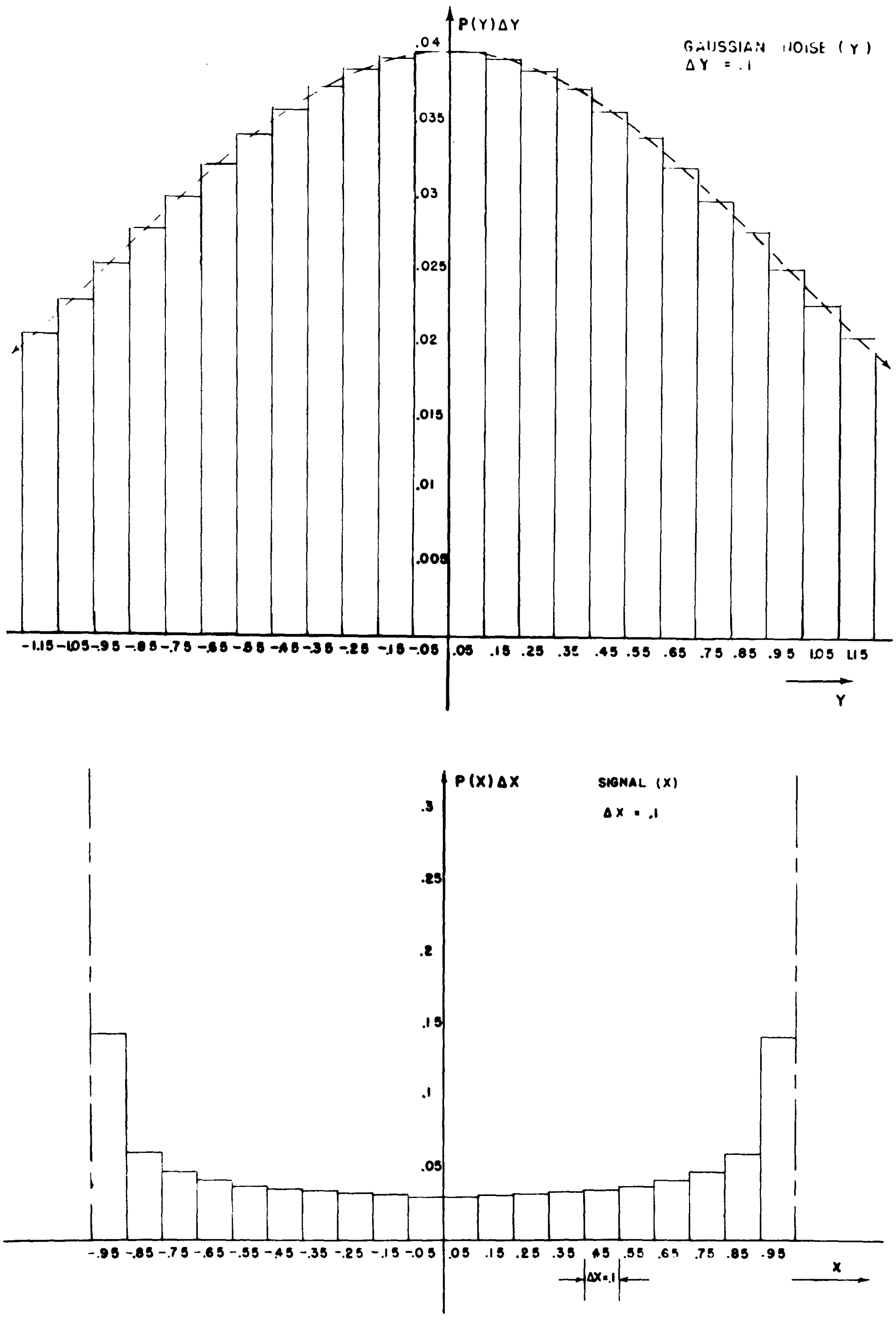


Fig. 53 Discrete functions used to approximate the probability density functions of a sinusoidal signal (x) of amplitude E = 1, and gaussian noise (Y).

Table 1: Illustrative example of method of computation.

$W_1 = 0.2, \qquad \psi_o = 1, \qquad E = 1$

1	2	3	4	5	6	7	8
x_m	$p(x)$	$p_x(w_1)$	$p(x)p_x(w_1)$	$p_{w_1}(x)$	$\frac{1}{p_{w_1}(x)}$	$-\log_{10}p_{w_1}(x)$	$-p_{w_1}(x)\log_{10}p_{w_1}(x)$
0.95	0.14356	0.02060	0.0029573	0.094407	10.59	1.01490	0.09581
0.85	0.06127	0.02299	0.0014086	0.044967	22.24	1.34713	0.06058
0.75	0.04833	0.02540	0.0012276	0.039189	25.52	1.40688	0.05513
0.65	0.04202	0.02780	0.0011682	0.037292	26.82	1.42830	0.05326
0.55	0.03814	0.03010	0.0011481	0.036651	27.28	1.43584	0.05263
0.45	0.03565	0.03229	0.0011511	0.036747	27.21	1.43473	0.05272
0.35	0.03407	0.03429	0.0011683	0.037297	26.81	1.42830	0.05327
0.25	0.03287	0.03604	0.0011848	0.037822	26.44	1.42226	0.05379
0.15	0.03222	0.03751	0.0012085	0.038580	25.92	1.41363	0.05454
0.05	0.03185	0.03865	0.0012310	0.039299	25.45	1.40569	0.05524
0.05	0.03185	0.03943	0.0012559	0.040092	24.94	1.39690	0.05600
0.15	0.03222	0.03983	0.0012833	0.040967	24.41	1.38757	0.05684
0.25	0.03287	0.03983	0.0013094	0.041799	23.92	1.37876	0.05763
0.35	0.03407	0.03943	0.0013435	0.042888	23.32	1.36773	0.05866
0.45	0.03565	0.03865	0.0013779	0.043986	22.73	1.35660	0.05967
0.55	0.03814	0.03751	0.0014308	0.045674	21.89	1.34025	0.06122
0.65	0.04202	0.03604	0.0015145	0.048348	20.68	1.31555	0.06360
0.75	0.04833	0.03429	0.0016573	0.052907	18.90	1.27646	0.06753
0.85	0.06127	0.03229	0.0019785	0.063160	15.83	1.19948	0.07576
0.95	0.14356	0.03010	0.0043211	0.137944	7.25	0.86028	0.11867

$p(w_1) = \sum_x p(x)p_x(w_1) = 0.031325$

$\sum_x p_w(x) =$ 1.00001

$H_{w_1}(x) = p_{w_1}(x)\log_{10}p_{w_1}(x) =$ 1.26257

$H_{w_1}(x)$, base 2 = 4.19416

an a priori knowledge of E) increases for larger and larger W. It would appear even at this stage that by selecting our measurements W we could increase our information on x accordingly.

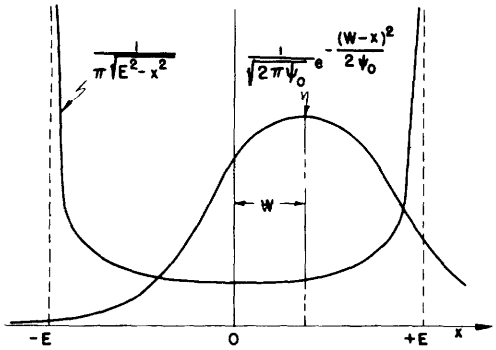


Fig. 54 Graphical representation of computation of p(w).

Figure 56 plots the functions p(w) for various rms signal-to-noise ratios. The curves are actually discrete with $\Delta w = 0.1$, but have been shown as continuous for ease of presentation.

Figure 57 is of particular interest since it plots the information associated with the $p_W(x)$ functions of Fig. 55 for the various signal-to-noise ratios. Each of the conditional entropy curves $H_W(x)$ are monotonic and their behavior with respect to the two reference lines H(x) and Max $H_W(x)$ is important. H(x) is the negative entropy associated with the curve p(x), i.e. it is equal to $-\sum p(x)\log_2 p(x)$. Max $H_W(x)$ represents the maximum negative conditional entropy which any disturbance can cause. Its finite value is due to the discrete and finite number of levels of our model. If p(x) could be treated as continuous, then Max $H_W(x)$ would be infinite. Max $H_W(x)$ serves as an upper-bound normalizing factor. The curves $H_W(x)$ are of course always less than Max $H_W(x)$, but they are for certain values of W greater than H(x). That is, whereas it is always true that on the average, entropy always increases (or what is the same; that negative entropy (information) always decreases), for some values of W our knowledge of x is less (more vague) than our a priori measure. (That the entropy on the average increases for our case is shown in Fig. 58.) Therefore, it would appear here that we have the possibility of a macroscopic Maxwell demon (3), which physically might be quite practical to construct. By constructing a clipping circuit which passed only those values of the sum W which exceeded the point of intersection of H(x) and $H_W(x)$, we would isolate those values of W that contained some measure of positive or confirming information about x. A reversed Maxwell demon is then constructed; it is actuated by energy level and stores up on one side of the gate greater information as to the presence of our signal.

There does not appear to be any obvious inconsistency here, since the construction of the Maxwell-demon clipper requires an a priori knowledge to be build into it. The success of the demon will be directly related to the preciseness of this knowledge.

The improvement in signal-to-noise ratio that should be possible is related to twice the area under the curves $H_W(x)$ of Fig. 57 from 0 to the point of intersection with H(x). These values normalized with respect to the mean of $H_W(x) = \sum_W p(w)H_W(x)$ are plotted on Fig. 59. It should be noted that Fig. 59 represents a gross or theoretical improvement. The clipping operation suggested here will change the character of the desired signal, and an additional operation is required to restore its original character. This

