XI. STATISTICAL COMMUNICATION THEORY

Prof. Y. W. Lee	K. L. Jordan	J. Tierney
A. G. Bose	C. S. Lorens	R. E. Wernikoff
J. Y. Hayase	A. H. Nuttall	H. E. White
	K. H. Powers	

RESEARCH OBJECTIVES

This group is interested in a variety of problems in statistical communication theory. Current research work is primarily concerned with: a theory of nonlinear systems, a unified theory of information, preparing an outline of Lebesgue theory for engineers, Markoff processes and flow graphs, second-order correlation functions, probability distribution analyzers, and experimental work on no-memory nonlinear systems.

1. The Wiener method of characterizing nonlinear systems is an effective tool for attacking the problem of optimum nonlinear filtering and prediction. A theory of the experimental determination of optimum nonlinear filters and predictors is being developed. Preliminary work on the extension of the theory to multiple nonlinear filtering and prediction is reported in Section XI-A.

2. The probabilistic theory of information has been generalized in such a direction that the discrete and continuous theories form special cases of a unified theory in which the fundamental information process is regarded as a change of probability measure on an abstract space. In this formulation it is found that the theory of optimum mean-square prediction plays a central rôle in the evaluation of information rates. A study is being made of the loss of information in linear systems.

3. A project is in progress to prepare a heuristic introduction for engineers to measure theory and Lebesgue integration. Because of their importance in information theory, probability theory, and ergodic theory, it seems desirable to prepare an introduction of this kind for the communication engineer who is not a mathematician.

4. Discrete Markoff processes hold an important place in communication problems. The properties of these processes are being investigated with the aid of flow graphs.

5. The project on second-order correlation functions continues with particular emphasis on the properties of these functions.

6. In several phases of our work this group needs to have a means of measuring probability distributions. A survey of existing techniques indicates that further development in achieving greater bandwidth and minimum drift is necessary. Both digital and analog machines are under development.

7. Along with the theoretical work on optimum nonlinear filtering and prediction an experimental investigation on no-memory nonlinear systems will be conducted as a first test of the theory.

Y. W. Lee

A. MULTIPLE NONLINEAR PREDICTION

The problem of multiple nonlinear prediction is that of predicting a time series from a knowledge of the past of related time series. An example, cited by Wiener, is the prediction of weather at one location from the knowledge of the past of the weather at that and other surrounding locations. Wiener's approach to the multiple nonlinear prediction problem is to form all possible nonlinear combinations of the given time functions and then perform a multiple linear prediction on the resulting combinations. The approach described below, on the other hand, first performs linear operations on the given time functions and then performs a multiple no-memory prediction on the resulting

time functions. This latter order of operations is the one used by Wiener in his theory of nonlinear system classification.

As Wiener does in his theory of nonlinear system classification, we shall confine our attention to nonlinear operators whose behavior depends less and less upon the remote past as we push the past back in time. More precisely, we are concerned with those nonlinear operators for which it is sufficient to represent the past of the functions on which they operate by a complete set of orthogonal functions, such as the Laguerre functions. In addition we shall, without further restriction in the practical case, consider only bounded continuous time functions.

Let $z(t+\alpha)$ be the function that we desire to predict and let $x_1(t)$ through $x_p(t)$ be the functions on whose pasts we operate to form our prediction. The set of functions $x_1(t)$ through $x_p(t)$ may, indeed, include $z(t)$. Since our prediction is to be formed from the knowledge of the past of the functions $x_1(t)$ through $x_p(t)$, it is convenient to characterize the past of these functions by the coefficients of complete sets of orthogonal functions. We choose Laguerre coefficients because they are realized by convenient circuitry. Let $u_{1j}, u_{2j}, \ldots, u_{sj}$, be the Laguerre coefficients of $x_j(t)$. For convenience in notation we represent the past of each $x(t)$ by the same number, s, of Laguerre coefficients.

Now let us think in terms of a function space (1) of the past of the inputs $x_1(t)$ through $x_p(t)$. This function space will have ps dimensions and the points in the space are uniquely determined by the Laguerre coefficients of the $x_j(t)$ functions. Hence the problem of prediction becomes that of choosing the best output for each point in the input function space. As we saw in the optimum filter problem (2), if we divide the function space into nonoverlapping cells, and assign an output to each cell, we can represent the output by a series of terms that are mutually orthogonal in time. The orthogonality of the terms is independent of the input time functions and of any weighting factor used in the minimum mean-square-error prediction.

Following the procedure and notation discussed in reference 2, we represent the actual predictor output by $y(t)$, which is given formally by the relation

$$y(t) = F(u_{11}, u_{21}, \ldots, u_{s1}, \ldots, u_{1p}, u_{2p}, \ldots, u_{sp}) \tag{1}$$

in which F is an arbitrary continuous real function of its arguments. Expanding F in terms of gate functions whose arguments are the Laguerre coefficients we have

$$y(t) = \sum_{i_1} \sum_{j_1} \cdots \sum_{h_1} \cdots \sum_{i_p} \sum_{j_p} \cdots \sum_{h_p} a_{i_1, j_1, \ldots, h_1, \ldots, i_p, j_p, \ldots, h_p}$$

$$\times \phi_{i_1}(u_{11})\, \phi_{j_1}(u_{21}) \cdots \phi_{h_1}(u_{s1}) \cdots \phi_{i_p}(u_{1p})\, \phi_{j_p}(u_{2p}) \cdots \phi_{h_p}(u_{sp}) \tag{2}$$

