X. STATISTICAL COMMUNICATION THEORY

Prof. Y. W. Lee	K. L. Jordan	S. D. Pezaris
A. G. Bose	C. S. Lorens	K. H. Powers
J. Y. Hayase	S. G. Margolis	R. E. Wernikoff

A. NOISE FIGURES FROM THE VIEWPOINT OF POWER TRANSMISSION EFFICIENCY

A method has been developed by which the noise figure F at the output of a general bilateral network is calculated through the use of reciprocity relations. It is found that 1/F is equal to the efficiency of weighted power transmission from output terminals to source. Weighted power is defined as the power dissipated in each resistor multiplied by the relative noise temperature of the resistor. In addition to being a useful calculating tool the method gives a better insight into the problem of minimizing the noise figure of a network.

Consider, at a frequency f, a linear bilateral network N. Let the noise in any resistor R_i of the network be represented, for a unit bandwidth around the frequency f, by a current source of the same frequency and magnitude $(4kT\rho_i/R_i)^{1/2}$ in shunt with the resistor R_i, and let there be no correlation between the noises generated in any two different resistors of the network. Further, let $E_1, E_2, \ldots, E_n$ be the open-circuit voltages at n terminal pairs 11', 22', ..., nn' of the network, and let $x = a_1E_1 + a_2E_2 + \ldots + a_nE_n$, where $a_1, a_2, \ldots, a_n$ are complex constants, be considered as the "output" of the network. Note that the output x is not necessarily a voltage or current actually appearing in the network.

Then we can prove

THEOREM 1: The mean square output $|x|^2$ per unit bandwidth produced by the noise source of any resistor R_i in the network N is equal to $4kT\rho_i P_i$, where k is Boltzmann's constant, T is the resistor temperature in degrees Kelvin, ρ_i is the relative noise temperature of resistor R_i, and P_i is the power dissipated in resistor R_i when the network is driven by current sources $I_1 = a_1, I_2 = a_2, \ldots, I_n = a_n$ at terminal pairs 11', 22', ..., nn', and all other sources of the network are set equal to zero.

PROOF: Let $z_{ik} = z_{ki}$, k = 1, 2, ..., n be the transfer impedances between the terminals of resistor R_i and terminal pairs 11', 22', ..., nn'. Then the current source $(4kT\rho_i/R_i)^{1/2}$ across R_i will produce an output x given by

$$x = a_1 z_{i1} (4kT\rho_i/R_i)^{1/2} + a_2 z_{i2} (4kT\rho_i/R_i)^{1/2} + \ldots + a_n z_{in} (4kT\rho_i/R_i)^{1/2}$$

$$= (a_1 z_{i1} + a_2 z_{i2} + \ldots + a_n z_{in}) (4kT\rho_i/R_i)^{1/2} \qquad (1)$$

But $a_1 z_{i1} + a_2 z_{i2} + \ldots + a_n z_{in}$ is the voltage E_{R_i} across resistor R_i when the network is driven by the current sources $I_1 = a_1, I_2 = a_2, \ldots, I_n = a_n$. Hence

$$x = E_{R_i} (4kT\rho_i/R_i)^{1/2} \tag{2}$$

that is,

$$|x|^2 = \left|E_{R_i}\right|^2 (4kT\rho_i/R_i) = 4kT\rho_i \frac{\left|E_{R_i}\right|^2}{R_i} = 4kT\rho_i P_i \qquad \text{Q.E.D.} \tag{3}$$

Define now as "weighted power" dissipated in a resistor R_i the actual power dissipated in R_i times the relative noise temperature ρ_i of the resistor. Also define as weighted power dissipated in a network the sum of the weighted powers dissipated in the resistors of the network. Then we can prove

THEOREM 2: The single frequency noise figure F of the output $x = a_1 E_1 + \ldots + a_n E_n$ of network N is equal to $1/\eta$, where η is the efficiency of transmission of weighted power from the sources $I_1 = a_1, \ldots, I_n = a_n$ to the signal source resistance.

PROOF: Let R_s be the signal source resistance, R_i $(i = 1, \ldots, m)$ the remaining resistances in the network. Then

$$F = \frac{\text{mean square output resulting from all noise sources}}{\text{mean square output resulting from signal noise source}} \tag{4}$$

Since the noise sources in different resistors are uncorrelated to each other the mean square output resulting from all noise sources can be found by summing over the individual contributions of these sources. Applying Theorem 1, we get

$$F = \frac{4kT\rho_s P_s + \sum_{i=1}^{m} 4kT\rho_i P_i}{4kT\rho_s P_s} = \frac{\rho_s P_s + \sum_{i=1}^{m} \rho_i P_i}{\rho_s P_s} \tag{5}$$

In the transmission of weighted power from the sources $I_1 = a_1, \ldots, I_n = a_n$ to R_s, the weighted power $\rho_s P_s$ dissipated in R_s has obviously to be considered as output; the weighted powers $\rho_i P_i$ dissipated in the R_i's as losses. Hence

$$F = \frac{\text{output + losses}}{\text{output}} = \frac{1}{\eta} \qquad \text{Q.E.D.} \tag{6}$$

When n = 1 the output is the voltage appearing across a single terminal pair 11'. Then, quite obviously, the efficiency of transmission from output terminals to signal source is the same, irrespective of whether we drive the output with a voltage or a current source. For this special case Theorem 2 can be restated as

THEOREM 2a: The noise figure of a network N having a single terminal pair output is equal to the reciprocal of the efficiency of transmission of weighted power from output to signal source.

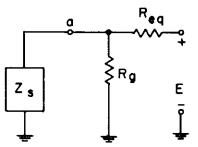

Fig. X-1

Low-frequency noise equivalent circuit for a triode.

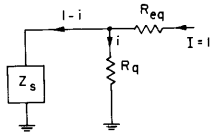

Fig. X-2

The circuit of Fig. X-1 driven at the output by a unit current source.

Theorems 2 and 2a show that the problem of minimizing the noise figure of a network is not one of obtaining maximum power in the network from the source; rather, it is one of obtaining maximum power in the source from the network, for a given power input at the output terminals of the network.

For example, consider the low-frequency noise equivalent circuit of a triode (Fig. X-1). The triode output is proportional to the voltage E, and the resistors R_g, R_{eq} have the relative noise temperatures ρ_g ($\rho_{eq} = 1$). Z_s is the source impedance, referred to point a.

Driving the network by a current source I = 1 at the terminals of E (Fig. X-2), we find that:

(a) If $\rho_g = 0$ then the loss in the weighted power transmission from output to source is fixed (= R_{eq}). Hence the minimum noise figure will be obtained when maximum power is dissipated in Z_s. But Z_s is fed from a source of internal impedance R_g, so that $Z_s = R_g$ will give the minimum noise figure; and this minimum noise figure will be

$$F_{min} = 1 + \frac{\text{losses}}{\text{output}} = 1 + \frac{R_{eq}}{\rho_s R_g \left(\frac{1}{2}\right)^2} = 1 + 4\,\frac{R_{eq}}{\rho_s R_g} \tag{7}$$

(b) If $\rho_g > 0$ then Z_s has to have a smaller value than R_g to decrease the voltage across, and consequently reduce the loss in, R_g. To find the optimum Z_s value and the resulting F_{min} consider Fig. X-2. Let $i = \lambda e^{j\phi}$ where λ is real. Then

$$F = 1 + \frac{R_{eq} + \rho_g R_g |i|^2}{\rho_s R_g \operatorname{Re}\left[i(1-\bar{i})\right]} = 1 + \frac{R_{eq} + \rho_g R_g \lambda^2}{\rho_s R_g (\lambda \cos\phi - \lambda^2)} \tag{8}$$