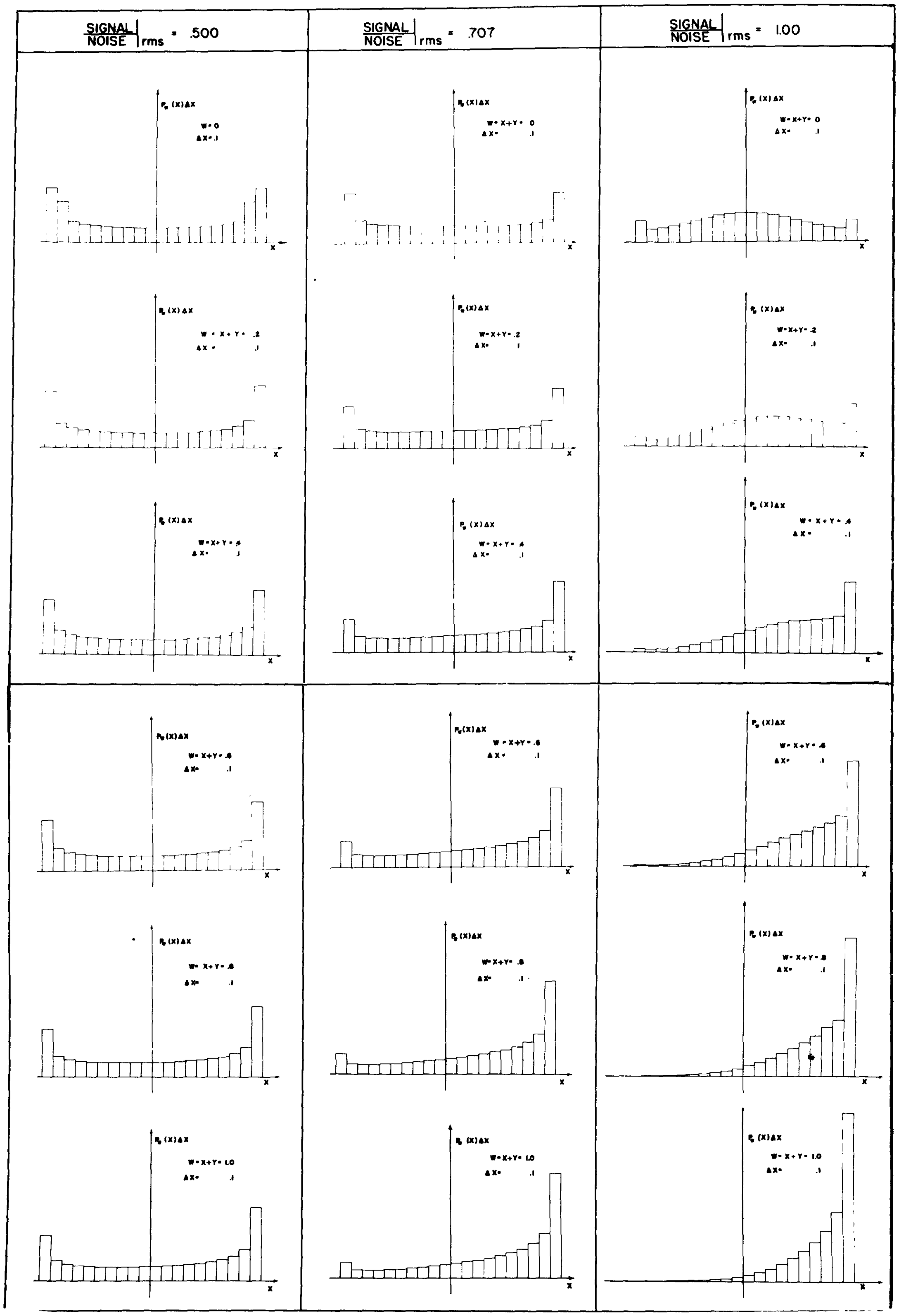


Fig. 55 Conditional probability functions, $p_w(x)$, for a sinusoidal signal (x) in gaussian noise (Y) where a sum $w = x + Y$ is measured. Both x and Y are approximated by discrete functions.

operation may lose information and the net gain will be the gross minus the loss occasioned by the retransformation.

Figure 60 plots optimum Maxwell-demon-clipping levels as a function of rms signal-to-noise ratio. It should be observed that with decreasing signal-to-noise ratios, the clipping level normalized with respect to σ_W - the rms value of the sum W, increases; this means that p(w) for $W > W_{clip}$ will be less, and hence a greater length of time will be required to build up the same amount of information or degree of assurance as to the presence of the sinusoid. This fact is again consistent with our notions of information theory.

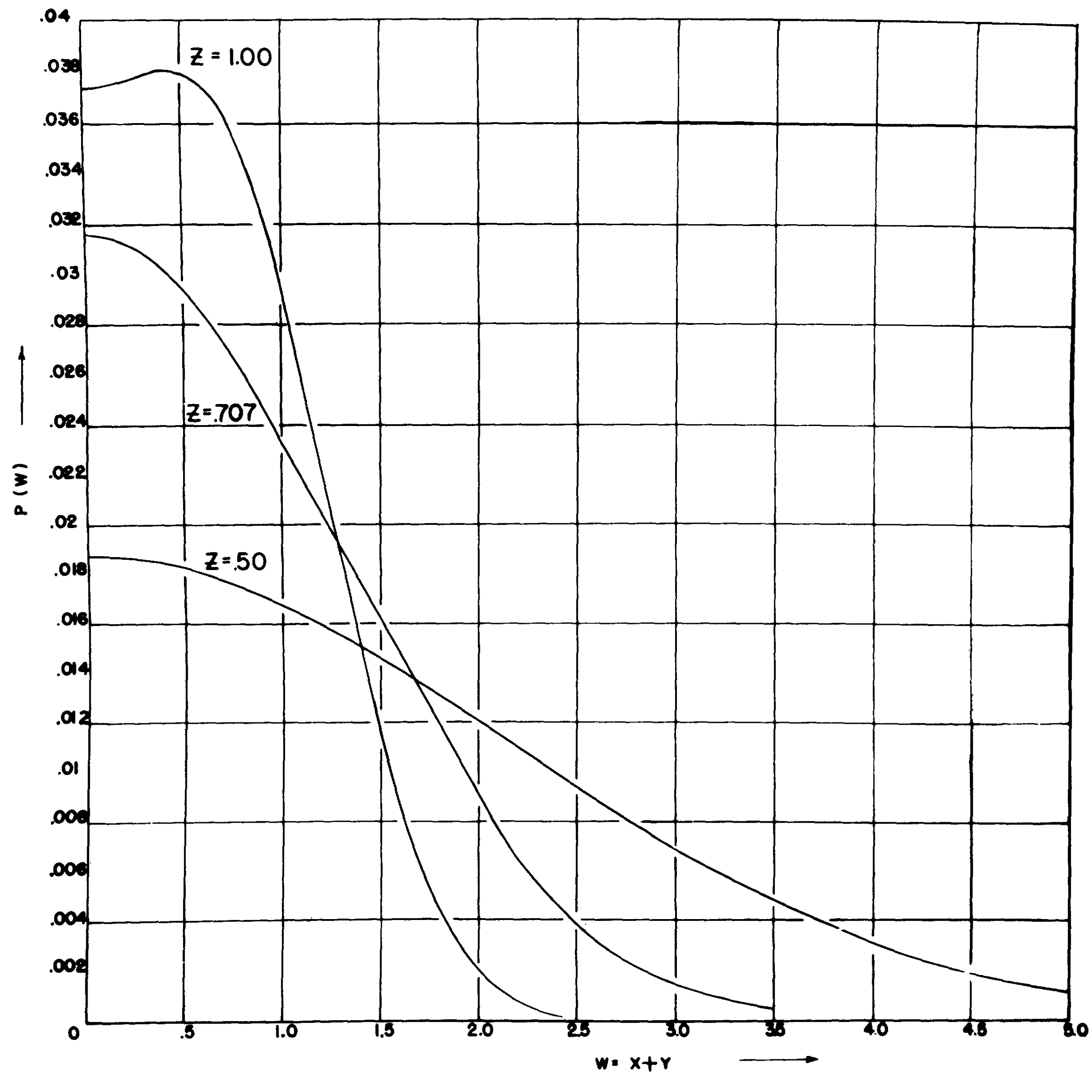


Fig. 56 Probability density functions p(w) for the sum (w) of a sinusoidal signal (x) in gaussian noise (Y). Both x and Y are approximated by discrete functions. $z = \frac{\text{signal}}{\text{noise}}$ rms .

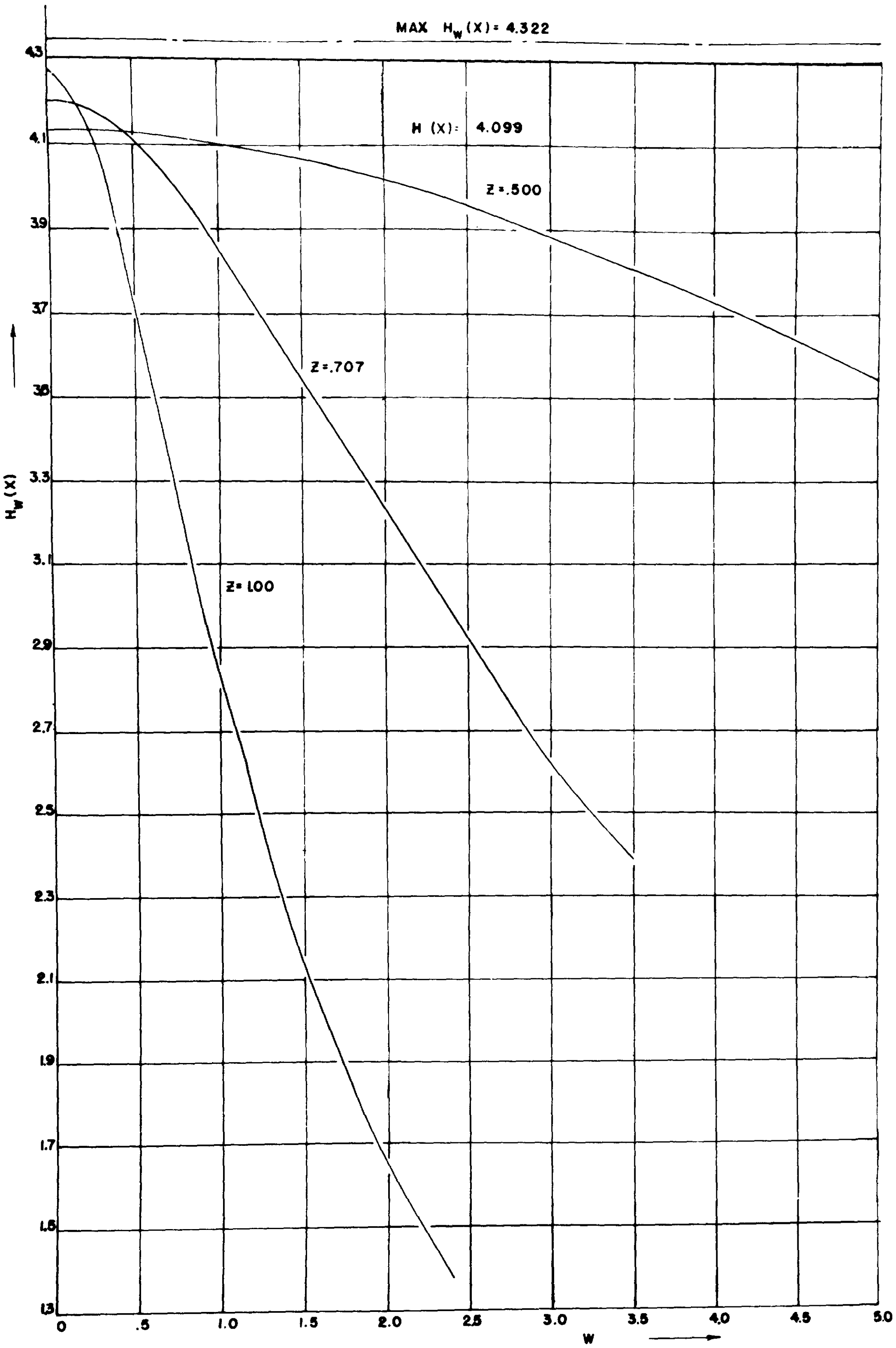


Fig. 57 Conditional entropy curves $H_w(x)$ for a sinusoidal signal (x) in gaussian noise (Y) where w = x + Y. Both x and Y are approximated by discrete functions.