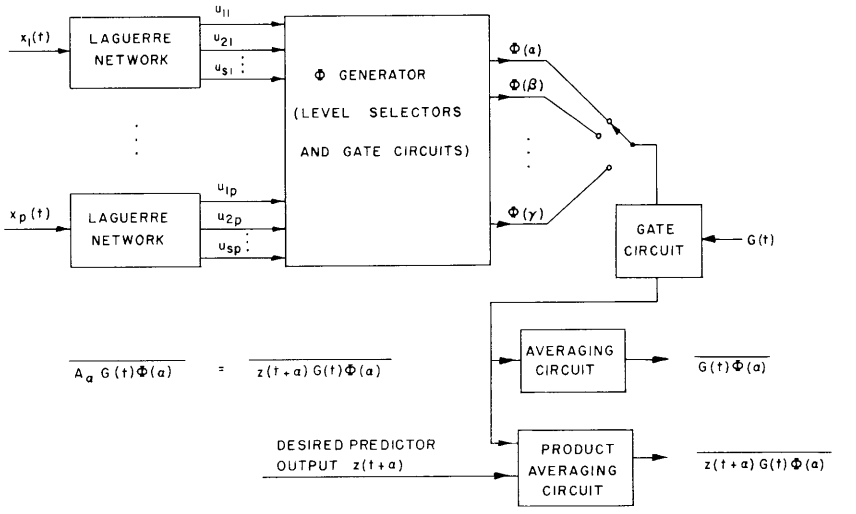

Fig. XI-1. Experimental setup for the determination of the optimum multiple predictor.

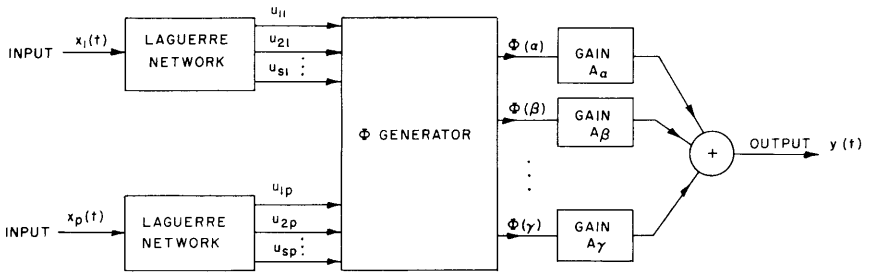

Fig. XI-2. Synthesis of the optimum multiple predictor.

If we associate a $\Phi(\alpha)$ with each product of ϕ's in Eq. 2 and let A_α be the corresponding coefficient $a_{i_1, j_1, \ldots, h_1, \ldots, i_p, j_p, \ldots, h_p}$, Eq. 2 takes the simplified form

$$y(t) = \sum_{\alpha} A_\alpha \, \Phi(\alpha) \tag{3}$$

We adopt a weighted mean-square-error criterion and minimize the integral

$$\mathscr{E} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} G(t) \left\{ z(t+\alpha) - \sum_{\alpha} A_\alpha \, \Phi(\alpha) \right\}^2 dt \tag{4}$$

in which G(t) is a weighting function at our disposal. The $\Phi(\alpha)$ form a nonoverlapping set of functions covering the input function space, hence they are orthogonal in time.

The result of minimizing Eq. 4 and taking advantage of the time-domain orthogonality of the $\Phi(a)$ takes the form of the following relation for the optimum predictor coefficients.

$$A_a \overline{G(t)\,\Phi(a)} = \overline{z(t+a)\,G(t)\,\Phi(a)} \tag{5}$$

The bars in this expression indicate time averages.

Equation 5 is recognized to have the same form as Eq. 12 in reference 2. In Eq. 5, however, the $\Phi(a)$ are products of gate functions of the Laguerre coefficients of more than one input variable. The block diagram for the experimental determination of the optimum coefficients is shown in Fig. XI-1. Having determined the coefficients, the optimum multiple nonlinear predictor can be synthesized in accordance with Eq. 3, as indicated in Fig. XI-2.

A. G. Bose

References

1. Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., Oct. 15, 1955, p. 47.
2. Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., Oct. 15, 1955, pp. 43-49.

B. AN APPLICATION OF PREDICTION THEORY TO INFORMATION RATES

An intimate relationship between information theory and prediction theory is reflected by the point of view (1) that regards a change of probability distribution as the fundamental information process. In such a process, the a priori and a posteriori distributions for the random variable in question are, in effect, conditional distributions, conditioned by our a priori and a posteriori knowledge concerning the variable. In order to evaluate the information gained from such a process, it is necessary to know these distributions, hence their determination is a first step in the evaluation procedure.

It is well known that the best estimation (in the mean-square sense) for a random variable is its conditional expectation, that is, the mean value of that distribution for the variable conditioned by our a priori knowledge. Conversely, that mean value is given by the optimum mean-square predicted value of the variable; herein prediction theory plays an important rôle in the information process. In many instances in which the distributions are not known, it may be possible to evaluate not only the best mean-square prediction, but also the distribution for the error resulting from such a prediction. The conditional distribution for the variable is then simply a translation of the error distribution by an amount equal to the difference between the value of that prediction and the mean value of the error. The resulting distribution has a mean equal

to the best mean-square predicted value and a variance equal to the least mean-square-error of prediction. In certain information processes, we may thus employ mean-square a priori and a posteriori predictions to determine the distributions of the process.