Letting $\phi = 0$ gives, for a given λ, the minimum possible F. $\phi = 0$ means that both i and 1 - i are in phase, that is, that $Z_s = i R_g/(1 - i)$ is real. And

$$F = 1 + \frac{(R_{eq}/R_g) + \rho_g \lambda^2}{\rho_s (\lambda - \lambda^2)} = 1 + \frac{\rho_g}{\rho_s} \frac{a + \lambda^2}{\lambda - \lambda^2} \tag{9}$$

where

$$a = \frac{R_{eq}}{\rho_g R_g} \tag{10}$$

F will be the minimum when

$$\frac{a + \lambda^2}{\lambda - \lambda^2} = \frac{\frac{d}{d\lambda}(a + \lambda^2)}{\frac{d}{d\lambda}(\lambda - \lambda^2)} = \frac{2\lambda}{1 - 2\lambda} \tag{11}$$

that is, when

$$\lambda = -a + (a^2 + a)^{1/2} \tag{12}$$

and

$$F_{min} = 1 + \frac{\rho_g}{\rho_s} \frac{2\lambda}{1 - 2\lambda} = 1 + 2 \frac{\rho_g}{\rho_s} \left[a + (a^2 + a)^{1/2}\right] \tag{13}$$

Therefore

$$F_{min} = 1 + \frac{2}{\rho_s} \frac{R_{eq}}{R_g} \left[1 + \left(1 + \frac{\rho_g R_g}{R_{eq}}\right)^{1/2}\right] \tag{14}$$

and

$$Z_{s_{opt}} = R_g \frac{\lambda}{1 - \lambda} = R_g \left(\frac{R_{eq}}{R_{eq} + \rho_g R_g}\right)^{1/2} \tag{15}$$

As a second example, assume that each of the outputs $E_1, E_2, \ldots, E_n$ is connected to N through a network N_i ($i = 1, 2, \ldots, n$) (Fig. X-3). $N_1, N_2, \ldots, N_n$ could be the noise equivalent circuits of n tubes and $E_1, \ldots, E_n$ their effective grid voltages. The outputs of the tubes are either added directly or (as in a distributed amplifier) added with phase shifts $a_1, \ldots, a_n$. The output of such a system would then be proportional to $a_1E_1 + a_2E_2 + \ldots + a_nE_n$. The reciprocal of the noise figure F of this output will be given by the efficiency of weighted power transmission from terminals $E_1, \ldots, E_n$ to the source. This efficiency cannot exceed all of the individual efficiencies of transmission through

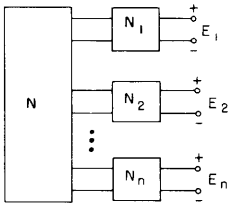

Fig. X-3
A circuit whose outputs $E_1, \ldots, E_n$ are connected to the network N through the networks $N_1, \ldots, N_n$.

networks $N_1, N_2, \ldots, N_n$. But these latter efficiencies represent the reciprocal of the noise figures of systems using only one of the tubes $N_1, \ldots, N_n$. Hence at least one of the tubes $N_1, \ldots, N_n$ can, when properly connected, give a system whose noise figure is $\leq F$.

(This result is a special case, for uncorrelated noise sources, of the noise figure theorem by A. G. Bose and S. D. Pezaris presented at the IRE National Convention of March 1955. A summary of this theorem is given in the Quarterly Progress Report, April 15, 1955, pages 29-31.)

S. D. Pezaris

B. THE WIENER THEORY OF NONLINEAR SYSTEMS*

The Wiener theory of nonlinear systems provides a method for characterizing nonlinear systems independent of their input (the input can be a stationary time function, a periodic function, or a transient) and a method for synthesizing the systems from their characterizing parameters. The synthesis, from the classifying parameters, is accomplished by specifying the parameters $A_\alpha, \ldots, A_\gamma$ in a system of the general form shown in Fig. X-4. The general system (Fig. X-4) is capable of producing an output y(t) which is an arbitrary function of the past of its input x(t) with the restriction that the output become less and less dependent upon the remote past of the input and that the output be continuous with respect to the input (1).

Since the Wiener theory provides us with a physically realizable arbitrary operator on the past of a time function, it includes within its scope a very large class of nonlinear systems. It is of interest to investigate a means for determining the optimum nonlinear filter from the class of systems considered in the Wiener theory. In this report we shall propose a procedure for obtaining this optimum filter for a particular error criterion.

Knowing the form of the general nonlinear system (Fig. X-4), we now have the

*See NOTE at end of Section B, page 42.

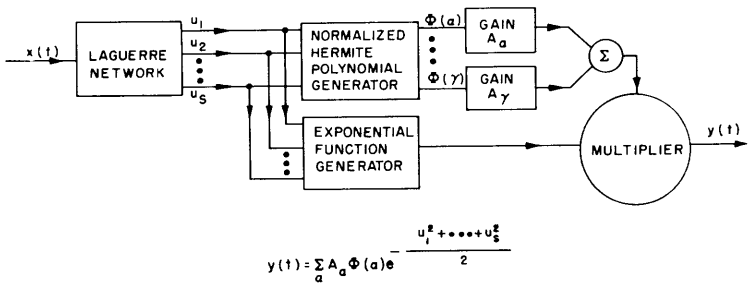

Fig. X-4

Block diagram for the synthesis of nonlinear systems.

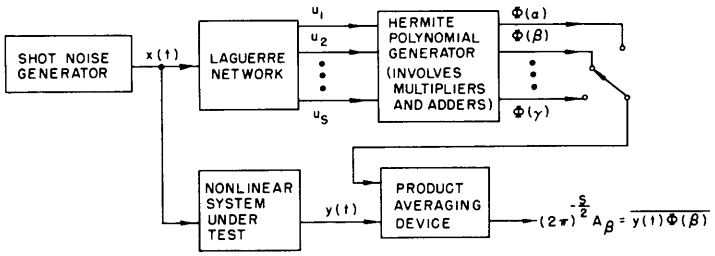

Fig. X-5

Block diagram of the circuitry for the classification of nonlinear systems.

problem of determining the parameters $(A_\alpha, \ldots, A_\gamma)$ for the optimum filter. In solving this problem we shall make use of the Wiener theory of classifying nonlinear systems as illustrated by the block diagram in Fig. X-5. The parameters $A_\alpha, \ldots, A_\gamma$ which classify the nonlinear system under test are obtained by averaging the product of the output of this system and the output of the Hermite polynomial generator when the system and the Laguerre network are driven by white gaussian noise. It is necessary in the Wiener theory of classification that the gaussian noise be white (that is, have a flat power spectrum) in order to establish the statistical independence of the Laguerre coefficients (u_1 through u_S of Fig. X-5) which is necessary to provide the orthogonality of the Hermite functions (1).

Now let us consider the optimum filter problem and see how it differs from the classification problem. Unlike the classification problem, in the determination of an optimum filter we do not have at our disposal the system labeled "Nonlinear System Under Test" in Fig. X-5. In the filter problem this system would be the optimum filter; exactly what we are searching for. It is clear, however, that if we did have the optimum nonlinear filter, then, by driving it with white gaussian noise in the setup of Fig. X-5, we could classify it and then synthesize it in the general form of Fig. X-4.

In other words, a knowledge of the response of the optimum filter to white gaussian excitation is sufficient to characterize and then to synthesize the filter.

In the filter problem we are generally given the input and the desired output time functions and asked to find the system that will produce the desired output with a minimum of error for a given error criterion. If the given input function x(t) happened to have a gaussian distribution with a flat power spectrum, the Wiener method of classification would be applicable to the filter problem. The coefficients $(A_\alpha, \ldots, A_\gamma)$ which characterize the optimum filter could be obtained by simply averaging the product of the desired filter output $y_d(t)$ with the output of the Hermite polynomial generator when the Laguerre network is driven by x(t) as shown in Fig. X-6. Note the difference in the setup of Figs. X-4 and X-5. In Fig. X-4 the "Nonlinear System Under Test" is a system whose output is y(t) when it is driven by the white gaussian input x(t). In general, for the filter problem of Fig. X-6 no such system exists. The characterizing coefficients obtained from the setup of Fig. X-6 correspond to the coefficients of a system which, when driven by the white gaussian x(t), yields $y_d(t)$ with a minimum error for an error criterion that we shall presently discuss.

In the optimum filter problem the given input x(t) is, in general, not gaussianly distributed with a flat power spectrum. Hence, we cannot directly apply the Wiener method of classification to this problem. We shall now discuss an extension of the Wiener method which enables us to determine an optimum filter in the general case when the given input x(t) is not gaussianly distributed with a flat power spectrum.