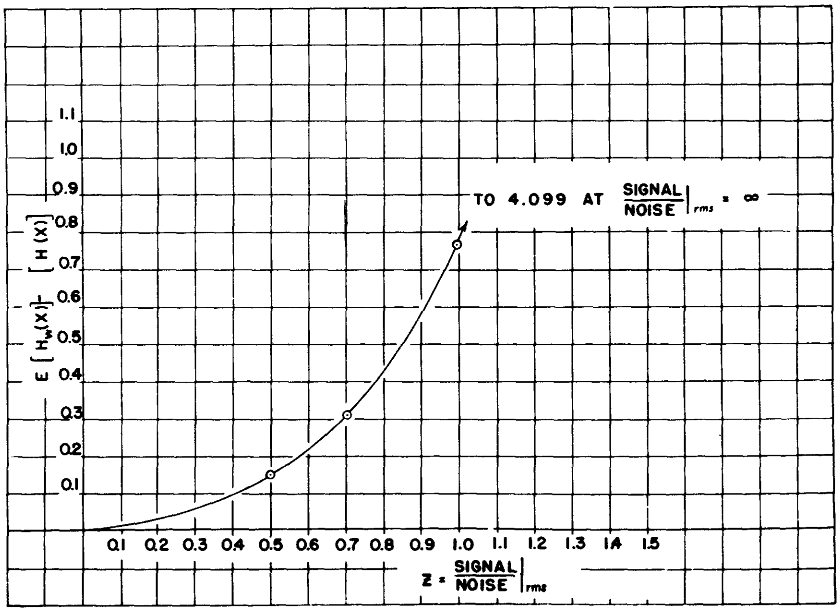


Fig. 58 Gain in information from $p_w(x)$ as a function of $\left.\frac{\text{signal}}{\text{noise}}\right|_{\text{rms}}$.

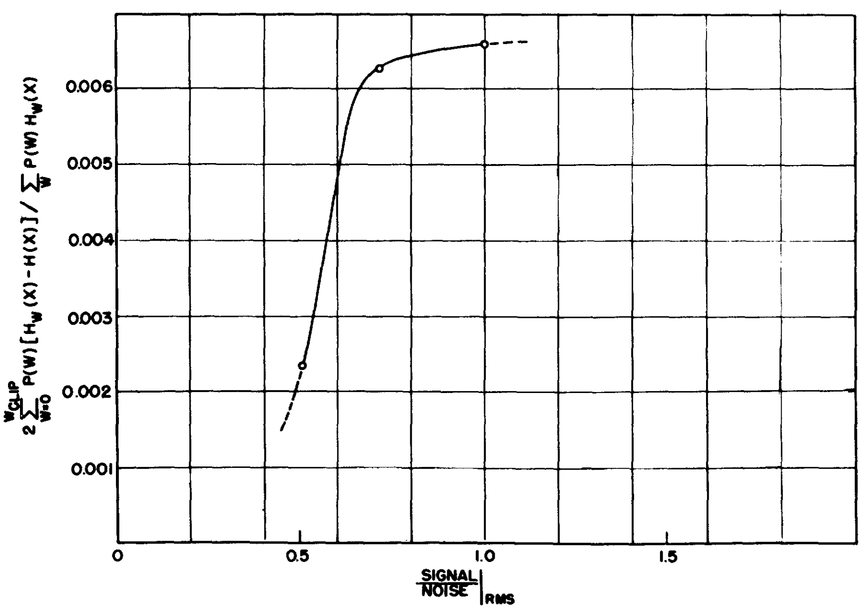


Fig. 59 Possible gain in information resulting from a clipping operation defined by $w_{clip} \cdot \left.\frac{\text{signal}}{\text{noise}}\right|_{\text{rms}}$.

The above evaluation and the discussion of this section do not leave us with as concrete an evaluation of an a priori knowledge of amplitude E as did the previous sections for frequency W and phase angle ϕ, but, qualitatively, it does indicate what should be possible, and it has given a possible physical means of transforming our evaluated negative entropy into an equivalent improvement in signal-to-noise ratio. Philosophically, the possibility of recognizing a Maxwell demon as being any device which has built into it a priori information is important.

5.2 Autocorrelation Functions of General Speech

A study of the statistical characteristics of speech including probability density distributions, conditional probability distributions and correlation functions under various conditions of voice, language and context is an important research problem in communication. A study of any one of these parameters, such as correlation functions, is certainly no less than a very long-term research study. No such exhaustive study has been attempted here, but a sufficient number of measurements have been made which, coupled with an available general knowledge of speech, allow certain conclusions to be drawn.

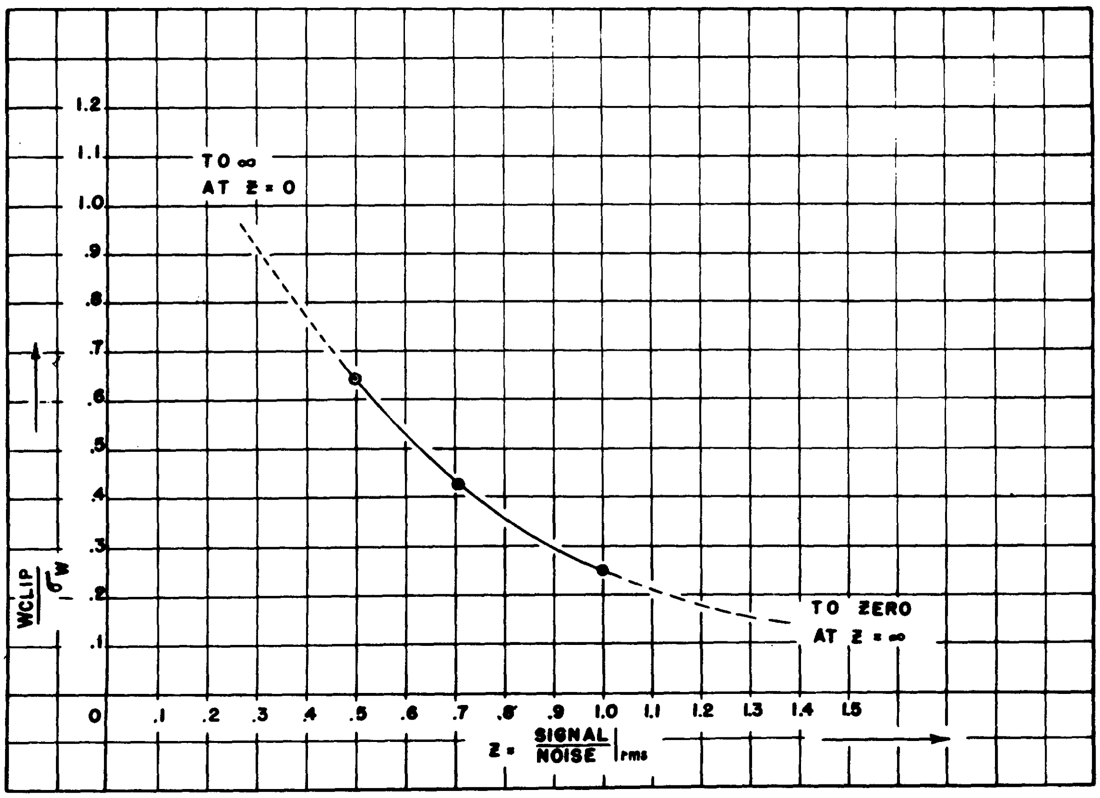


Fig. 60 Normalized Maxwell-demon clipping levels for the detection of a sinusoidal signal (x) in gaussian noise (Y), plotted as a function of $z = \frac{\text{signal}}{\text{noise}}\Big|_{rms}$. $\sqrt{w}$ = standard deviation of $w = x + Y = \sqrt{\frac{E^2}{2} + \psi_o}$.

From observation we know that the human ear operates in some sense as a statistical filter. If a person's voice is well known (that is, its inflections, intonations, mannerisms, etc.) one can pick it out of a high level of background interference with little trouble. We may also, having a particular stranger pointed out in a crowd of people, by concentrated listening gradually improve our ability to pick his conversation out from all others. We may describe this last phenomenon as the ability of the human ear* to learn. An attempt to explain the mechanism by which the human ear accomplishes the many wonderful things it does is a challenging research problem. It is too big a problem to be solved here, but it is important to show what possible role correlation functions could play in the more complex picture. We may ask if it is possible for the ear to differentiate or measure individual characteristics of speech on a spectra (or correlation) basis. A necessary condition for this possibility is that the correlation functions for a given selection of speech spoken by people of different voice characteristics should present measurable differences. Figure 61 shows the autocorrelation functions for an identical selection of speech spoken by a male and female voice in English and by a female voice after translation of the selection to Spanish. The speech selection was two minutes in length and the correlation function was computed out to 1.25 msec. The evidence shown in Fig. 61 is by no means conclusive, but the differences indicated by the curves are in agreement with our general feeling and knowledge of

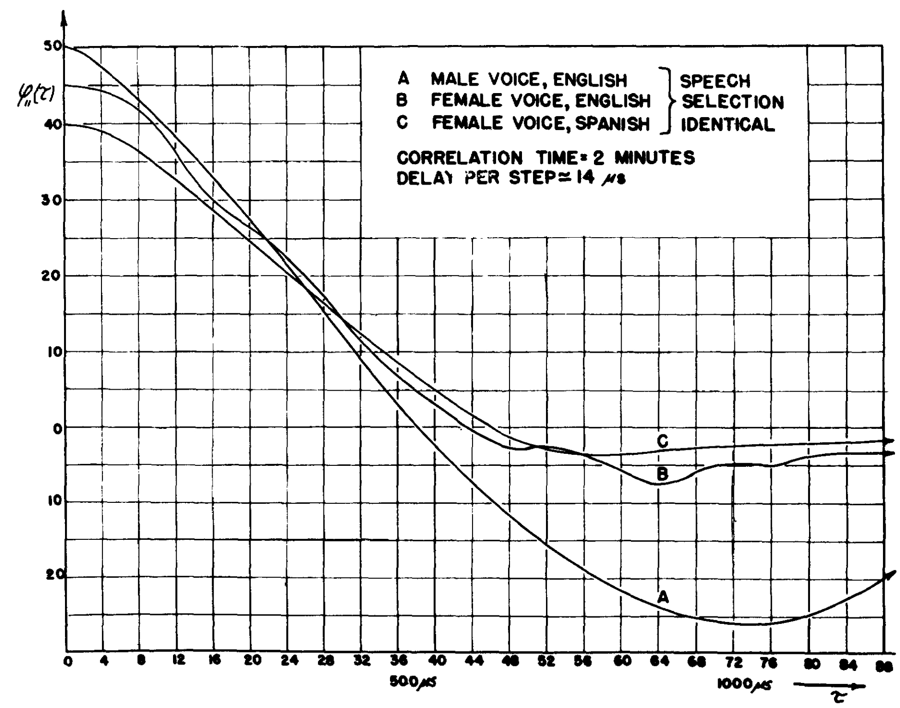


Fig. 61 Autocorrelation functions of typical speech.

*The human ear is not used here in the sense of it being a separate item by itself, but includes its integrated role in the functions of the entire human body.

speech characteristics (17). As the time of correlation is increased, the correlation functions of Fig. 61 should tend toward a limit of less difference, since we go from a one-one relation of correlation function to time series to a one-many relation. The correlation time of two minutes was selected on a rough estimate basis as being greater than the average significant learning time of the human ear.*

It should be noted that correlation functions have been discussed here with considerable freedom as to the original mathematical definition of Eq. 21, which defines a measure invariant with time. We have discussed here a statistical parameter which changes with time at a relatively slow rate with reference to the time function it measures. It is appropriate in this sense to speak of short-time correlation functions.

It is a reasonable estimate that an optimum Wiener filter for average speech designed on a basis of a time invariant approximation (long-time correlation) will offer little toward bridging the gap that lies between our present filter design and the properties attributed to the human ear. A more flexible filter is needed: perhaps one which is capable of adjusting itself to follow relatively slow changes in the statistical character of the speech wave. By defining a correlation function as a mean from $-\infty$ to $+\infty$, advantage is not taken of the difference in rates at which a message and disturbance approach statistical equilibrium. The possibility of a self-adjusting optimum filter based on measurements of short-time correlations is considered in the next section.