An interesting illustration of these ideas is the problem of evaluating the rate at which one gaussian time sequence gives information about another similar sequence correlated with it. These results will be extended to the case of random functions of a continuous time by a process of sampling in the time domain.

1. Symmetry and Additivity Properties of Information

The information rate problem may be greatly simplified by the application of certain symmetry and additivity properties of the average value of information. We shall state these properties in the form of lemmas that are valid for general information processes, but proofs will be supplied only under the additional assumption that the probability measures of the process are absolutely continuous with respect to Lebesgue measure. This restriction allows the proofs to be given in the common language of probability, while the proofs for the general case require the application of concepts concerning measures on product spaces.

Consider a random process consisting of a sequence $\{\xi_i\}$ $(i = \ldots, -2, -1, 0, 1, 2, \ldots)$ of real random variables such that for every finite subsequence $\{x_i\}$ of n elements of $\{\xi_i\}$ there is defined a probability measure μ on an n-dimensional Euclidean product space X. Let $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_m)$ be a pair of disjoint subsequences of $\{\xi_i\}$ taking values on n- and m-dimensional measure spaces $(X, \mathcal{S}, \mu)$ and $(Y, \mathcal{T}, \nu)$, respectively. Since the union of the sequences x and y represents another subsequence of $\{\xi_i\}$ containing m + n elements, there exist on X and Y for every fixed value y and x, conditional measures μ_y and ν_x, in addition to the absolute measures μ and ν.

Following the definition previously given (1), the information about the sequence x provided by the specification of a particular value y is given by

$$I(y) = \int_X \log \frac{d\mu_y}{d\mu} \, d\mu_y \tag{1}$$

while the average value of I(y) over all possible values of the subsequence y is given by

$$I(x;y) = \int_Y \int_X \log \frac{d\mu_y}{d\mu} \, d\mu_y \, d\nu \tag{2}$$

which by Fubini's theorem becomes

$$I(x;y) = \int_{X\times Y} \log \frac{d\mu_y}{d\mu}\, d\lambda \tag{3}$$

where λ is a measure on the product space $X \times Y$. We call $I(x;y)$ the average information about x provided by y.

Now let $z = (z_1, z_2, \ldots, z_\ell)$ be a subsequence of $\{\xi_i\}$ disjoint with both x and y and defined on a space $(Z, \mathscr{U}, \rho)$. For a given fixed z, known a priori, the information about x provided by a particular value y is

$$\int_X \log \frac{d\mu_{yz}}{d\mu_z}\, d\mu_{yz}$$

where μ_z and μ_{yz} are conditional measures on X for given values z and (y, z), respectively. The average value of the information over all possible values y for a fixed z becomes

$$\int_Y \int_X \log \frac{d\mu_{yz}}{d\mu_z}\, d\mu_{yz}\, d\nu_z$$

Averaging this over the Z-space, we then obtain the average information about x given by y when z is known.

$$I(x;y|z) = \int_Z \int_Y \int_X \log \frac{d\mu_{yz}}{d\mu_z}\, d\mu_{yz}\, d\nu_z\, d\rho \tag{4}$$

LEMMA I: $I(x;y) = I(y;x)$

PROOF: With the assumption that all probability measures considered are absolutely continuous with respect to Lebesgue measure, there exists a probability density distribution $p(x,y)$ on the $X \times Y$ space, and the Radon-Nikodym derivative in Eq. 2 becomes simply the ratio of the a posteriori to the a priori probability densities.

$$I(x,y) = \int_{XY} p(x,y) \log \frac{p(x|y)}{p(x)}\, dx\, dy$$

$$= \int_{XY} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)}\, dx\, dy$$

$$= \int_{XY} p(x,y) \log \frac{p(y|x)}{p(y)}\, dy\, dx$$

$$= I(y;x)$$

LEMMA II: $I(x;y|z) = I(y;x|z)$

PROOF:

$$I(x;y|z) = \int_{XYZ} p(x,y,z) \log \frac{p(x|y,z)}{p(x|z)} \, dx \, dy \, dz$$

$$= \int_{XYZ} p(x,y,z) \log \frac{p(y|x,z)}{p(y|z)} \, dx \, dy \, dz$$

$$= I(y;x|z)$$

LEMMA III: If w is a subsequence of $\{\xi_i\}$ disjoint with x, y, and z, then $I(x;y,z|w) = I(x;y|w) + I(x;z|w,y)$.

PROOF:

$$I(x;y,z|w) = \int_{WXYZ} p(w,x,y,z) \log \frac{p(x|w,y,z)}{p(x|w)} \, dw \, dx \, dy \, dz$$

$$= \int_{WXYZ} p(w,x,y,z) \log \frac{p(x|w,y,z)\, p(x|w,y)}{p(x|w)\, p(x|w,y)} \, dw \, dx \, dy \, dz$$

$$= \int_{WXY} p(w,x,y) \log \frac{p(x|w,y)}{p(x|w)} \, dw \, dx \, dy$$

$$+ \int_{WXYZ} p(w,x,y,z) \log \frac{p(x|w,y,z)}{p(x|w,y)} \, dw \, dx \, dy \, dz$$

$$= I(x;y|w) + I(x;z|w,y)$$

Although the validity of these lemmas does not require that the spaces W, X, Y, and Z be finite dimensional, the extension to the infinite case will be omitted in this report.

Before turning to the information rate problem, let us review some results in the spectral theory of a discrete stationary stochastic process.