As shown in Fig. X-7 the optimum filter is obtained by first operating on the given stationary random input $f_i(t)$ to form a white gaussian time function x(t). With x(t) as the input to the Laguerre network, the optimum nonlinear system which operates on x(t) to give the desired output $y_d(t)$ is determined as discussed above (see Fig. X-6). This optimum nonlinear system (W_{opt}) has the form shown in Fig. X-8. If it is desired, the Wiener method of classification could be applied to the system of Fig. X-8 and an equivalent system could be synthesized in the general form of Fig. X-4.

A proposed method for realizing the transformation ϕ of Fig. X-7 consists of cascading a no-storage nonlinear device with a minimum phase linear network. The nonlinear device transforms the probability distribution of $f_i(t)$ into a gaussian distribution. The transfer characteristic (output vs. input) of this device can always be synthesized as a monotonically increasing function. The minimum phase linear network operates upon the output of the nonlinear device to yield a time function with a flat spectrum. Since the network is linear, it does not alter the gaussian character of the time function upon which it operates. Hence its output is gaussianly distributed with a flat spectrum.

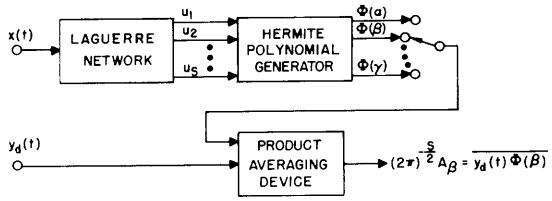

Fig. X-6

Block diagram of experimental setup to determine an optimum filter in the special case where the given filter input x(t) is gaussianly distributed with a flat power density spectrum. $y_d(t)$ is the desired filter output.

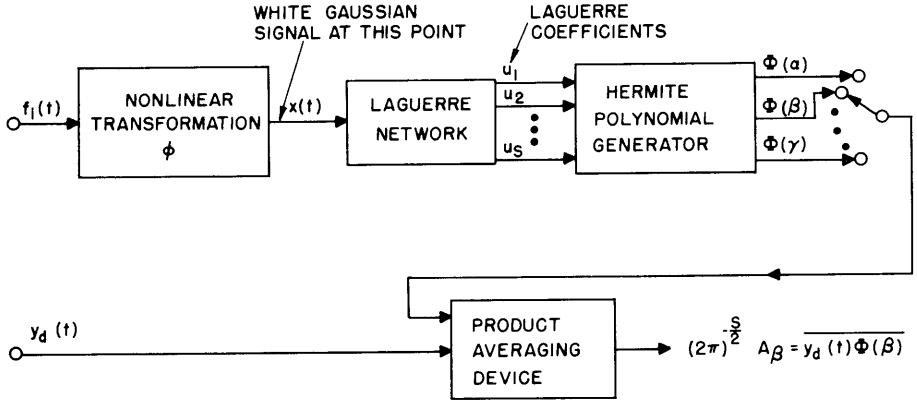

Fig. X-7

Experimental setup for the determination of an optimum nonlinear filter. $f_1(t)$ is the given random filter input; $y_d(t)$ is the desired filter output.

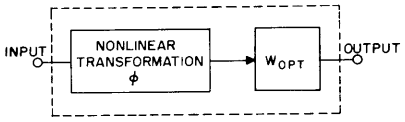

Fig. X-8

General form of optimum nonlinear filter.

Two measurements are necessary for the synthesis of the transformation ϕ. First we measure the probability density of the given filter input $f_i(t)$. This enables us to synthesize the no-storage nonlinear device. Then we measure the power density spectrum of the output of this device when its input is $f_i(t)$. This measurement enables us to synthesize the minimum phase linear network.

It can be shown that the optimum filter obtained by the process indicated above is an optimum filter for a weighted mean square error criterion. Let $y_o(t)$ be the actual output of the filter and $y_d(t)$ be the desired output. Then, for the optimum filter the time average of the quantity $\exp\left[\left(u_1^2 + \ldots + u_s^2\right)/2\right]\left[y_d(t) - y_o(t)\right]^2$ is a minimum. The u's in the exponential weighting factor are the time varying outputs of the Laguerre network. The weighting factor is such that errors occurring during time intervals over which the mean square input voltage to the Laguerre network is large are weighted more heavily than those errors occurring during time intervals over which the mean square input voltage to the Laguerre network is small. The same error criterion applies to the Wiener theory of classification when only a finite number of Laguerre and Hermite functions is used.

NOTE (added at press-time): Since the writing of this report, an error has been found in the proposed procedure for the synthesis of the transformation ϕ. This error will be corrected in the next report.

A. G. Bose

References

1. Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., Oct. 15, 1954, pp. 55-62.

C. A UNIFIED DEFINITION OF INFORMATION

The probabilistic theory of information as developed by Wiener and Shannon has been generalized in such a direction that the discrete and continuous theories form special cases of a unified theory. This first report concerns primarily a fundamental definition of information expressed in a unified form. To treat the problem from a general point of view, it has been necessary to resort to the theories of measure and of Lebesgue integration, although appeal has been made only to those measure-theoretic concepts that are absolutely necessary.

We consider a random variable f which takes on real values x in $(-\infty, \infty)$ according to some monotonic probability distribution function

$$\mu(x) \equiv \text{Probability that } f \leq x \qquad (1)$$

$\mu(x)$ is a nondecreasing function, bounded by zero and one, and well defined for every finite x. At points of discontinuity, it is continuous from the right.

We now introduce a Lebesgue-Stieltjes measure μ generated by $\mu(x)$ such that if E is any measurable subset of the real line, the measure of E is represented by the probability that the value of f belongs to E.

$$\mu(E) = \text{Probability that } f \in E \tag{2}$$

If the entire real line is denoted by the space X, clearly $\mu(X)$ equals 1.

Now let us consider an information process in which we know a priori that the value of a function f lies in a certain set A, and we are told in addition that $f \in B$. Our a posteriori knowledge is that f belongs to both A and B and hence lies in their intersection $A \cap B$. Following Wiener (1), a reasonable measure of the information we have obtained is

$$I = \log \frac{\mu(A)}{\mu(A \cap B)} \tag{3}$$

where the base of the logarithm determines the units of I. Clearly, since $A \cap B$ is included in A, the information gained is nonnegative.

A second more general information process is one in which f is known a priori to be distributed according to $\rho(x)$, and information is obtained which permits the formulation of a new, a posteriori distribution $\nu(x)$. Here, the information has not merely reduced to a subset the range of values of f, but has in fact changed the defining measure on the space. Later we shall see that this is not a different situation from the one considered above, but represents in reality a more general process of which the first is a special case.

Let us suppose that the true value of f is y. Then a reasonable measure of the information associated with the fact that f = y is

$$I(y) = \lim_{\epsilon \to 0} \log \frac{\nu(y+\epsilon) - \nu(y-\epsilon)}{\rho(y+\epsilon) - \rho(y-\epsilon)} \tag{4}$$

To obtain the total information received in the process, we take the average of I(y) with respect to the a posteriori distribution over all possible y. Thus

$$I = \int_{-\infty}^{\infty} \lim_{\epsilon \to 0} \log \frac{\nu(y+\epsilon) - \nu(y-\epsilon)}{\rho(y+\epsilon) - \rho(y-\epsilon)} \, d\,\nu(y) \tag{5}$$

It is apparent that under certain conditions, the integrand in Eq. 5 does not exist. First, consider the case in which $\rho(x)$ contains discontinuities not in common with $\nu(x)$. The integrand becomes infinitely negative at these points. However, since $\rho(x)$ is

monotonic and bounded, the set of all its discontinuities must be countable – hence, also the subset of those not in common with $\nu(x)$. Clearly, this subset is of ν-measure zero (a posteriori probability zero) and does not affect the value of the integral.

Second, consider the case in which $\nu(x)$ has discontinuities not in common with $\rho(x)$. The set of these discontinuities is of positive ν-measure and since the integrand of Eq. 5 becomes infinitely positive on this set, the integral diverges. Thus a necessary condition for the finiteness of I is that any point of discontinuity of $\nu(x)$ represent a discontinuity of $\rho(x)$ also. This condition as well as the form of Eq. 5 can be made more general by an appeal to the concept of absolute continuity and to a restricted form of the Radon-Nikodym theorem.