5.21 A Thinking Filter

It is appropriate to consider at this point the construction of a "thinking" filter as a possible analog to some portion of the human ear's complex operation. A feasible construction is that indicated in Fig. 62, where a short-time correlator, after correction by sources of a priori information, is used to control the shape of a variable optimum filter (constructed according to Wiener's least mean-square-error criterion). There are various techniques which might be used to obtain short-time correlations. The nature or means for compensation of the controlling function for a priori information will be largely determined by the nature of this information, but presumably can be accomplished. A more difficult engineering and theoretical problem is the construction of a filter whose shape can be arbitrarily changed as a function of the controlling signal from the computer. Analytical techniques for designing optimum filters are available directly from Wiener's theory (2, 11), but an electronic mechanization of the equivalent techniques has not been

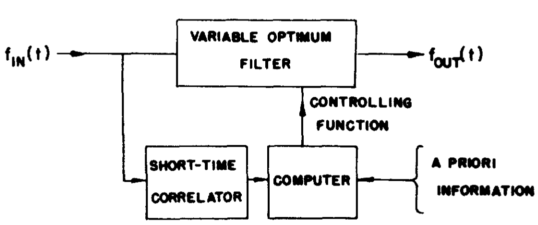


Fig 62 Block diagram of a thinking filter.

*Work on short-time correlation characteristics of speech is presently being undertaken at the Research Laboratory of Electronics, M.I.T., Acoustics Laboratory, M.I.T., and at Northeastern University.

accomplished to date. The computer and variable optimum filter taken together will represent an electronic solution of the Wiener-Hopf equation (2, 11).

5.22 Time-Domain Filters Using Arbitrary Delays

A possible engineering solution to the problem of constructing a variable filter lies in the time-domain synthesis of networks for a prescribed transient.

Some possible techniques for accomplishing this have grown out of joint discussion and experiment with G. D. Robertson of the Office of Naval Research, to determine the detection properties of a feedback amplifier having a delay τ in the feedback path, as indicated in Fig. 63.

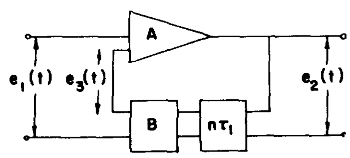


Fig. 63 Feedback amplifier with a delay τ in the feedback path.

As a result of an erroneous initial analysis* based on physical reasoning, it was at first thought that the circuit of Fig. 63 could be capable of a significant improvement in signal-to-noise ratio for a periodic signal (of frequency $f = \frac{1}{n\tau_1}$) located in broad-band noise. The general reasoning was that if the input noise were of a sufficiently broad bandwidth (so that the autocorrelation function of the noise was essentially zero at $\psi_{11}(n\tau_1)$), then the delay $n\tau_1$ would make the feedback noise signal independent of the incoming signal. At the input, the noise component terms would therefore add on a power basis, whereas the signal would be exactly in phase and hence add on an amplitude basis. If the networks A and B are assumed to be all-pass, then naive reasoning can lead to the conclusion that Fig. 63 represents a network which is broad-band for noise and narrow-band for the signal. This conclusion leads to an immediate inconsistency, in that properties are attached to a network of linear elements which give it an optimum position in a cascade of linear-system functions. The error of the reasoning is brought out in the following simple analysis of the circuit

$$e_2(t) = A\left[e_1(t) + e_3(t)\right] \tag{71}$$

$$e_3(t) = B\, e_2(t + \tau) \tag{72}$$

$$e_2(t) = A\left[e_1(t) + B\, e_2(t + \tau)\right] . \tag{73}$$

Taking the Laplace transform of Eq. 73, with $s = \sigma + jw$, we have

$$E_2(s) = A\left[E_1(s) + B\, E_2(s)e^{-n\tau s}\right] . \tag{74}$$

Solving for $E_2(s)$

$$E_2(s) = \frac{A\, E_1(s)}{1 - AB\, e^{-n\tau s}} \tag{75}$$

*Included here as an interesting mistake.

The system transfer function H(s) is therefore

$$H(s) = \frac{E_2(s)}{E_1(s)} = \frac{A}{1 - AB\, e^{-n\tau s}} . \tag{76}$$

Substituting s = jw, we find

$$H(jw) = \frac{A}{1 - AB\, e^{-jwn\tau}} . \tag{77}$$

The absolute magnitude $\left|H(jw)\right|$ of Eq. 71 is given by

$$\left|H(jw)\right| = \frac{A}{\sqrt{1 - 2AB \cos nw\tau + A^2B^2}} . \tag{78}$$

A plot of Eq. 78 is given in Fig. 64. The response of the network to a unit impulse, h(t), obtained from the inverse Laplace transform of Eq. 76, is plotted in Fig. 65. It should be clear from Figs. 64 and 65 that the circuit of Fig. 63 is a narrow-band circuit for any and all input time functions and its detection properties are those prescribed by Fig. 64.

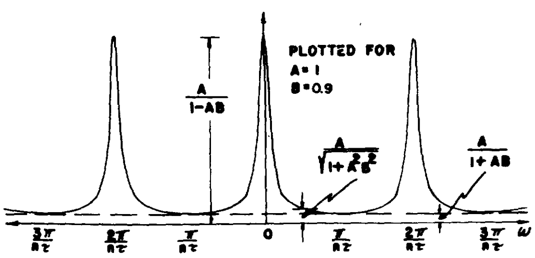


Fig. 64 Amplitude response of the network of Fig. 63.

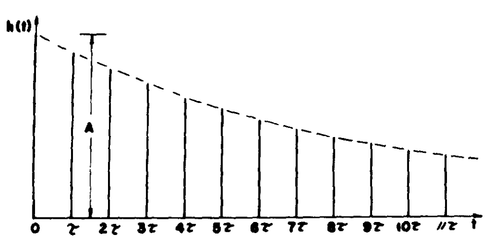


Fig. 65 Unit impulse response of network.

In verification of these results, autocorrelation and crosscorrelation functions were measured on a circuit corresponding approximately to the block diagram of Fig. 63. The actual circuit, designed by G. D. Robertson, is shown in Fig. 66. Two cascaded Ballantine meters were used as a noise source. The autocorrelation function of the noise source, shown in Fig. 67, is observed to be essentially zero at $\tau = 34$ μsec, the value of delay inserted in the feedback path. As a further check, the crosscorrelation between the input and the output of the B network was measured with the loop between A and B broken. The crosscorrelation was zero for all values of τ. With the loop closed, autocorrelation curves were measured at the output of the 20-40 kc band-pass filter (inserted to reject the harmonic passbands indicated in Fig. 64) for various values of feedback ratio AB. The curves measured were all very similar to Fig. 43 of Sec. 4.32, showing an essential single-tuned characteristic, as Fig. 64 would indicate.

The main importance of the above discussion lies, not in the circuit analysis of

Fig. 63, but rather in the extension of its basic concept to other more flexible networks which can be used as variable optimum filters. Two alternate but equivalent forms of time-domain filters are shown in Fig. 68. Their amplitude response is given by Eq. 79 and their response to a unit impulse is shown in Fig. 69.

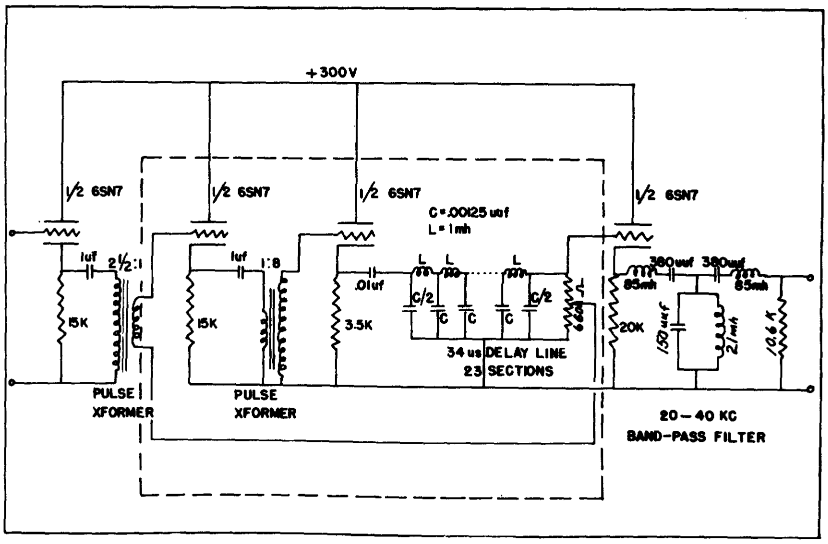


Fig. 66 Circuit diagram of feedback amplifier having a delay τ in the feedback path.

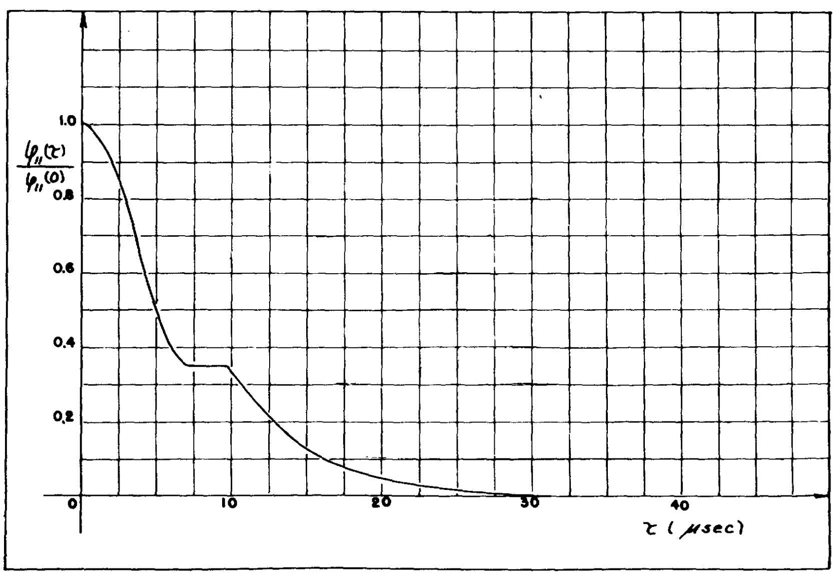


Fig. 67 Autocorrelation function of noise source (two cascaded Ballantine meters).

$$\left|H(jw)\right| = \sqrt{(1 + \cos w\tau + \cos 2w\tau + \ldots . \cos nw\tau)^2 + (\sin w\tau + \sin 2w\tau + \ldots . \sin nw\tau)^2} \, . \qquad (79)$$

A more practical and flexible development of Fig. 68 is shown in Fig. 70 where each τ delay point drives a linear amplifier with a unit impulse response, F(t), whose time constant and amplitude can be varied. In this fashion a prescribed transient response can be approximated with the greater flexibility of controls n, τ, a_n, and $F_n(t)$.

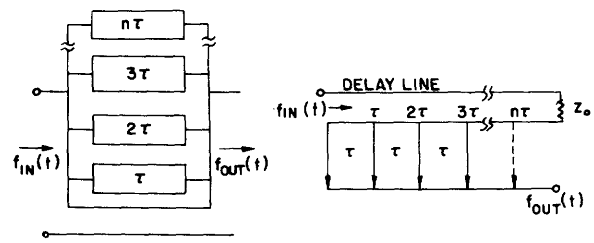


Fig. 68 Equivalent time-domain filters.