2. Spectral Theory

We consider an infinite random sequence $\{f_i\}$ $(i = \ldots, -1, 0, 1, \ldots)$ of real numbers which we assume to be stationary in the wide sense of Khintchine. Hence the correlation coefficient

$$R_m = \overline{f_k f_{k+m}} = R_{-m} \tag{5}$$

exists and depends only on m. According to well-known results of Wiener (2) and Khintchine (3), there exists a bounded nondecreasing spectral function $W(\theta)$ defined on $(-\pi, \pi)$ with $W(-\pi) = 0$, so that

$$R_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-im\theta}\, dW(\theta) \tag{6}$$

The function $W(\theta)$ may be decomposed uniquely into the sum of two nondecreasing functions

$$W(\theta) = W_1(\theta) + W_2(\theta) \tag{7}$$

where $W_1(\theta)$ is absolutely continuous (with respect to Lebesgue measure), and $W_2(\theta)$ is a function whose derivative vanishes almost everywhere. This spectral decomposition represents a corresponding decomposition of the random sequence into the sum of two parts, one of which has an absolutely continuous spectrum, while the other has an almost-everywhere vanishing spectral density. Wold (4) has shown that sequences with the latter type of spectrum are deterministic; that is, their future elements are determined completely in terms of those of their past. Furthermore, Kolmogorov (5) has shown that a sequence with an absolutely continuous spectrum is deterministic if and only if the integral

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |\log W'(\theta)|\, d\theta$$

diverges. Sequences with absolutely continuous spectra for which the integral given above is finite are termed "regular" by Kolmogorov, and only these sequences are useful as information carriers. In the remainder of this report, we shall concern ourselves only with regular sequences.

Since the spectrum of a regular sequence is absolutely continuous, it is completely specified by its derivative, which exists almost everywhere in $(-\pi, \pi)$ and is equivalent to the Fourier series development of the correlation coefficients:

$$W'(\theta) \sim \sum_{m=-\infty}^{\infty} R_m e^{im\theta} \tag{8}$$

The series on the right may be regarded as the boundary values on the unit circle of a function $\Lambda(z)$ of the complex variable $z = re^{i\theta}$. The Fourier coefficients R_m are obtained from $\Lambda(z)$ by

$$R_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Lambda(e^{i\theta})\, e^{-im\theta}\, d\theta = \frac{1}{2\pi i} \oint \Lambda(z)\, \frac{dz}{z^{m+1}} \tag{9}$$

where the contour integral is performed on the unit circle.

Since the sequence is assumed to be regular, the integral

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |\log W'(\theta)|\, d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\log\, \Lambda(e^{i\theta})|\, d\theta$$

is finite. Furthermore, since the correlation coefficients are even, $\Lambda(z) = \Lambda(1/z)$, and we may apply a theorem of Szegö (6) to factor $\Lambda(z)$ into

$$\Lambda(z) = \lambda(z)\, \lambda\left(\frac{1}{z}\right)$$

where $\lambda(z)$ is analytic and nonvanishing inside the unit circle. This is the discrete analog of the well-known spectrum factorization technique employed by Wiener (7) for continuous stationary time functions.

3. The Information Rate of Random Sequences

We now consider a pair of regular sequences $\{f_i\}$ and $\{g_i\}$ that are assumed to be correlated and distributed according to a set of multivariate gaussian distribution functions. It is of interest to determine the rate (per element) at which the sequence f conveys information about the sequence g. That is, given a priori the values of the elements $(\ldots, f_{k-2}, f_{k-1})$, how much additional information about the entire sequence g is given on the average by the next element f_k? Let us denote the past history of the sequence f by the subsequence $p = (\ldots, f_{k-2}, f_{k-1})$. According to Lemma III, the average information about the sequence g given by the pair of elements (f_k, f_{k+1}) when p is known is

$$I(g; f_k, f_{k+1}|p) = I(g; f_k|p) + I(g; f_{k+1}|p, f_k) \tag{10}$$

But the pair (p, f_k) is the past of the sequence $\{f_{k+1}\}$; thus the average information about g given by (f_k, f_{k+1}) when the past p is known is simply the information given by f_k when its past is known plus that given by f_{k+1} when its own past is known. By iteration of Lemma III, it is seen that the average information about g given by a block of N successive elements of f when the past of the block is known is simply N times the average information provided by each element of the block when the past of that element is known. We thus need to determine the average information provided by only a single element in order to obtain the average rate of the sequence.

In the gaussian case, this problem is simply one of linear prediction. We can

determine the average information provided by f_k about a particular element g_{k+p} by obtaining the a priori and a posteriori distribution functions for that element. These distributions will be gaussian and will have means given by the optimum mean-square predicted values of g_{k+p} and variances given by the mean-square-errors of prediction. However, in order to evaluate the information about the entire sequence g, the a priori and a posteriori distributions become infinite dimensional, and if the autocorrelation of the sequence g is taken into consideration, the problem becomes quite formidable. Here we may use Lemma II to great advantage. For the problem at hand,

$$I(g; f_k | p) = I(f_k; g | p) \tag{11}$$

From the right-hand side, we see that if we make predictions of f_k by linear operations on p and (g, p), the a priori and a posteriori distributions are one-dimensional. Thus Lemmas II and III have reduced an infinite-dimensional problem to a single-dimensional one.