Definition: A measure ν is said to be absolutely continuous with respect to the measure ρ (in symbols, $\nu \sim\!\!\!\!\rightarrow \rho$) if, given an $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $\rho(E) < \delta$, $\nu(E) < \epsilon$. More simply, whenever E is a set of ρ-measure zero, then $\nu(E)$ equals 0 also. It should be noted that the symbol ($\sim\!\!\!\!\rightarrow$) is not in general symmetric. Whenever we have both $\nu \sim\!\!\!\!\rightarrow \rho$ and $\rho \sim\!\!\!\!\rightarrow \nu$, we write $\nu \leftrightsquigarrow \rho$.

Radon-Nikodym Theorem: Given a measure space X and a pair of measures ν, ρ with $\nu \sim\!\!\!\!\rightarrow \rho$, there exists a finite-valued, nonnegative function θ, integrable with respect to ρ, such that for every measurable set $E \subset X$

$$\nu(E) = \int_E \theta \, d\rho \tag{6}$$

The function θ, which is defined uniquely everywhere except on a set of ρ-measure zero, is called the Radon-Nikodym derivative and is frequently written $d\nu/d\rho$. The proof of a less restricted form of this theorem may be found in Halmos (2).

With the interpretation of

$$\lim_{\epsilon \to 0} \frac{\nu(y+\epsilon) - \nu(y-\epsilon)}{\rho(y+\epsilon) - \rho(y-\epsilon)}$$

as a derivative in the sense of Radon-Nikodym, Eq. 5 can be written in the more general form

$$I = \int_X \log \frac{d\nu}{d\rho} \, d\nu \tag{7}$$

providing a fundamental definition of information applicable to discrete, continuous, and mixed information processes. A necessary (but not sufficient) condition for the finiteness of I is that $\nu \sim\!\!\!\!\rightarrow \rho$.

Now let us show that Eq. 7 is a valid generalization of Eq. 3 and hence reduces to

the fundamental definition of Wiener. We consider again a random function f distributed by $\mu(x)$. If $\mu(x)$ is differentiable, the fact that $f \in A$ may be expressed by an a priori distribution density

$$\rho'(x) = \begin{cases} \mu'(x)/\mu(A) & x \in A \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Or, more generally, we may relax the differentiability condition on $\mu(x)$ and write

$$\frac{d\rho}{d\mu} = \begin{cases} 1/\mu(A) & x \in A \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Similarly, the additional knowledge $f \in A \cap B$ is formulated in terms of the a posteriori measure ν:

$$\frac{d\nu}{d\mu} = \begin{cases} 1/\mu(A \cap B) & x \in A \cap B \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

From Eq. 7, the information obtained is

$$I = \int_X \log \frac{d\nu}{d\rho}\, d\nu = \int_X \frac{d\nu}{d\mu} \left[\log \frac{d\nu}{d\mu} - \log \frac{d\rho}{d\mu}\right] d\mu$$

$$= \int_{A \cap B} \frac{1}{\mu(A \cap B)} \left[\log \frac{1}{\mu(A \cap B)} - \log \frac{1}{\mu(A)}\right] d\mu$$

$$= \log \frac{\mu(A)}{\mu(A \cap B)}$$

as before. Hence Eq. 7 includes the definition of Wiener.

Applying a fundamental logarithmic inequality, we find that

$$I = \int_X \log \frac{d\nu}{d\rho}\, d\nu = \int_X \frac{d\nu}{d\rho} \log \frac{d\nu}{d\rho}\, d\rho$$

$$\geqslant \int_X \left(\frac{d\nu}{d\rho} - 1\right) d\rho = \nu(X) - \rho(X) = 0$$

with the result that information as defined in Eq. 7 is nonnegative.

It may be shown also that if f and g represent two independent events, the

information provided about the joint event (f, g) is the sum of the information about f and g separately. The proof of this result will be omitted here.

It is perhaps of value to show by means of certain examples just what part is played by the Radon-Nikodym derivative in the evaluation of the integral, Eq. 7. Let us first consider a discrete case in which $\rho(x)$ is a step-function having a countable number of discontinuities of magnitude p_k on a set S of elements x_k. We further assume $\nu(x)$ to be a similar function except that its discontinuities have magnitude q_k and occur on a subset M of S. Clearly $\nu \sim \rho$, since every discontinuity of ν occurs in common with one of ρ. The Stieltjes integral in Eq. 7 becomes in this case a simple summation

$$I = \sum_{x_k \in M} q_k \log \frac{q_k}{p_k} \tag{11}$$

over all the discontinuities of $\nu(x)$.

Now let us consider the case in which both $\rho(x)$ and $\nu(x)$ are continuous. In this case the Radon-Nikodym derivative becomes a derivative in the ordinary sense, hence is simply the ratio of the a posteriori and a priori probability densities. The information is then given by the ordinary integral

$$I = \int_{-\infty}^{\infty} \nu'(x) \log \frac{\nu'(x)}{\rho'(x)} \, dx \tag{12}$$

For a third example, we treat a mixed process where $\rho(x)$ and $\nu(x)$ are the monotonic distributions of Fig. X-9. For simplicity, we have chosen a small number of discontinuities. Since every discontinuity of $\nu(x)$ occurs in common with one of $\rho(x)$, and since the interval $(x_3 < x \leq x_4)$ is both of ρ-measure and ν-measure zero, it is easy to verify that $\nu \sim \rho$. Thus the Radon-Nikodym derivative is defined almost everywhere with regard to ρ. This function, which we call $\theta(x)$, is plotted in Fig. X-9. We note first of all that $\theta(x)$ must be well defined at all points of discontinuity of $\rho(x)$ and must have the value p_k/q_k at such points x_k. These values are indicated by the heavy dots in the figure. With the recognition that any monotonic function may be expressed as the sum of a continuous function and a step-function, $\theta(x)$ is given at all other points by the ratio $\nu_c'(x)/\rho_c'(x)$ where the subscript c denotes the continuous part. In this manner we define $\theta(x)$ at all points except perhaps on a set of ρ-measure zero. From Fig. X-9 one sees that $\theta(x)$ is left undefined in the interval $(x_3 < x \leq x_4)$ which is indeed of zero ρ-measure. However, from the absolute continuity condition, any set of ρ-measure zero is also of ν-measure zero — hence the integral in Eq. 7 with respect to ν is unaffected by any values we might assign to $\theta(x)$ in the undefined interval. After we determine the Radon-Nikodym derivative, it is easy to verify that the Stieltjes integral

$$\nu(x) = \int_{-\infty}^{x} \theta(\xi)\, d\rho(\xi) \tag{13}$$

is valid.

It is of interest also to note that the assertion in the Radon-Nikodym theorem that the function $\theta(x)$ be finite-valued does not imply boundedness. In the example of Fig. X-9, it may be seen that $\theta(x)$ does become unbounded in the neighborhood to the left of x_3. However, $\theta(x)$ has the value zero at the point x_3 and is indeed finite-valued everywhere.

For the mixed case, the information obtained in the process of going from $\rho(x)$ to $\nu(x)$ is given by the Stieltjes integral

$$I = \int_{-\infty}^{\infty} \log \theta(x)\, d\nu(x) = q_2 \log \frac{q_2}{p_2} + q_5 \log \frac{q_5}{p_5} + \int_{-\infty}^{\infty} \nu_c'(x) \log \frac{\nu_c'(x)}{\rho_c'(x)}\, dx \tag{14}$$

Thus the information received in a mixed process is simply the sum of a discrete and a continuous part.