The conception of time-domain filtering is not new; and one of the early investigations in this field was made by Wiener and Lee (18). In general, however, synthesis procedures in the time domain have not made use of delays as a parameter. An important requirement in the construction of a variable optimum filter is that it be composed of a finite number of parts and a desirable feature is that the adjustment of these parts be as near independent as possible. The synthesis of h(t) in sections which do not overlap would accomplish this, and a condition which allowed the Kth adjustment to be independent of all previous (0 to K - 1) adjustments is a near condition of independence. The use of delays in the synthesis procedure is one means of obtaining this property.

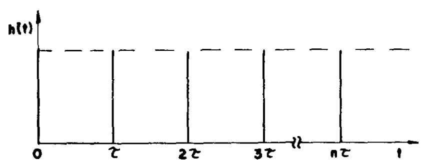


Fig. 69 Unit impulse response of networks in Fig. 66.

An example of a time-domain filter for use at high frequencies which allows for independent adjustment of portions of h(t) is shown in Fig. 71, where transit time* between the deflection plates of an electron gun is used as an equivalent delay parameter.

5.3 Autocorrelation Functions of Spoken Numbers

The autocorrelation functions of the spoken numbers, "zero", "one", "two", "three", "four", "five", "six", "seven", "eight", "nine" and the consecutive numbers to "one hundred and fifty" are shown in Fig. 72. Each of these short-time correlation curves was computed by the correlator on a relatively long-time basis by first recording the designating word or words of each curve on two and one-half minutes of magnetic tape. The tape was then played in a continuous loop**

*The use of an electron gun in this manner was derived from a similar transit-time problem originated by Prof. M. F. Gardner, M.I.T.

**Using a method designed at the Research Laboratory of Electronics, M.I.T., by C. W. Callahan and J. B. Angell.

on a Magnecord recorder and fed directly to the correlator for computation.

The first ten curves are short-time correlations of a single word and substantiate the general comments of Sec. 5.2. Considerable variation is indicated for the different numbers even for a maximum delay of only 1.25 msec. The last curve for numbers "zero" to "one hundred and fifty" was included for comparison purposes.

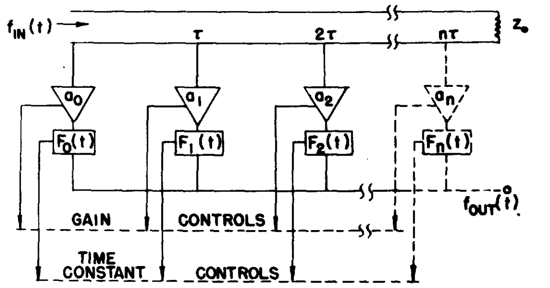


Fig. 70 A possible extension of Fig. 66.

A problem which has been receiving an increasing amount of general attention in research during the past two years is the possibility of constructing devices which will respond to spoken words (19). Two of the simpler and immediate applications that have been suggested are (1) a typewriter capable of responding to direct dictation and (2) a telephone circuit which would respond to a spoken number. To accomplish such engineering feats, a measure of individual words, which is invariant under conditions of change in accent, sex, age, etc. of the speaker, is necessary.

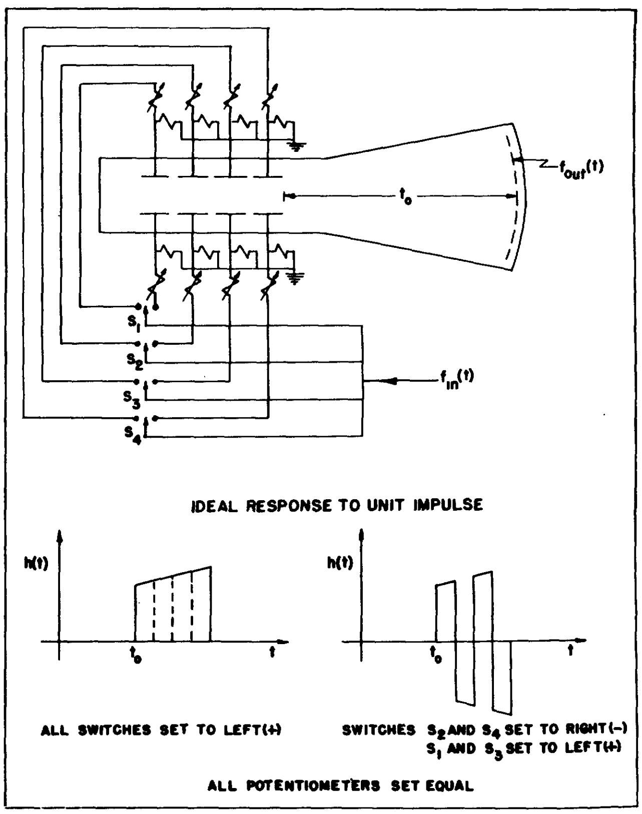


Fig. 71 Electron gun used as a variable time-domain filter.

The Bell Telephone Laboratories have reported (17) results applicable to this general problem. A plot of frequency versus time, obtained from a spectrograph, has been shown to be sufficient information for a trained person to differentiate words and syllables from a visual-scope presentation of the data. However, some difficulty is encountered in distinguishing between similar words of high-frequency content such as "nine" and "five". The analog in the time domain of the spectrograph is a short-time correlator. It was early thought that since the correlation function is more accurate at high frequencies (due to the infinite-local properties of the Fourier transform), the difference between similar high-frequency content words would show more clearly in their autocorrelation curves. This prediction is seen to be borne out in Fig. 72 where the "nine" and the "five" are practically identical except for the presence and absence of a peak at $\approx$300 μsec in the "nine" and "five" respectively. It would appear that a combination spectrograph and short-time correlator would increase the practical accuracy which this method of measure may have for distinguishing words.

An important additional amount of data could be obtained through the use of correlation functions in this general study. Crosscorrelation functions could well be used to determine the differences introduced in a spectra measure of a word as a result of difference in accent, sex, age, or other variations relevant to the character of a word.

5.4 Correlation Functions in the Study of Random Noise through Linear Passive Filters

Detailed quantitative measurements of amplitude probability density distributions, zero crossings, correlation functions and power density spectra of various noise sources through broad- and narrow-band linear filters have been made and described by N. Knudtzon (15). The correlation curves presented in this report were measured, using the electronic correlator described in Sec. IV. Figures 73 and 74 have been reproduced from Mr. Knudtzon's report and presented here as illustrations of the use of the correlator in the study of filtered random noise. The following explanations of Figs. 73 and 74 are taken from Mr. Knudtzon's report. Figure 73 shows the various transfer characteristics through which the noise sources were measured. In addition to these transfer characteristics, identified in the last column by a code number, a single tuned variable Q-circuit centered at 20.4 kc was provided, which could be inserted if desired.

In Fig. 74 the left-hand columns show autocorrelation curves as measured by the correlator for two different ranges of maximum delay. These curves were computed for the Si-crystal and 884 gas-tube noise sources for various filter transfer characteristics, as indicated in the last column and identified by reference to Fig. 73.

It was pointed out in Sec. III that the Fourier transform of the autocorrelation function will give the power density spectrum $\Phi_{11}(w)$ of a time series. The process of fitting an analytical curve to an experimental autocorrelation curve and the evaluation of its Fourier transform is a task which has only to be attempted once in order to demonstrate a convincing argument for the expenditure of time and effort on the mechanization of the process. Fortunately, an electronic differential analyzer capable of performing Fourier transformations has been developed at the Research Laboratory of Electronics, M.I.T., by Dr. A. B. Macnee (20). The differential analyzer was used by Mr. Knudtzon to transform some of the autocorrelation curves of Fig. 74. The curves to be transformed were first plotted in linear coordinates and reproduced photographically as masks for the function generator in the differential analyzer (20). Pictures of the corresponding traces on the scope are shown in Fig. 74 together with the power spectra obtained.

A detailed analysis and interpretation of these curves is given by Mr. Knudtzon in

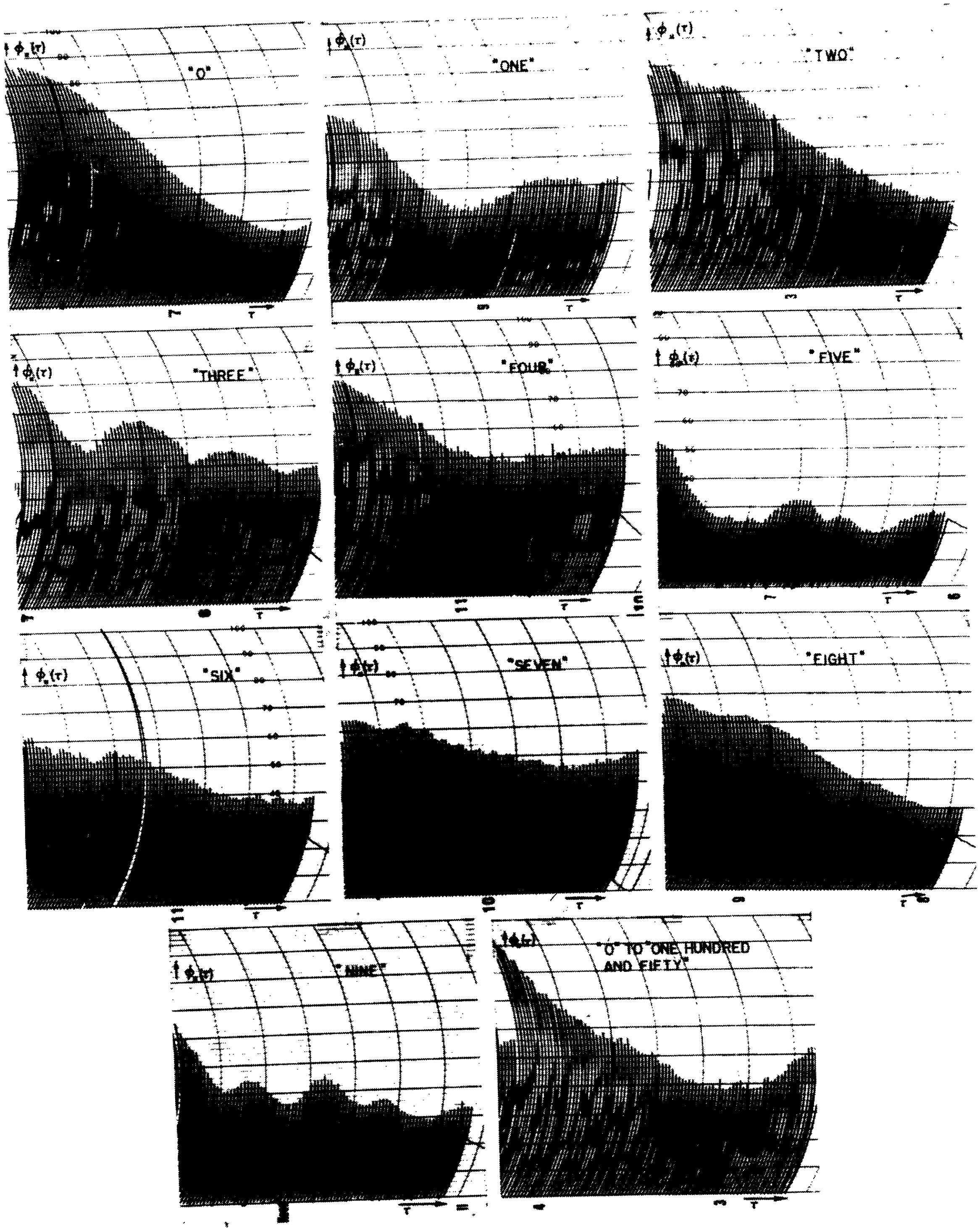


Fig. 72 Experimental autocorrelation functions for spoken numbers. Range of delay τ 0 to 1.25 msec. Delay per step $\sim$ 14μsec. Correlation time = 2 minutes.