The a priori distribution of f_k, that is, its distribution conditioned by a knowledge of its past, is obtained by finding the set of coefficients $\{a_i\}$ $(i = 0, 1, \ldots)$ which minimizes the mean-square-error

$$\overline{\left| f_k - \sum_{i=0}^{\infty} a_i f_{k-1-i} \right|^2} \tag{12}$$

The a priori distribution is gaussian and has density

$$\rho'(x) = \frac{1}{\sqrt{2\pi \sigma_1^2}} \exp \left\{ - \frac{(x - \alpha_k)^2}{2\sigma_1^2} \right\} \tag{13}$$

where σ_1^2 is the minimum value of the mean-square-error, and α_k is the predicted value of f_k

$$\alpha_k = \sum_{i=0}^{\infty} a_i f_{k-1-i} \tag{14}$$

Similarly, the a posteriori density is given by

$$\nu'(x) = \frac{1}{\sqrt{2\pi \sigma_2^2}} \exp \left\{ - \frac{(x - \beta_k)^2}{2\sigma_2^2} \right\} \tag{15}$$

where

$$\sigma_2^2 = \min_{b_i, c_i} \overline{\left| f_k - \sum_{i=0}^{\infty} b_i f_{k-1-i} - \sum_{i=-\infty}^{\infty} c_i g_{k-i} \right|^2} \tag{16}$$

and

$$\beta_k = \sum_{i=0}^{\infty} b_i f_{k-1-i} + \sum_{i=-\infty}^{\infty} c_i g_{k-i} \tag{17}$$

Note that the a posteriori distribution for f_k is that one conditioned by a knowledge of the entire past of f, and the past, present, and future of g. Thus the index of the coefficient c_i runs over all positive and negative integers.

The average information given by a particular element f_k is then

$$I_k = \int_{-\infty}^{\infty} \nu'(x) \log \frac{\nu'(x)}{\rho'(x)} dx = \frac{1}{2} \log \frac{\sigma_1^2}{\sigma_2^2} - \frac{\sigma_1^2 - \sigma_2^2}{2\sigma_1^2} + \frac{(\beta_k - a_k)^2}{2\sigma_1^2} \tag{18}$$

Taking the average of this expression over all k, it is found that

$$\overline{(\beta_k - a_k)^2} = \sigma_1^2 - \sigma_2^2 \tag{19}$$

Hence the average information about the sequence g provided by each element of the sequence f, which is, of course, the rate we have set out to evaluate, becomes simply

$$R(g;f) = \frac{1}{2} \log \frac{\sigma_1^2}{\sigma_2^2} \tag{20}$$

The set of coefficients $\{a_i\}$ that minimizes expression 12 is found quite readily to be the solutions of the relation

$$\sum_{i=0}^{\infty} a_i R_{m-i}^{(ff)} = R_{m+1}^{(ff)} \qquad m \geq 0 \tag{21}$$

Letting

$$A(z) = \sum_{i=0}^{\infty} a_i z^i$$

we can express Eq. 21 in terms of the complex spectrum $\Lambda_{ff}(z)$ of the sequence f.

$$\frac{1}{2\pi i}\oint A(z)\,\Lambda_{ff}(z)\,\frac{dz}{z^{m+1}} = \frac{1}{2\pi i}\oint \Lambda_{ff}(z)\,\frac{dz}{z^{m+2}} \qquad m \geqslant 0 \tag{22}$$

Using methods analogous to the solution of the Wiener-Hopf integral equation, we find that the solution of Eq. 22 is

$$A(z) = \frac{1}{\lambda_{ff}(z)}\sum_{\ell=0}^{\infty} z^{\ell}\left[\frac{1}{2\pi i}\oint \frac{\Lambda_{ff}(\zeta)}{\lambda_{ff}\left(\frac{1}{\zeta}\right)}\,\frac{d\zeta}{\zeta^{\ell+2}}\right] = \frac{1}{z}\left[1 - \frac{\lambda_{ff}(0)}{\lambda_{ff}(z)}\right] \tag{23}$$

where $\lambda_{ff}(z)$ is analytic and nonvanishing in $|z| < 1$. The minimum mean-square-error becomes

$$\sigma_1^2 = \lambda_{ff}^2(0) = \frac{1}{2\pi i}\oint \log\,\Lambda_{ff}(z)\,\frac{dz}{z} \tag{24}$$

The minimization in Eq. 16 yields coefficients (b_i, c_i) which satisfy simultaneously the relations

$$\sum_{i=0}^{\infty} b_i\,R_{m-i}^{(ff)} + \sum_{i=-\infty}^{\infty} c_i\,R_{m+1-i}^{(fg)} = R_{m+1}^{(ff)} \qquad m \geqslant 0 \tag{25}$$

$$\sum_{i=0}^{\infty} b_i\,R_{m-1-i}^{(gf)} + \sum_{i=-\infty}^{\infty} c_i\,R_{m-i}^{(gg)} = R_m^{(gf)} \tag{26}$$

where Eq. 26 must hold for all m. $R_n^{(fg)}$ is the crosscorrelation coefficient of f and g

$$R_n^{(fg)} = \overline{f_k\,g_{k+n}} = R_{-n}^{(gf)} \tag{27}$$

Since Eq. 26 holds for all m, we may write it in terms of the complex variable z as

$$z\,B(z)\,\Lambda_{gf}(z) + C(z)\,\Lambda_{gg}(z) = \Lambda_{gf}(z) \tag{28}$$

Expressing Eq. 25 in spectral form, and making use of Eq. 28 to eliminate C(z), B(z) must satisfy

$$\frac{1}{2\pi i}\oint B(z)\left[\Lambda_{ff}(z) - \frac{\Lambda_{fg}(z)\,\Lambda_{gf}(z)}{\Lambda_{gg}(z)}\right]\frac{dz}{z^{m+1}}$$

$$= \frac{1}{2\pi i}\oint \left[\Lambda_{ff}(z) - \frac{\Lambda_{fg}(z)\,\Lambda_{gf}(z)}{\Lambda_{gg}(z)}\right]\frac{dz}{z^{m+2}} \qquad m \geqslant 0 \tag{29}$$