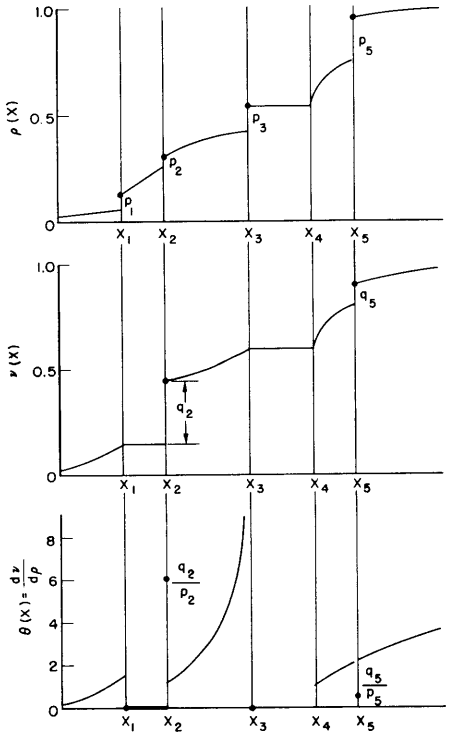

Fig. X-9

The formulation of the Radon-Nikodym derivative in a mixed information process.

Since $\theta(x)$ is unbounded, whether or not the integral in Eq. 14 converges depends upon the behavior of log $\theta(x)$ in the neighborhood to the left of x_3. It is an easy matter to draw examples for which divergence results even though ν is assumed absolutely continuous with respect to ρ. Thus we see that the condition $\nu \sim \rho$ is not a sufficient one for the finiteness of I.

K. H. Powers

References

1. N. Wiener, Cybernetics (John Wiley and Sons, Inc., New York, 1948), p. 75.
2. P. R. Halmos, Measure Theory (D. Van Nostrand Company, Inc., New York, 1950), p. 128.

D. A COMPENSATOR FOR THE DIGITAL ELECTRONIC CORRELATOR

The analog section of the digital correlator periodically samples an input signal and converts the samples to binary numbers. The correlator utilizes a compensator to minimize zero drifts in this section. The binary number generator has a ten digit capacity; that is, it can count up to 1111111111 in the binary system (1023 in the decimal system). 1000000000 is used as a reference zero; that is, for zero input voltage, the number generator will generate 1000000000 (512 in the decimal system). If the correlator is ac coupled, the average value of the signal will be zero, and the average number should be 1000000000. The present compensator operates on the principle that the tenth counter should then be, on the average, on half of the time and off half of the time. Any deviation from this is corrected by controlling the dc level of the input signal with a long time-constant feedback circuit.

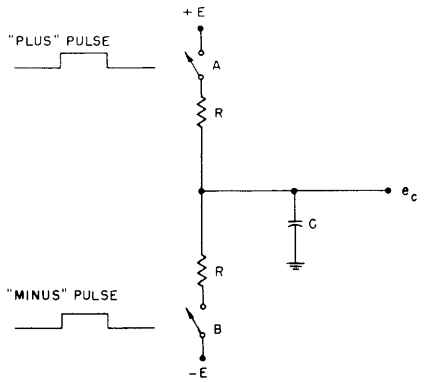

Fig. X-10
Simplified compensator circuit.

This compensator is unsatisfactory, however, for correlation of certain signals. One example is a symmetrical square wave, whose average value could drift within the limits of the peak-to-peak amplitude without compensation. Another example is an unsymmetrical square wave which is never, on the average, half the time on either side of zero.

A scheme has been proposed which provides compensation for any signal. It uses the whole number rather than only one digit. A simplified circuit of this compensator is shown in Fig. X-10. The switches have replaced electronic circuitry to simplify the diagram. For each sample a

pulse is formed; its duration is proportional to the difference between the number and the reference zero. If the difference is positive, switch A is closed for the duration of the pulse; if the difference is negative, switch B is closed for the duration of the pulse.

If P is the average "plus" pulse length and M is the average "minus" pulse length, e_c can be shown to approach a final value of

$$e_{c_{final}} = \frac{E(P-M)}{P+M}$$

with a time constant depending on R and C. This voltage, e_c, will adjust the dc level of the input signal.

This scheme has two distinct advantages over others that were considered. The first is that if a balanced power supply is used, any variations in E are cancelled out when e_c is near zero. The second is that the correction voltage is proportional to E.

This scheme is now being tested.

K. L. Jordan, A. G. Bose

E. MEASUREMENT OF INDUSTRIAL PROCESS BEHAVIOR

Earlier reports (1, 2) stated that the heat exchanger was operated with the random variation of the temperature of the water from the mains as its input, and that the shell and tube temperature variations were recorded through two-stage, RC, highpass networks. The resulting data have been converted to digital form and the autocorrelation function and power-density spectrum of the shell temperature computed. The results are shown in Figs. X-11 and X-12. The highpass filtering has produced an autocorrelation function with a desirable fast-decaying spike in the neighborhood of $\tau = 0$, but with quite irregular behavior for $|\tau| > 5$ sec. It is probable that the dips and peaks in the power-density spectrum in the region above 0.1 cps do not reflect the actual variations in input temperature, but represent the effects of errors in the measuring process. Consequently, a spectrum-shaping digital filter was designed which would preserve the spectral components below 0.1 cps and suppress those above. It was intended that the result of passing the shell data through such a filter would be a time function with an autocorrelation function of the form $\phi_{11} = e^{-|\tau|}$, corresponding to a power-density spectrum flat from zero frequency to 0.16 cps (1 rad/sec). The magnitude of the required filter transfer ratio was found directly from the power-density spectrum; this transfer function is shown in Fig. X-13. The minimum phase corresponding to the magnitude function was computed by a semigraphical method (3). The real part of the transfer function was computed for finding the impulse response of the filter (Fig. X-14), by Guillemin's method (4). As a check on the computations, the magnitude of the Fourier

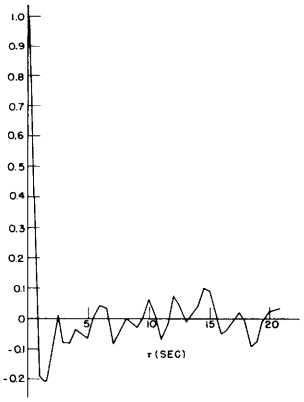

Fig. X-11

Normalized autocorrelation function for use in filter design.

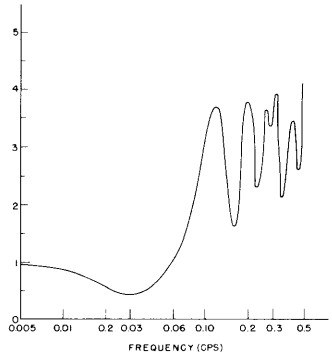

Fig. X-12

Normalized power density spectrum from autocorrelation function of Fig. X-11 for use in filter design.

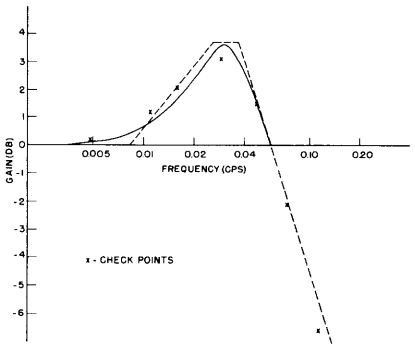

Fig. X-13

Magnitude of the transfer function of digital filter.

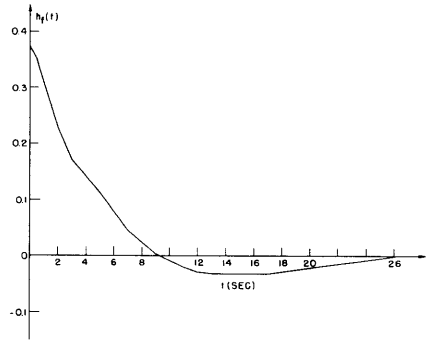

Fig. X-14

Impulse response of digital filter.

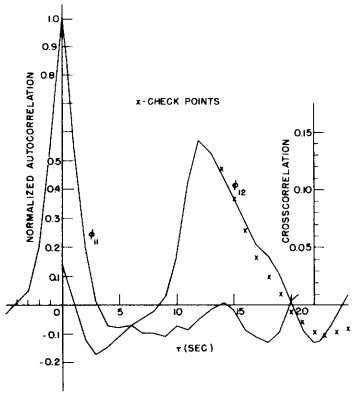

Fig. X-15

Correlation functions computed after two-stage analog filtering and one stage of digital filtering.

transform of the impulse response was computed; a few values are plotted on Fig. X-13. The shell and tube temperature records were each convolved with the impulse response of Fig. X-14, and the autocorrelation and crosscorrelation functions ϕ_{11} and ϕ_{12} were computed in accordance with the modified definitions given by Eqs. 3 and 4 on page 43 of the April report (2). The results are plotted in Fig. X-15. These computations were carried out by the M. I. T. Office of Statistical Services, under the supervision of Mr. W. B. Thurston, using an I. B. M. Model 604 Electronic Calculating Punch. The computed ϕ_{11} has a spike symmetrical about $\tau = 0$ and six sec wide at its base. This agrees with the expected behavior, but small irregularities remain for larger values of τ.