TABLE I-2: TRANSFER CHARACTERISTICS

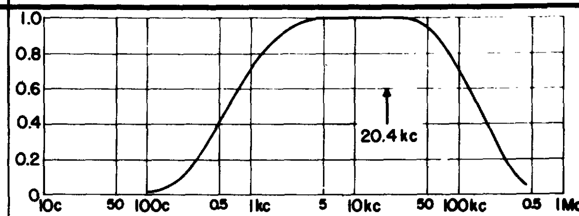
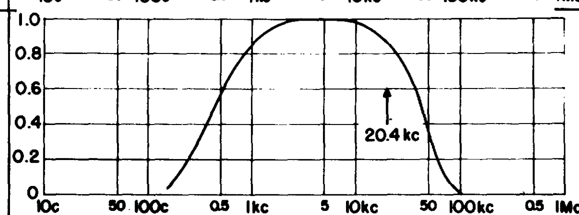
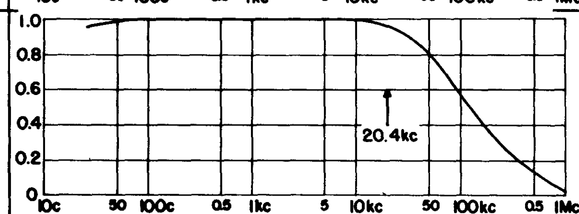
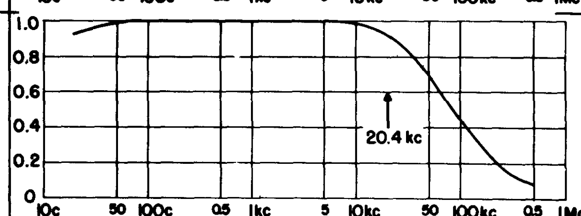
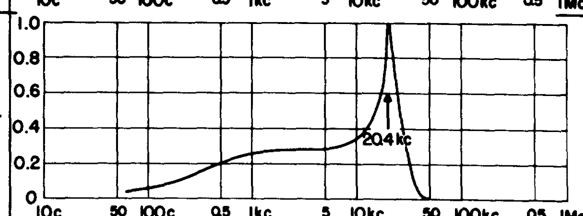
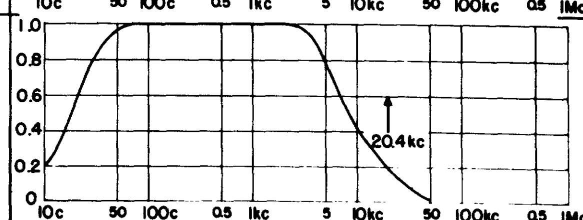
	EQUIPMENT	NORMALIZED CURVES	CODE NUMBER
	AMPLIFIER WITHOUT BUTTERWORTH BAND-PASS FILTER (FIG. I-8)		1
	AMPLIFIER WITH BUTTERWORTH BAND-PASS FILTER (FIG. I-8)		2
PREAMPLIFIERS	BALLANTINE AMPLIFIER MODEL 300 (TABLE I-1)		3
PREAMPLIFIERS	BALLANTINE DECADE AMPLIFIER MODEL 220 (TABLE I-1)		4
PREAMPLIFIERS	TRANSFORMER x BALLANTINE DECADE AMPL. x BALLANTINE AMPLIFIER (TABLE I-1)		5
NOISE SOURCE	G.R. AMPLIFIER TYPE 1231-B (TABLE I-1)		6

Fig. 73 Transfer characteristics used in a study of filtered random noise. (Reprinted from Technical Report No. 115, Research Laboratory of Electronics, M.I.T. by permission of Mr. N. Knudtzon.)

his report, to which the reader is referred for further details. It is important here to mention that the results obtained from the autocorrelation curves were important in explaining certain peculiar properties betrayed by the Si-crystal and 884 gas-tube noise sources in the other statistical measurements made, and served to complete an integrated picture of the general and specific statistical properties of various noise sources.

5.5 Correlation Functions of Random Noise through Nonlinear Devices

As pointed out in Sec. 3.3, the experimental measurement of correlation functions provides a possible strong initial method of attack on the general problem of nonlinear systems. As a qualitative example of this method of analysis, Fig. 75 shows

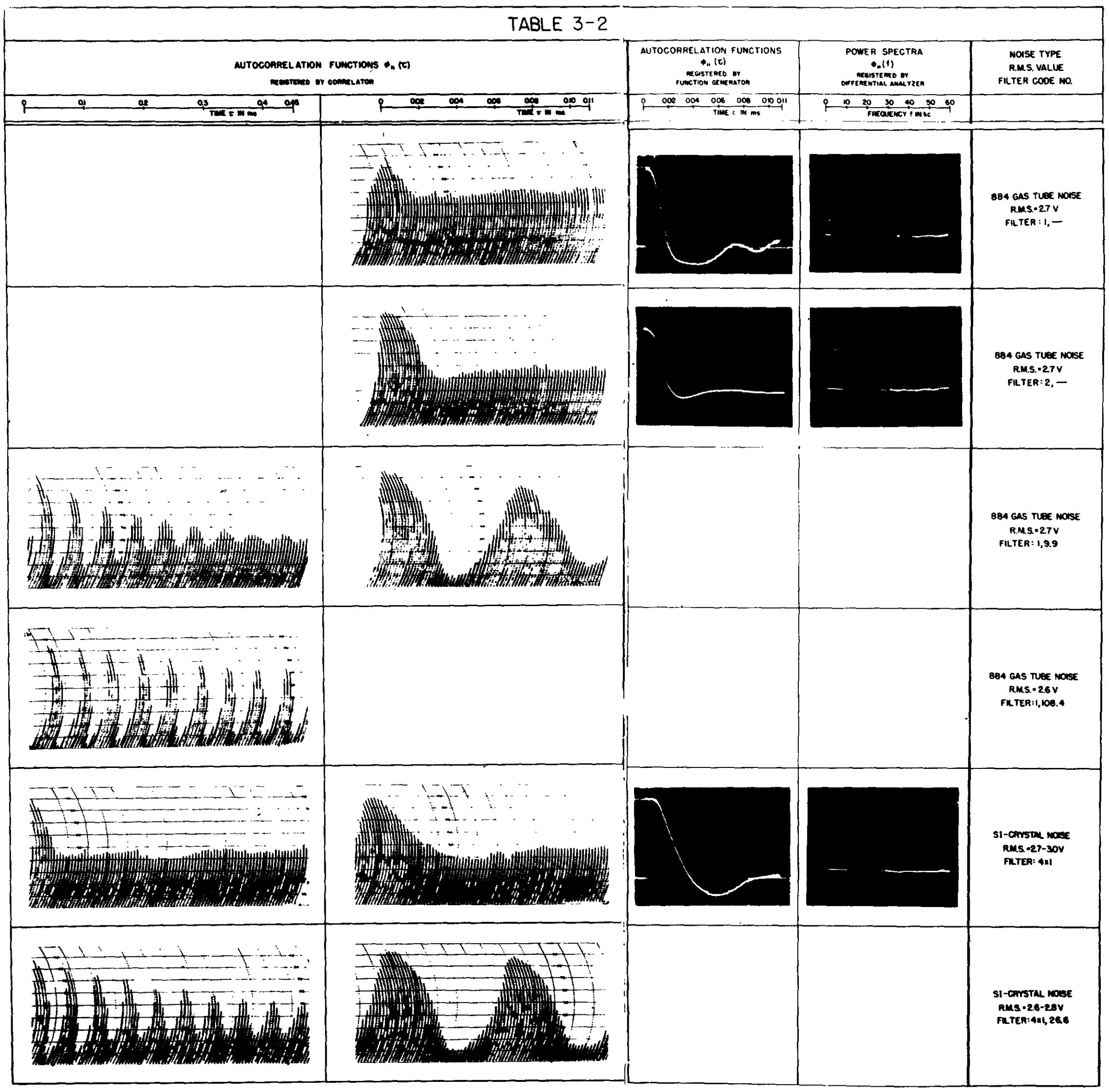


Fig. 74 Autocorrelation functions and power sensity spectra for filtered random noise. (Reprinted from Technical Report No. 115, Research Laboratory of Electronics, M.I.T. by permission of Mr. N. Knudtzon.)

autocorrelation functions and oscillograms of noise through linear and nonlinear circuits. The noise source used was two cascaded Ballantine meters having a relatively broad bandwidth as indicated by their autocorrelation curve, shown in Fig. 67, Sec. 5.22. The noise was passed through a single tuned circuit having a Q of about 15 and then limited by two front-to-back biased 1N34 crystals for the case of symmetrical limiting and by a single biased 1N34 crystal for the case of asymmetrical limiting. Photographs of the input time functions and the multiplying pulses to the integrator are also shown in Fig. 75 for qualitative comparison. The autocorrelation curves are in general qualitative agreement with the theoretical results derived by Middleton (21). The autocorrelation function of broad-band noise through a narrow-band single tuned circuit is a damped cosine wave, whose frequency is very nearly that of the center frequency of the tuned circuit and whose damping is proportional to the bandwidth of the tuned circuit (11, 22). This general character is shown by the first autocorrelation