Since the function

$$\Lambda(z) = \Lambda_{ff}(z) - \frac{\Lambda_{fg}(z)\ \Lambda_{gf}(z)}{\Lambda_{gg}(z)} \tag{30}$$

is non-negative and even on the unit circle, it satisfies necessary and sufficient conditions for a spectrum; hence if Eqs. 29 and 22 are compared, the solution for B(z) will be

$$B(z) = \frac{1}{z}\left[1 - \frac{\lambda(0)}{\lambda(z)}\right] \tag{31}$$

where $\lambda(z)$ is the factor of $\Lambda(z)$ that is analytic and nonvanishing in $|z| < 1$. The a posteriori minimum mean-square-error of Eq. 16 becomes simply

$$\sigma_2^2 = \lambda^2(0) = \frac{1}{2\pi i}\oint \log\left[\Lambda_{ff}(z) - \frac{\Lambda_{fg}(z)\ \Lambda_{gf}(z)}{\Lambda_{gg}(z)}\right]\frac{dz}{z} \tag{32}$$

and the desired rate of Eq. 20 is given by

$$R(g;f) = \frac{1}{4\pi i}\oint \log\left[\frac{\Lambda_{ff}(z)\ \Lambda_{gg}(z)}{\Lambda_{ff}(z)\ \Lambda_{gg}(z) - \Lambda_{fg}(z)\ \Lambda_{gf}(z)}\right]\frac{dz}{z} \tag{33}$$

The symmetry of this equation reveals that g provides information about f at the same rate as that provided by f about g.

4. The Information Rate of Random Time Functions

Let us consider a pair of random time functions f(t) and g(t) which are multivariate gaussian and hence are described statistically by correlation functions $\phi_{ff}(\tau)$, $\phi_{gg}(\tau)$, and $\phi_{fg}(\tau)$. If we imagine the functions to be sampled in time at equal intervals T, the values of the functions at sample points form discrete sequences with correlation coefficients given by

$$R_m^{(fg)} = \phi_{fg}(m\,T) \tag{34}$$

and spectra

$$\Lambda_{fg}(e^{i\theta}) = \sum_{m=-\infty}^{\infty} \phi_{fg}(m\,T)\ e^{im\theta} \tag{35}$$

Letting $\theta = \omega T$, we note that as $T \to 0$

$$\lim_{T\to 0} T\ \Lambda_{fg}(e^{i\omega T}) = \int_{-\infty}^{\infty} \phi_{fg}(\tau)\, e^{i\omega\tau}\, d\tau = \Phi^{*}_{fg}(\omega) \tag{36}$$

where $\Phi^{*}_{fg}(\omega) = \Phi_{gf}(\omega)$ is the cross spectral density. The average information per sample point is

$$R_T(g;f) = \frac{T}{4\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} \log \left[\frac{\Lambda_{ff}(e^{i\omega T})\ \Lambda_{gg}(e^{i\omega T})}{\Lambda_{ff}(e^{i\omega T})\ \Lambda_{gg}(e^{i\omega T}) - |\Lambda_{fg}(e^{i\omega T})|^2} \right] d\omega \tag{37}$$

Dividing by T and taking the limit as $T \to 0$, we obtain the average time rate at which the function f(t) gives information about g(t):

$$R(g;f) = \frac{1}{4\pi} \int_{-\infty}^{\infty} \log \left[\frac{\Phi_{ff}(\omega)\ \Phi_{gg}(\omega)}{\Phi_{ff}(\omega)\ \Phi_{gg}(\omega) - |\Phi_{fg}(\omega)|^2} \right] d\omega \tag{38}$$

If we consider the special case treated by Shannon (8) in which g(t) = m(t) is a gaussian message limited to a frequency band (0, W), and f(t) = m(t) + n(t) is that message disturbed by an additive gaussian noise uncorrelated with the message, the rate of Eq. 38 becomes

$$R(m;m+n) = \frac{1}{2\pi} \int_{0}^{2\pi W} \log \left[1 + \frac{\Phi_{mm}(\omega)}{\Phi_{nn}(\omega)} \right] d\omega \tag{39}$$

in complete agreement with the channel capacity given by Shannon for this case.

K. H. Powers

References

1. K. H. Powers, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., July 15, 1955, p. 42.
2. N. Wiener, Generalized harmonic analysis, Acta Math. 55 (1930).
3. A. Khintchine, Korrelationstheorie der stationären stochastischen Prozesse, Math. Ann. 109 (1934).
4. H. Wold, A Study in the Analysis of Stationary Time Series (Almquist and Wiksells, Uppsala, 1938).
5. A. N. Kolmogorov, Stationary sequences in Hilbert space, Bull. State Univ., Moscow, Ser. Math., 2 (1941).
6. G. Szegö, Über die Randwerte einer analytischen Funktion, Math. Ann. 84 (1921).

7. N. Wiener, Extrapolation, Interpolation and Smoothing of Stationary Time Series (John Wiley and Sons, Inc., New York, 1949).