A first approximation to the impulse response relating ϕ_{12} to ϕ_{11} was made by drawing a smooth curve through the positive values of ϕ_{12} from 8 sec to 20 sec, and assuming the response to be zero from 0 sec to 8 sec. It was assumed that the impulse response could assume only positive values because of the resistance-capacitance nature of the heat transfer process. This approximation to the impulse response was convolved with the measured ϕ_{11} to get an approximate ϕ_{12}. A constant multiplier which would minimize the mean-square error between the measured and approximate values of ϕ_{12} was computed. The product of the constant multiplier and the first approximation to the impulse response was assumed to be the impulse response of the heat exchanger. As a check, the assumed impulse response was convolved with ϕ_{11}; the resulting points are plotted on Fig. X-15 for comparison with the measured ϕ_{12}.

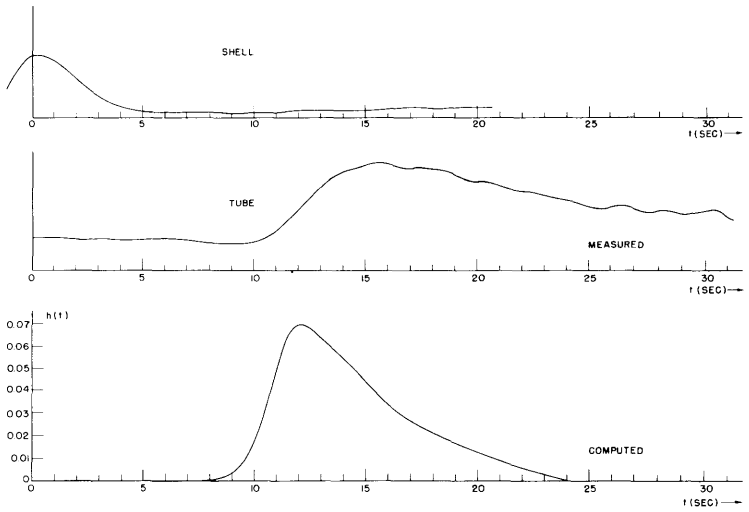

Fig. X-16

Comparison of measured impulse response with impulse response computed from the correlation functions of Fig. X-15.

One method of measuring the impulse response of a system is to apply a step function to the input and measure the input and output through differentiating filters. This test was performed on the heat exchanger, and a mixing valve was used to produce a fast rise of shell temperature. The results are shown at the top of Fig. X-16. The curve marked "tube" can be considered to be the response of the heat exchanger to the differentiated step of temperature occurring at the beginning of the "shell" trace. The impulse response computed from the random-input results is shown at the bottom of the figure for comparison. The two curves are similar in form although a difference in transportation lag (pure time delay) and a smaller difference in time constant are apparent. The discrepancies can be attributed to the fact that in the conventional transient test the rise of shell temperature produced by the mixing valve was accompanied by a slight decrease in flow rate. It should be pointed out that the spike at the origin in Fig. X-15 represents a much faster variation in shell temperature than could conveniently be produced in the physical system.

The experimental data were obtained in the Process Control Laboratory through the cooperation of Professors D. P. Campbell and L. A. Gould.

S. G. Margolis

References

1. Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., Oct. 15, 1954, pp. 75-76.
2. Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., April 15, 1955, pp. 42-46.
3. D. E. Thomas, Bell System Tech. J. 26, 4, 870 (1947).
4. E. A. Guillemin, Technical Report 268, Research Laboratory of Electronics, M.I.T., Sept. 2, 1953 (reprinted from Proc. NEC 9, 513 (1954)).

F. FIELD MAPPING BY CROSSCORRELATION

In the Quarterly Progress Report, October 15, 1954 (pp. 66-69), a scheme using second-order crosscorrelation to locate random signal sources was presented. An example has been worked out showing the technique for locating independent sources which generate Poisson square waves. To simplify the calculation, only two independent sources are considered in illustrating the general procedure.

Figure X-17 shows the area of interest in the xy plane, and the three listening posts L_1, L_2, and L_3 located at points

$$\left(0, \frac{a}{3^{1/2}}\right), \quad \left(-\frac{a}{2}, -\frac{a}{2(3)^{1/2}}\right), \quad \text{and} \quad \left(\frac{a}{2}, -\frac{a}{2(3)^{1/2}}\right)$$

respectively. If $f_1(t)$ and $f_2(t)$ denote the random signals generated by sources S_1 and S_2, the signals received at L_1, L_2, and L_3 are

$$g_1(t) = f_1(t - kd_{11}) + f_2(t - kd_{12}) \tag{1}$$

$$g_2(t) = f_1(t - kd_{21}) + f_2(t - kd_{22}) \tag{2}$$

and

$$g_3(t) = f_1(t - kd_{31}) + f_2(t - kd_{32}) \tag{3}$$

respectively, where d_{ij} is the distance in kilometers between the i^{th} listening post and the j^{th} source, and k is the inverse of the velocity of light in air, that is, $k = (10/3)(\mu sec/km)$.

Let the signal $f_1(t)$ be the sum of a Poisson square wave $f_{1a}(t)$, that takes the values +E and -E, and $f_{1b}(t) = E$; and let $f_2(t)$ be the sum of another Poisson square wave $f_{2a}(t)$, that takes the values +E' and -E', and $f_{2b}(t) = E'$. Then the second-order crosscorrelation function of $g_1(t)$, $g_2(t)$, and $g_3(t)$ is

$$\begin{aligned}\overline{g_1(t)\, g_2(t+\tau_1)\, g_3(t+\tau_2)} = (E+E') \Big\{ &(E+E')^2 + E^2 \Big[e^{-2k_1 |\, k(d_{21}-d_{31})-\tau_1+\tau_2 |} \\ &+ e^{-2k_1 |\, k(d_{11}-d_{31})+\tau_2 |} + e^{-2k_1 |\, k(d_{11}-d_{21})+\tau_1 |} \Big] \\ &+ E'^2 \Big[e^{-2k_2 |\, k(d_{12}-d_{32})-\tau_1+\tau_2 |} + e^{-2k_2 |\, k(d_{12}-d_{32})+\tau_2 |} \\ &+ e^{-2k_2 |\, k(d_{12}-d_{22})+\tau_1 |} \Big] \Big\}\end{aligned} \tag{4}$$

where k_1 and k_2 are the average numbers of zero-crossing per microsecond of the Poisson waves $f_{1a}(t)$ and $f_{2a}(t)$.

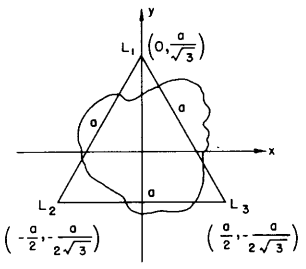

Fig. X-17
Location of listening posts L_1, L_2, L_3 and area of interest; a = 30 km.