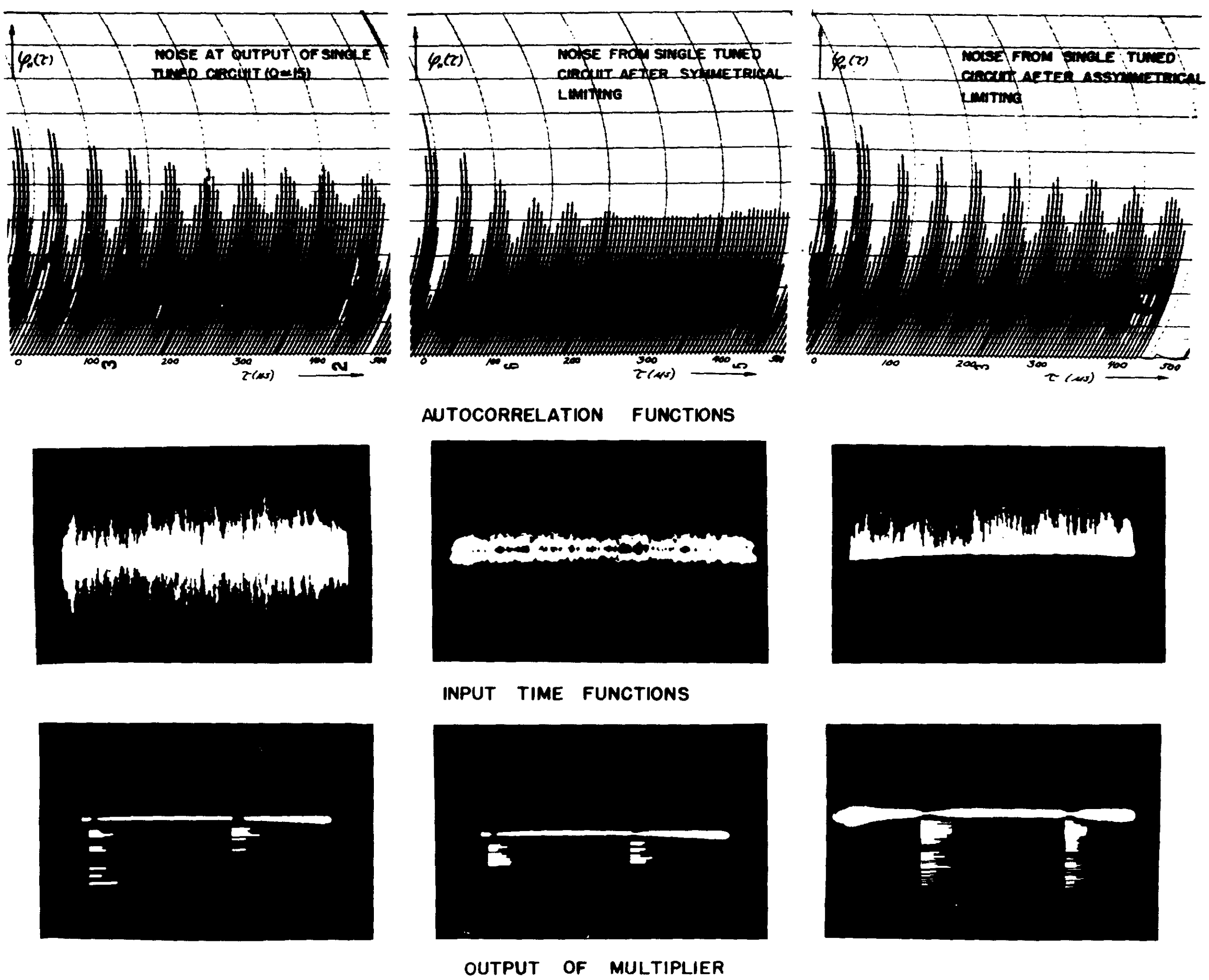


Fig. 75 Autocorrelation functions and oscillograms of noise through linear and nonlinear circuits.

curve of Fig. 75. Middleton's theoretical result that clipping always spreads the spectrum is illustrated in the second curve by the much increased damping factor. The generation of smaller harmonic noise bands is not an obvious result to be seen directly by inspection of the autocorrelation curve, but a Fourier transformation of the curve would also place this result of Middleton's in evidence. In the third curve, the spectrum spread is noted to be less than for the second curve, but the asymmetrical shape of the autocorrelation curve shows the presence of a d-c and low-frequency component in the resulting spectrum, as predicted by Middleton.

An immediate practical application of this technique of analysis and study in the field of communication is in the study of the operation of the various types of limiters (and discriminators) used in f-m receivers in the presence of thermal and impulse noise, both with and without a signal present.

5.6 Correlation Functions of Heartbeats for Use in the Detection of a Murmur

There has been a steadily increasing interest on the part of the medical profession in the possible applications of the electronic correlator, not only in the classification and analysis of channels and regions of information flow, as mentioned in Sec. 3.21, but also in its use as an early warning detector for the presence of malfunctioning of particular organs of the body.

As a result of various discussions and conferences with several heart specialists, principally Dr. R. Streiper, House of Good Samaritan Hospital, Boston, Mass., the writer has come to appreciate the difficulty involved in the early detection and classification of a heart murmur. In general, today's technique relies very heavily on a doctor's stethoscope and his years of experience. There is very little evidence of an objective criterion of measure. A reliable economic means of screening and examining a large number of people quickly, as desired by the Army Medical or Public Health Service, is not available. In addition, most methods that are actually used to test the reserve strength and endurance of a suspected weakened heart require placing the suspected heart under a moderate strain, which has often proved fatal due to an inaccurate initial estimate.

A heart murmur is, in general, caused by a faulty operation of the heart valves. It was reasonable to expect that this faulty operation would set up a transient sound that, if detectable, could be used as a characteristic measure. The detection properties of the correlator discussed in the first part of this section afford a practical method of testing this hypothesis.

The experiments described here were of an exploratory and optimistic nature. First, the maximum delay of the correlator without the modifications of Sec. IV, was only of the order of 2.5 msec, whereas the basic heart beat is roughly one per second. For the autocorrelation function to show anything significant, the transient would have to be in the range of about 400 cps. The experiments and measurements were performed, however, with the viewpoint that even with a negative result an upper bound, of which later experiments could take advantage, would be established.

As a first experiment, the autocorrelation function of an EKG signal of a normal heart (left wrist to right leg) was computed. The recording of the signal was made late at night in a noise cage with considerable care taken to reduce all hum and pickup as much as possible. The recorded signal was three minutes and forty seconds in length and the correlation time set on the correlator was two minutes. The autocorrelation curve, as measured and replotted on linear coordinates, is shown in Fig. 76. The

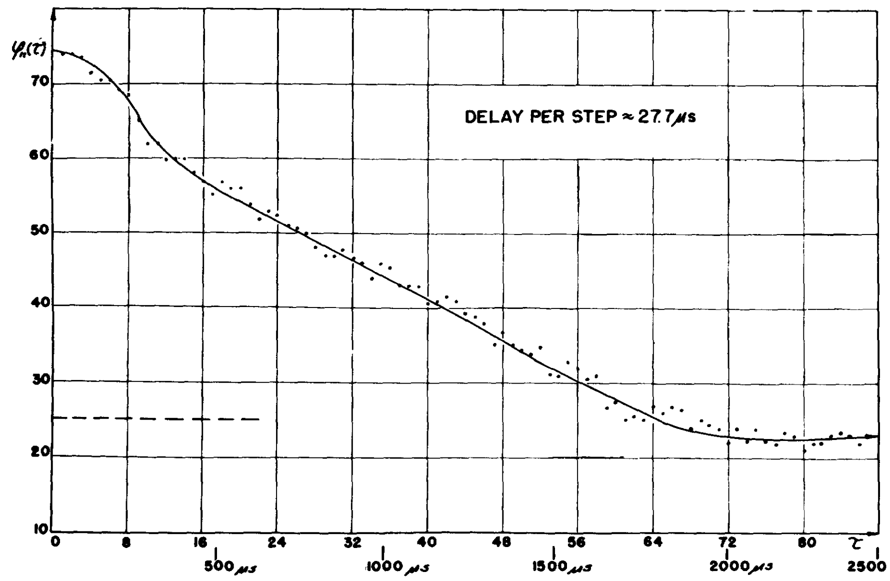


Fig. 76 Autocorrelation function EKG normal heart (left wrist to right leg).

shape of this curve was surprising, since it shows the presence of random-frequency components up to possibly 1 kc. This curve gave evidence of high-frequency components in the EKG signal, but it did not prove that they came from the heart, and the possibility of the measured result being caused by muscle reactions or emissions from other organs could not be ruled out.*

A second measurement was then taken in an anti-echo chamber, using a very sensitive microphone to pick up the direct heart sounds. This time the subject was a patient having a known and severe heart condition. According to Dr. R. Streiper, the pertinent medical facts were that the patient had suffered from rheumatic heart fever which later developed into a so-called Grade 3 heart murmur, classified as aortic regurgitation with a mitral involvement. This type of murmur is of relatively high pitch and therefore

*Professor Wiener has suggested that the crosscorrelation of signals from two points be computed in order to reduce the effect of local and uncorrelated signals. Professor J. B. Wiesner has initiated the construction of a twin-track and variable speed recording unit for specific use in computing the crosscorrelation of such signals. This equipment was not available, however, at the time of this experiment.

well suited for the experiment. The appearance of the heart sounds on an oscilloscope was much quieter between heart beats than for the EKG signal. There was, however, a distinct transient noticeable at the end of each heartbeat, although an eye estimate did not place it above 200 cps. The autocorrelation curve shown in Fig. 77 is in agreement with these observations and shows that the interesting content of Fig. 76 is not a result of the heart.

The maximum delay of Fig. 77 is 2 msec. It is recommended that further experiments in this direction be with the computation of both autocorrelation and crosscorrelation functions carried out to approximately 10 msec.

5.7 Crosscorrelation as a Means of Determining the Prediction or Delay Time between Two Time Series

The importance of this property of crosscorrelation functions and the ability of the electronic correlator to compute these functions have already been described. As an illustration of the experimental applications of the electronic correlator in such measurements, the reader is referred to Technical Report No. 129, Research Laboratory of Electronics, M.I.T., "Statistical Prediction of Noise", by Y. W. Lee and C. A. Stutt. This report describes the synthesis, design and construction of an optimum predictor for filtered noise using Wiener's least mean-square-error criterion. Experimental results are shown in the form of a simultaneous display on film of the input and output waveforms of the predictor circuit, and in measured crosscorrelation functions between the input and output. Some twelve crosscorrelation functions are shown as a function

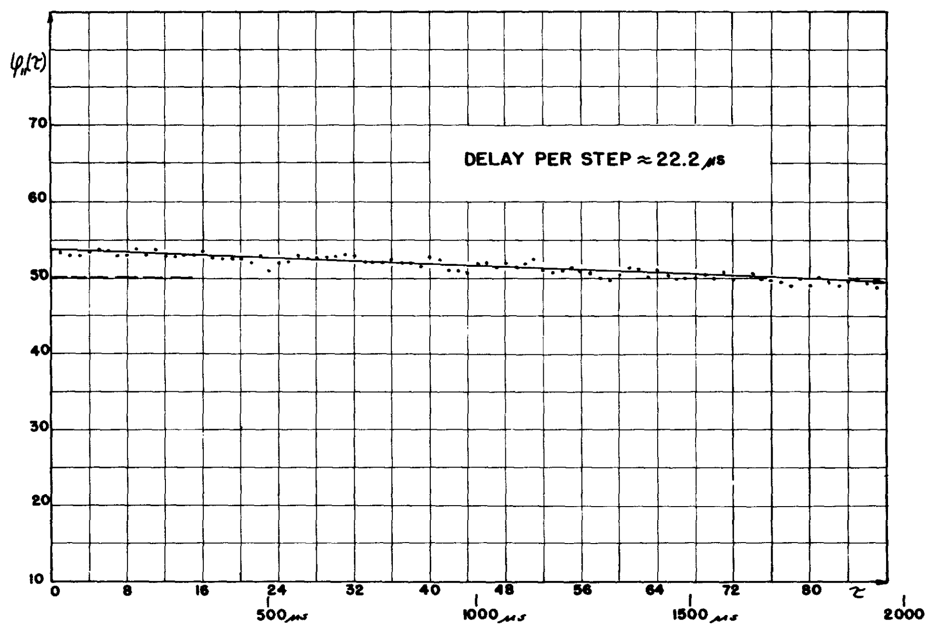


Fig. 77 Autocorrelation function of heart sounds recorded with sensitive microphone in anti-echo chamber. Patient known to have Grade 3 murmur.

of both the Q of the filtered noise and the designed prediction values. Prediction times, as recorded by the correlator in these curves, together with averages computed from direct readings of the film, are compared with the design values.