8. C. E. Shannon, Communication in the presence of noise, Proc. IRE 37, 10 (Jan. 1949).

C. AN OUTLINE OF LEBESGUE THEORY FOR ENGINEERS

The object of this project is to produce a motivated, heuristic introduction to the principal concepts of the modern theory of measure and integration, with emphasis on the interrelations between this theory and ordinary Riemann theory. The outline should serve either as a source for meaningful definitions of terms most frequently encountered in statistical work or as a guide to nonmathematicians in interpreting the rather concise modern presentations of measure and integration. The motivation for writing the outline is the importance of measure theory and Lebesgue integration in rigorous treatments of such subjects as ergodic theory, representation of functions by orthonormal series, Hilbert space, information theory, probability theory, and so forth.

The treatment is not restricted to Lebesgue measure of the real line. Instead, it follows the contemporary trend (1, 2) of developing the theory in completely general terms. The main reason for this choice is that it facilitates the application of the theory to practical problems. For example, having developed measure theory in general terms, it is very easy to specialize it to the case of probability measure, thus making the whole body of theorems available for the study of information theory. On the other hand, after a few initial complications, it is no more difficult to develop integration theory in terms of general measures than it is to develop it for Lebesgue measure only.

The scope of the outline is summarized in the following list of section headings and the accompanying short abstracts of the contents of each section.

Section 1. Introduction. The Riemann integral. Limitations of the Riemann definition. Preliminary sketch of the construction of a more general integral.

Section 2. Simple functions. Some useful notations. Simple functions and their approximation properties. Integrals of simple functions and their use in the definition of a general integral.

Section 3. Measure theory. I. Set theoretical concepts and notations. Set functions. Rings and σ-rings, with examples. Definition of measure on a ring. Examples. Properties of measures.

Section 4. Measure theory. II. The search for a suitable domain for measure functions. Measurable sets and measurable functions. Their relation to integration theory. Properties of measurable functions. Extension of measures to σ-rings.

Section 5. Integration. Rigorous definition of integrals using measure theory. The

determination of integrals and their use in theoretical work in engineering. Properties of integrals. Equivalence classes of functions.

Section 6. Convergence. Sequences of functions. Pointwise and almost-everywhere convergence. Examples. Convergence in measure and convergence in the mean. Interrelations between various types of convergence. Limits of convergent sequences and equivalence classes of functions.

Section 7. Limits and integration. Interchange of limits and integration. Term-by-term integration of series. Requirements with Riemann integration. Requirements with general integrals. Lebesgue monotone and bounded convergence theorems.

Section 8. Absolute continuity and differentiation. The integral as a set function. Lebesgue-Stieltjes measures. Which set functions are integrals? Absolute continuity. The Radon-Nikodym theorem. Radon-Nikodym derivatives.

The work on the outline is substantially completed. It will be published as a technical report.

R. E. Wernikoff

References

1. P. R. Halmos, Measure Theory (D. Van Nostrand Co., New York, 1950).
2. M. E. Monroe, Introduction to Measure and Integration (Addison-Wesley Publishing Company, Inc., Cambridge, Mass., 1953).

D. SYNTHESIS FOR INTERFERENCE FILTERING

A design procedure for interference filtering was completed in a form that greatly facilitates the synthesis of the networks involved. The procedure produces the mean-square optimum under infinite-delay conditions without necessarily involving infinite delay in the synthesis of physical networks. It is worked out in a form in which approximate methods of synthesis can be easily applied.

The term "interference filtering" applies to filtering problems involving the separation or generation of some desired signal from a set of signals. Wiener (1) went a long way toward solving the problem. Costas (2) produced an infinite-delay solution and tried to apply the theory to a physical system.

The formulation of the general interference filtering problem is shown in Fig. XI-3. It is assumed that N statistically stationary signals (3) are passed through N linear networks and added together to give one output signal. The networks are so designed that the output signal is as close (in the mean-square sense) as possible with linear

networks to some desired signal. It is assumed that the correlations among the N signals and between the N signals and the desired signal are known.

The mean-square-error is

$$\mathscr{E} = \left\langle \left[f_d(t) - f_o(t) \right]^2 \right\rangle \tag{1}$$

Variation of the error with respect to the impulse response $h_q(t)$ of the q^{th} network gives the familiar Wiener-Hopf equation

$$\phi_{qd}(\tau) = \sum_{p=1}^{N} \int_0^\infty h_p(\sigma)\, \phi_{qp}(\tau-\sigma)\, d\sigma \qquad \tau \geq 0 \qquad q = 1, \ldots, N \tag{2}$$

The ϕ terms are the standard correlation functions.

Following the standard techniques for solving the Wiener-Hopf equation, it can be shown that the solution to this equation for any of the networks is

$$H_q(\omega) = \frac{1}{\Phi^+_{qq}(\omega)} \frac{1}{2\pi} \int_0^\infty e^{-j\omega t}\, dt \int_{-\infty}^\infty \frac{\Phi_{qd}(\lambda) - \sum_{\substack{p=1 \\ p\neq q}}^{N} H_p(\lambda)\, \Phi_{qp}(\lambda)}{\Phi^-_{qq}(\lambda)} e^{j\lambda t}\, d\lambda$$

$$q = 1, 2, \ldots, N \tag{3}$$

The Φ terms represent the Fourier transforms of the corresponding correlation function. The Φ^+_{qq} and Φ^-_{qq} terms represent the separation of the Φ_{qq} term into its upper-half and lower-half plane poles and zeros.