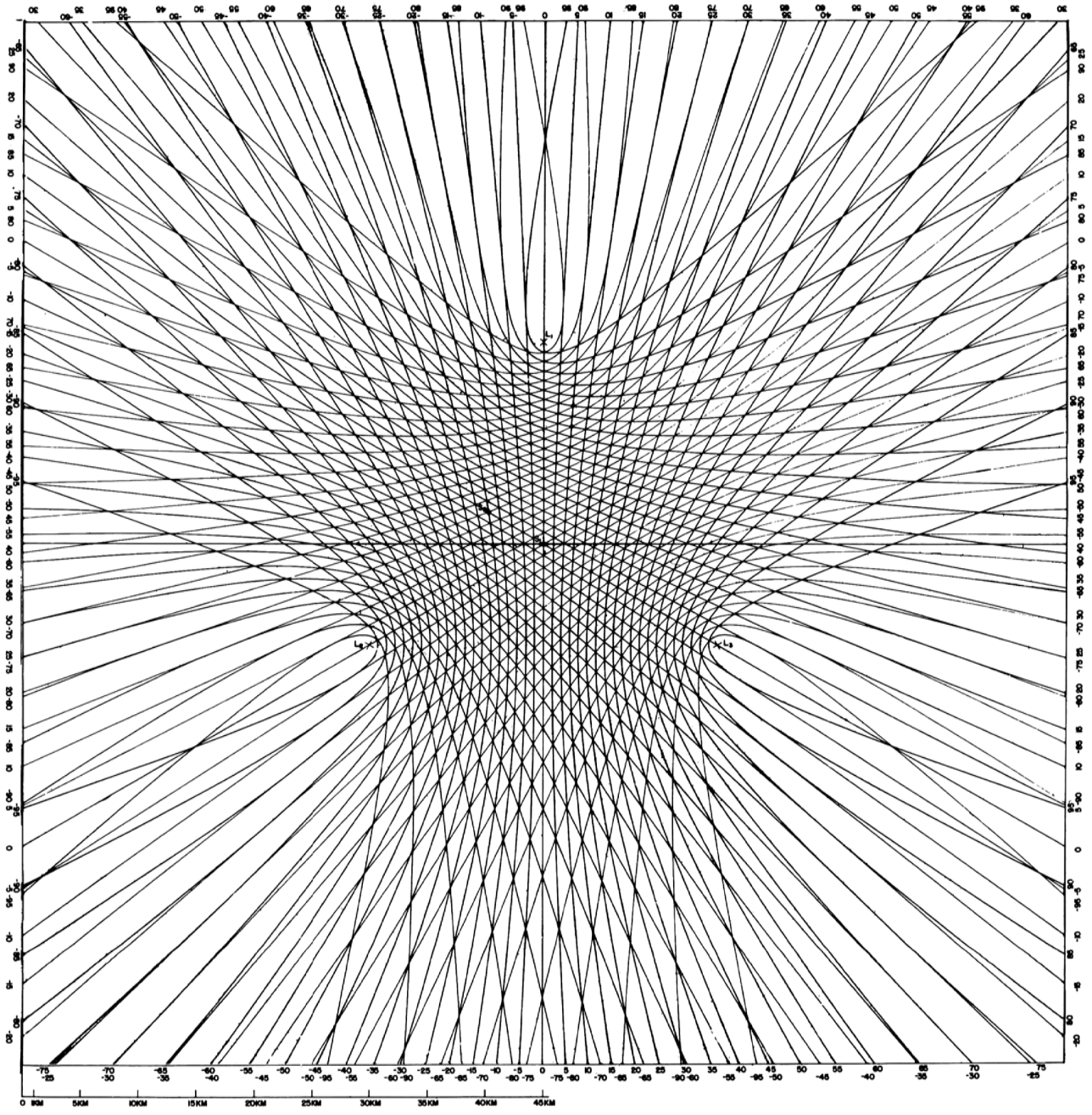

Fig. X-18
Hyperbolic grid system.

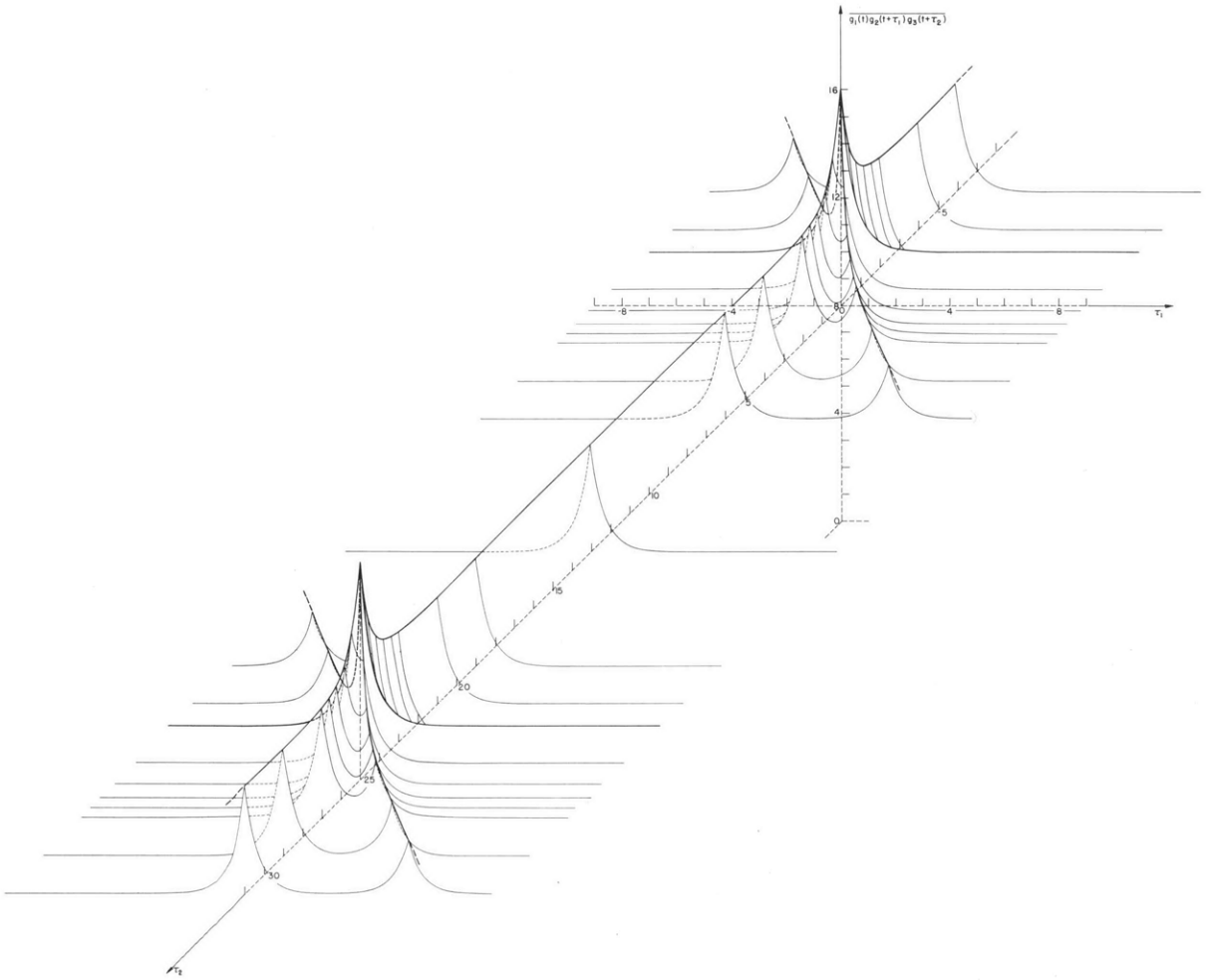

Fig. X-19

Plot of $\overline{g_1(t)\, g_2(t+\tau_1)\, g_3(t+\tau_2)}$.