Appendix I

The expectation of -log p(x) is a maximum if the variable x is distributed at random, i.e. if p(x) = constant. The following proof parallels that given by Fry (8, p. 201). We want to make

$$E(-\log p(x)) = -\int_{-\infty}^{\infty} p(x) \log p(x) dx = \text{a maximum} \tag{1}$$

subject to the constraint that

$$\int_{-\infty}^{\infty} p(x) dx = 1 \ . \tag{2}$$

The desired curves satisfy the Euler equation for minimizing

$$\int \left[F(y) - \lambda G(y) \right] dy, \text{ where } \lambda = \text{a constant} \ . \tag{3}$$

The necessary condition is

$$\frac{\partial F}{\partial y} - \lambda \frac{\partial G}{\partial y} - \frac{d}{dx}\left[\frac{\partial F}{\partial y'}\right] + \lambda \frac{d}{dx}\left[\frac{\partial G}{\partial y'}\right] = 0 \tag{4}$$

where

$$F = y \log y \qquad \frac{\partial F}{\partial y} = 1 + \log y \qquad \frac{\partial F}{\partial y'} = 0$$

$$G = y \qquad \frac{\partial G}{\partial y} = 1 \qquad \frac{\partial G}{\partial y'} = 0 \ .$$

By substitution in Eq. 4, we have

$$1 + \log y - \lambda = 0$$

or

$$\log y = \lambda - 1 \ . \tag{5}$$

Since log y is a constant, then y is a constant also. This proves that log y = constant is an extremal. To show that it defines a maximum, let the solution to the problem be p(x) = P. Since this function makes the E(-log p(x)) a maximum, it follows that any change made in p(x) must decrease E(-log p(x)).

For instance, a substitution of $p(x) = P + \delta(x)$ should decrease E(-log p(x)).

$$-\int_{-\infty}^{\infty} \left[P + \delta(x)\right] \left[\log\left[P + \delta(x)\right] - \lambda\right] dx \ . \tag{6}$$

Equation 6 can be analyzed by expanding the term $\log\left[P + \delta(x)\right]$ in a series

$$\log\left[P + \delta(x)\right] = \log P + \frac{\delta(x)}{P} - \frac{\delta^2(x)}{2!\,P^2} + \frac{2\delta^3(x)}{3!\ P^3} - \frac{6\delta^4(x)}{4!\ P^4} \ldots \qquad (7)$$

Then

$$- \left[P + \delta(x)\right] \left[\log(P + \delta(x)) - \lambda\right] = - P(\log P - \lambda) - \delta(x) + \frac{\delta^2(x)}{2P} - \frac{\delta^3(x)}{3P^2} + \frac{\delta^4(x)}{4P^3} - \ldots - \delta(x)\,(\log P - \lambda) - \frac{\delta^2(x)}{P} + \frac{\delta^3(x)}{2P^2} - \frac{\delta^4(x)}{3P^3} + \ldots \qquad (8)$$

Since $\delta(x)$ is very small, we may neglect terms in $\delta(x)$ higher than 2nd order and write as an approximate expression for Eq. 6

$$- \int_{-\infty}^{\infty} P(\log P - \lambda)dx - \int_{-\infty}^{\infty} (\log P + 1 - \lambda)\ \delta(x)dx - \frac{1}{2} \int_{-\infty}^{\infty} \frac{\delta^2(x)}{P}\, dx + \ldots \qquad (9)$$

The first term is a maximum by hypothesis and the third and all succeeding terms can only decrease the sum. However, unless the second term is zero, a proper choice of $\delta(x)$ can make the entire result larger than the first, which is contradictory. Hence the second term must be zero and we arrive at the same condition of Eq. 5, i.e. $\log y = \lambda - 1$ = constant, but with the additional specification that it defines a maximum.

Appendix II

Probability Density Distribution for the Product of Two Normally Distributed Variables

Let X and Y be two normally distributed variables, having a standard deviation σ_1 and σ_2 respectively.

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{x^2}{2\sigma_1^2}} \qquad (1)$$

$$p(y) = \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{y^2}{2\sigma_2^2}} .$$

We desire $p(z)$, where $z = xy$. Since x and y are independent, the probability of x, y is

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_1^2} + \frac{y^2}{\sigma_2^2}\right)} . \tag{2}$$

To transform the above distribution function into a function of the defined variable z, an additional variable of integration, u, may be defined as follows

$$z = xy$$
$$u = \frac{x}{y} . \tag{3}$$

The Jacobian of this transformation becomes

$$\frac{\partial(z, u)}{\partial(x, y)} = \begin{vmatrix} y & x \\ \frac{1}{y} & \frac{-x}{y^2} \end{vmatrix} \begin{matrix} = -2\frac{x}{y} = -2u \\ = +2u \end{matrix}$$ (since the negative sign has no significance).

The following relation holds between the old and new variables of the distribution function

$$p(z, u) = \frac{p(x, y)}{\frac{\partial(z, u)}{\partial(x, y)}} \tag{4}$$

and therefore

$$p(z, u') = \frac{1}{2\pi\sigma_1\sigma_2} \frac{e^{-\frac{1}{2\sigma_1\sigma_2}(zu' + \frac{z}{u'})}}{2u'} , \text{ where } u' = \frac{\sigma_2}{\sigma_1} u .$$

To determine p(z), we have only to integrate p(z, u) over all possible values of u. Since $u' = x/y \ \sigma_2/\sigma_1^2$, its limits are $-\infty$ to $+\infty$.

$$p(z) = \frac{1}{4\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2\sigma_1\sigma_2}(u' + \frac{1}{u'})}}{u'} du' . \tag{5}$$

The signs of u' and z are always the same. Therefore, if we restrict u' to run from 0 to ∞, we restrict z also to run from 0 to ∞; since the function p(z) is symmetrical, we may rewrite it as

$$p(z) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{0}^{\infty} \frac{e^{-\frac{z}{2\sigma_1\sigma_2}(u' + \frac{1}{u'})}}{u'} du', \quad z > 0 . \tag{6}$$

Now let $u' = e^y$ limits $u' = 0$ $y = -\infty$

$du' = e^y dy = u' dy$ $u' = \infty$ $y = \infty$.

Then

$$u' + \frac{1}{u'} = e^{y} + e^{-y} = 2 \cosh y$$

$$p(z) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{z}{\sigma_1\sigma_2} \cosh y} \, dy \ , \quad z > 0 \ . \tag{7}$$

The above integral was recognized by Dr. M. Cerrillo as being in Watson: Bessel Functions, Ch. VI, pp. 181-183, Macmillan Company, 1945, where it is found that

$$K_{\nu}(z) = \frac{1}{2} \int_{-\infty}^{\infty} e^{-z \cosh t - \nu t} \, dt \ . \tag{8}$$

Therefore

$$K_o\left(\frac{z}{\sigma_1\sigma_2}\right) = \frac{1}{2} \int_{-\infty}^{\infty} e^{-\frac{z}{\sigma_1\sigma_2} \cosh y} \, dy \tag{9}$$

and the solution to Eq. 6 is seen to be

$$p(z) = \frac{1}{\pi\sigma_1\sigma_2} K_o\left(\frac{z}{\sigma_1\sigma_2}\right) . \tag{10}$$

If the variables x and y are from the same source, but sufficiently separated in time so that independence holds, then σ_1 can be taken as equal to σ_2, and p(z) can be written

$$p(z) = \frac{1}{\pi\sigma^2} K_o\left(\frac{z}{\sigma^2}\right) . \tag{11}$$

All moments of p(z) can be evaluated from the following integral obtained by Dr. M. Cerrillo

$$\int_0^{\infty} K_{\nu}(x)\, x^{u-1} dx = 2^{u-2} \Gamma\left(\frac{u+v}{2}\right) \Gamma\left(\frac{u-v}{2}\right) \tag{12}$$

and, as a particular example, the second moment M_2 is found to be

$$M_2 = \frac{2}{\pi\sigma^2} \int_0^{\infty} z^2 K_o\left(\frac{z}{\sigma^2}\right) dz$$

$$M_2 = \frac{2\sigma^4}{\pi} \int_0^{\infty} x^2 K_o(x) dx = \frac{2(2\sigma^4)}{\pi} \Gamma\frac{3}{2} \Gamma\frac{3}{2}$$

$$M_2 = \sigma^4$$

Acknowledgment

The writer wishes to express his gratitude to Professor Y. W. Lee for supervising this research and to Professor N. Wiener and Professor J. B. Wiesner for their suggestions, comments and encouragement. The writer is particularly indebted to Dr. Keith Boyer, Dr. E. R. Kretzmer, Mr. C. A. Stutt, Mr. Walter Pitts, Dr. M. Cerrillo, Prof. H. J. Zimmermann, Mr. N. Knudtzon, Dr. D. Middleton, Mr. G. Robertson, Mr. H. E. Singleton and Mr. L. G. Kraft for their generous time spent in many stimulating discussions relevant to the general topic of statistical communications.

References

1. E. Nagel: Principles of the Theory of Probability (International Encyclopedia of Unified Science), 1, 6, University of Chicago Press, 1947
2. N. Wiener: The Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications, John Wiley, New York, 1949
3. N. Wiener: Cybernetics, or Control and Communications in the Animal and the Machine, John Wiley, New York, 1949
4. C. E. Shannon: B.S.T.J. 27, 379-423, July, 1948; ibid. 623-656, October, 1948
5. W. G. Tuller: Doctoral Thesis, Electrical Engineering, M.I.T. 1948
6. N. Wiener: Acta Mathematica 55, 117-125, 1930
7. R. M. Fano: The Transmission of Information, Technical Report No. 65, Research Laboratory of Electronics, M.I.T. March, 1949
8. T. C. Fry: Probability and Its Engineering Uses, Van Nostrand, New York, 1928
9. G. Birkhoff, S. MacLane: A Survey of Modern Algebra, Macmillan, New York, 1948
10. J. V. Uspensky: Introduction to Mathematical Probability, McGraw-Hill, New York, 1937
11. Y. W. Lee: Theory of Optimum Linear Systems, Class Notes, Course 6.563, M.I.T. 1948 (unpublished)
12. G. I. Taylor: Proc. London Math. Soc. 196-212, 1920
13. E. R. Kretzmer: Interference Characteristics of Pulse-Time Modulation, Technical Report No. 92, Research Laboratory of Electronics, M.I.T. May, 1949
14. J. M. Ham: Master's Thesis, Electrical Engineering, M.I.T. 1947
15. N. Knudtzon: Experimental Study of Statistical Characteristics of Filtered Random Noise, Technical Report No. 115, Research Laboratory of Electronics, M.I.T. July, 1949
16. S. O. Rice: B.S.T.J. 27, 109-157, January, 1948
17. R. K. Potter, G. A. Kopp, H. C. Green: Visible Speech, Van Nostrand, New York, 1947
18. N. Wiener, Y. W. Lee: U.S. Patent 2,128,257; granted 1931
19. Northeastern University: Visual Message Presentation, Progress Report No. 1, Contract No. AF 19(122)-7, July, 1949; Progress Report No. 4 and 5, Contract No. W28-o99-ac-386, January, 1949 and March, 1949
20. A. B. Macnee: An Electronic Differential Analyser, Technical Report No. 90, Research Laboratory of Electronics, M.I.T. December, 1948

21. D. Middleton: J. App. Phys. 17, 10, 778-801, October, 1946

22. Y. W. Lee, C. A. Stutt: Statistical Prediction of Noise, Technical Report No. 129, Research Laboratory of Electronics, M.I.T. July, 1949

23. A. M. Turing: Proc. London Math. Soc. ser. 2, 42, 230-265, 1936-1937

24. M. Cerrillo: Technical Report No. 55, Research Laboratory of Electronics, M.I.T. (to be published)