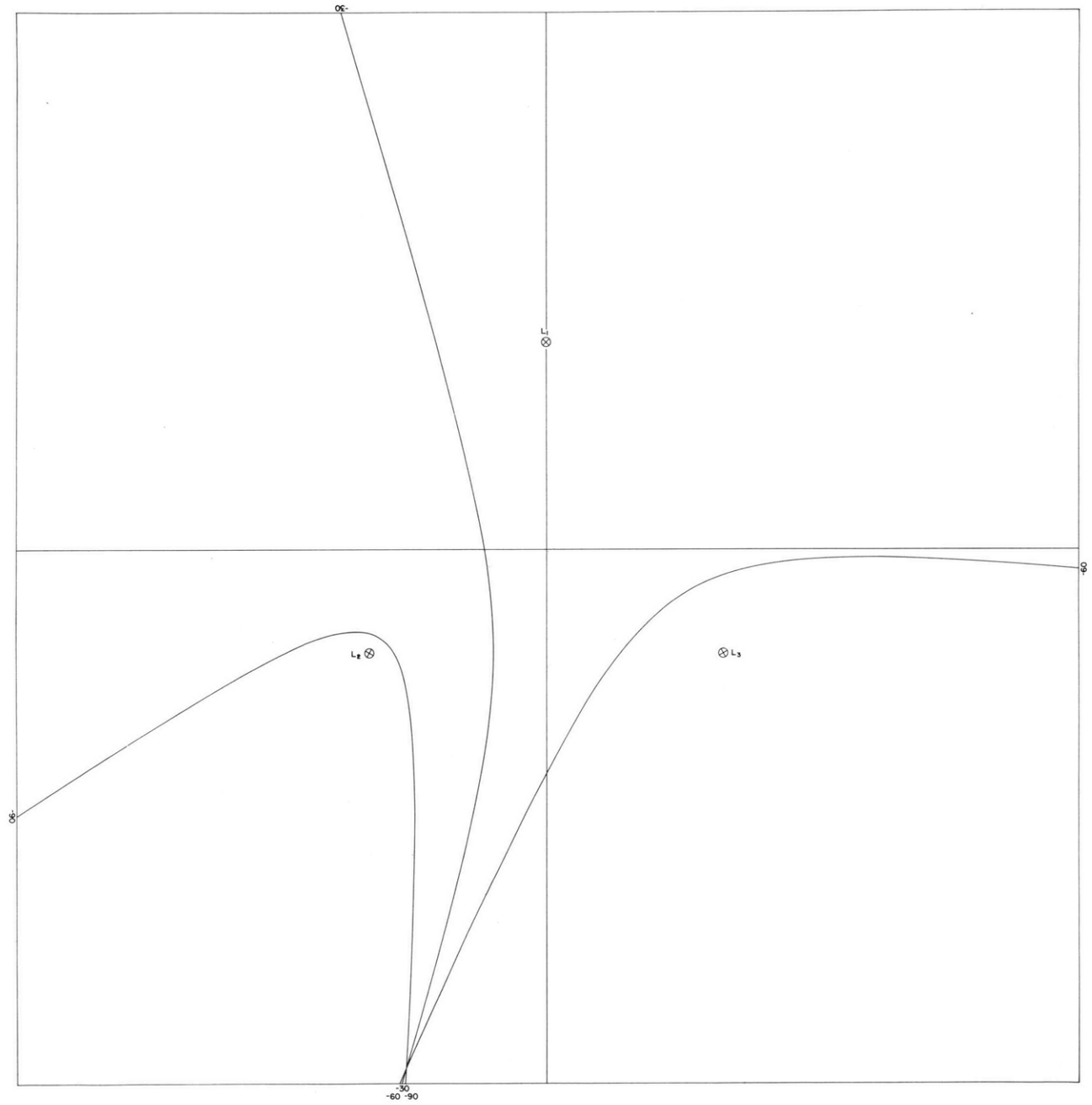

Fig X-20

Trace of hyperbolas from Fig. X-18. τ = -60, -30, and -90.

The values of τ_1 and τ_2 that correspond to the peak values of Eq. 4 determine the positions of the sources. These values of τ_1 and τ_2 for the j^{th} source are denoted by $\left(\tau_1^{(j)}, \tau_2^{(j)}\right)$. To work an example, we must assume a set of values for the d_{ij}'s. For these values we can obtain the values of $\left(\tau_1^{(j)}, \tau_2^{(j)}\right)$ from Eq. 4. In the actual case, however, the values of $\left(\tau_1^{(j)}, \tau_2^{(j)}\right)$ are determined directly by crosscorrelating $g_1(t)$, $g_2(t)$, and $g_3(t)$ without any knowledge of the d_{ij}'s.

For the example, assume that

$$d_{11} = d_{21} = d_{31} = 17.32 \text{ km} \tag{5}$$

and

$$d_{12} = d_{22} = 15.35 \text{ km}, \quad d_{32} = 22.85 \text{ km} \tag{6}$$

These values specify points S_1 and S_2 in Fig. X-18. To simplify Eq. 4, let

$$E = E' = 1 \quad \text{and} \quad k_1 = k_2 = 1 \tag{7}$$

Equation 4 becomes

$$\overline{g_1(t)\, g_2(t+\tau_1)\, g_3(t+\tau_2)} = 2\left(4 + 2e^{-2|\tau_1|} + e^{-2|\tau_2|} + e^{-2|\tau_2-\tau_1|} + e^{-2|\tau_2-25.0|} + e^{-2|\tau_2-\tau_1-25.0|}\right) \tag{8}$$

Equation 8 is plotted in Fig. X-19. From this figure we see that

$$\left(\tau_1^{(1)} = 0,\ \tau_2^{(1)} = 0\right), \quad \left(\tau_1^{(2)} = 0,\ \tau_2^{(2)} = 25.0\right) \tag{9}$$

With these values the sources are located by means of the hyperbolic grid system shown in Fig. X-18. This system consists of three families of hyperbolas superimposed on each other. Points L_1 and L_2 of Fig. X-17 are the foci of the family of hyperbolas given by

$$\left[\left(x + \frac{a}{2}\right)^2 + \left(y + \frac{a}{2(3)^{1/2}}\right)^2\right]^{1/2} - \left[x^2 + \left(y - \frac{a}{3^{1/2}}\right)^2\right]^{1/2} = 0.3\tau \tag{10}$$

When points L_1, L_3, and L_2, L_3 are the foci, the equations of the two families of hyperbolas are

$$\left[\left(x - \frac{a}{2}\right)^2 + \left(y + \frac{a}{2(3)^{1/2}}\right)^2\right]^{1/2} - \left[x^2 + \left(y - \frac{a}{3^{1/2}}\right)^2\right]^{1/2} = 0.3\tau \tag{11}$$

and

$$\left[\left(x+\frac{a}{2}\right)^2+\left(y+\frac{a}{2(3)^{1/2}}\right)^2\right]^{1/2}-\left[\left(x-\frac{a}{2}\right)^2+\left(y+\frac{a}{2(3)^{1/2}}\right)^2\right]^{1/2}=0.3\tau \qquad (12)$$

respectively. In Eqs. 10, 11, and 12, $0 \leq \tau < 100$ μsec and a = 30 km. In Fig. X-18, the values of the parameter τ are shown in the border beside each of the hyperbolas. For instance, if $\tau = -60$, -30, and -90 μsec are the parameter values of the hyperbolas given by Eqs. 11, 12, and 10, respectively, then these hyperbolas are as shown in Fig. X-20 which has been traced from Fig. X-18 for clarity.

In Eq. 8 of the Quarterly Progress Report, October 15, 1954 (p. 68), the quantities $d_{2j} - d_{1j}$, $d_{3j} - d_{1j}$, and $d_{2j} - d_{3j}$ are expressed in terms of $\tau_1^{(j)}$ and $\tau_2^{(j)}$. The first relation in the equation shows that $\tau_1^{(j)}$ is the difference in transmission time of a signal from the j^{th} source to L_2 and L_1. The parameter τ of the family of hyperbolas given by Eq. 10 similarly gives the difference in time for a signal from the j^{th} source to reach L_2 and L_1. Hence, the j^{th} source lies on the hyperbola given by Eq. 10 with $\tau = \tau_1^{(j)}$. By a similar argument, the j^{th} source must also lie on the hyperbola given by Eq. 11 with $\tau = \tau_2^{(j)}$. Hence the intersection of the two hyperbolas determines the location of the j^{th} source. Clearly the source S_1 is at the intersection of two hyperbolas given by Eqs. 10 and 11 with $\tau = 0$. The source S_2 is at the intersection of the hyperbolas specified by Eq. 10 with $\tau = 0$, and the hyperbola given by Eq. 11 with $\tau = 25.0$. The positions of S_1 and S_2 determined this way agree with the positions specified by Eqs. 5 and 6.

This example can be solved by means of the first-order crosscorrelation functions $\overline{g_1(t)\, g_2(t+\tau)}$, $\overline{g_1(t)\, g_3(t+\tau)}$, and $\overline{g_2(t)\, g_3(t+\tau)}$ to determine $\tau_1^{(j)}$ and $\tau_2^{(j)}$, but in general when many sources are involved this procedure becomes highly complicated. For the general problem, second-order crosscorrelation provides a straightforward solution.

J. Y. Hayase