By permitting a delay α in the desired signal and a delay α in each of the channels we have the situation of Fig. XI-4. As the delay α approaches infinity the solution may be expressed:

$$\sum_{p=1}^{N} H_p(\omega)\, \Phi_{qp}(\omega) = \Phi_{qd}(\omega) \qquad q = 1, 2, \ldots, N \tag{4}$$

From this linear set of equations, we are able to solve for $H_n(\omega)$ in terms of the cross-power spectra without becoming involved in factoring the power spectra of the N signals. Permitting the infinite delay α allows the linear networks to perform a better minimization of the mean rms error. It is generally felt that this limiting minimization is the best that can be attained with linear networks.

By defining the error as in Eq. 5 it is possible to calculate the minimum-error power spectrum, Eq. 6, by using the solution of Eq. 4.

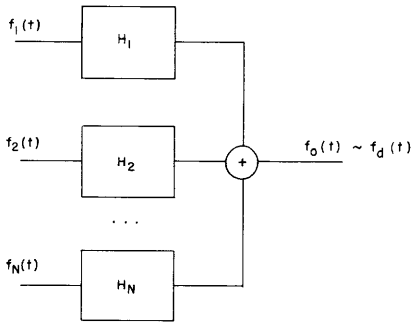

Fig. XI-3. Interference filtering.

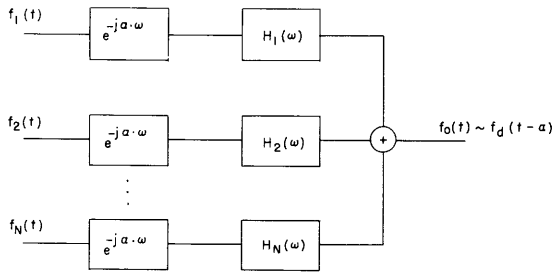

Fig. XI-4. Interference filtering with sufficient delay.

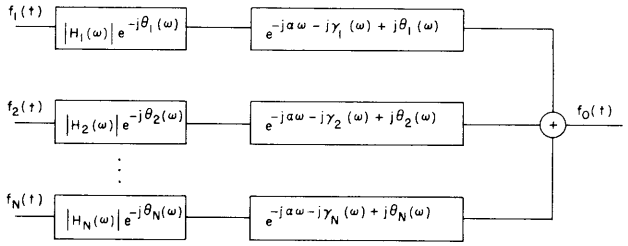

Fig. XI-5. A semidisjoint method of synthesis.

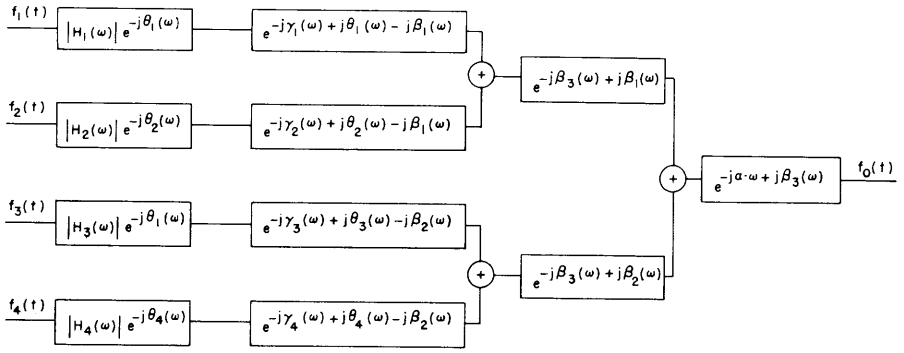

Fig. XI-6. Four-channel interference filtering.

$$\mathscr{E}(t) = f_d(t - a) - f_o(t) \tag{5}$$

$$\Phi_{\mathscr{E}\mathscr{E}}(\omega) = \Phi_{dd}(\omega) - \sum_{n=1}^{N} H_n(\omega)\, \Phi_{dn}(\omega) \tag{6}$$

Any system that produces this minimum-error power spectrum is equivalent to the optimum linear infinite-delay system.

The general solution for $H_n(\omega)$ may be expressed in the form

$$H_n(\omega) = |H_n(\omega)|\, e^{-j\gamma_n(\omega)} \tag{7}$$

where the phase $\gamma_n(\omega)$ may have little or no relation to the magnitude $|H_n(\omega)|$.

"Brute-force" synthesis of the networks $H_n(\omega)\, e^{-ja\cdot\omega}$ is usually not very easy and lacks an appreciation of the real purpose of the networks. The following discussion presents some of the more practical methods of realizing these networks.

Assume the phase $\theta_n(\omega)$ to be such that $|H_n(\omega)|\, e^{-j\theta_n(\omega)}$ is a realizable minimum-phase-transfer network. Then for sufficiently large a it is possible to construct the network shown in Fig. XI-5. This is one way of realizing the system of networks. It has the advantage of breaking up the synthesis into two semidisjoint parts: the realization of the magnitude, and the phase.

The synthesis of the transfer networks $|H_n(\omega)|\, e^{-j\theta_n(\omega)}$ is usually quite straightforward with present-day synthesis techniques. The phase $\theta_n(\omega)$ should be determined from the synthesized $|H_n(\omega)|$ and not from the theoretical condition between the magnitude and the phase. The phase-network synthesis is much more critical and subsequently more difficult.

Once the correct spectral magnitudes are obtained, we are interested in phasing sinusoids against each other in such a way that certain components are enhanced while others are destroyed. An indication of this critical phase is shown by the fact that two signals of equal amplitude must be within one degree out-of-phase in order to get a 35-db suppression by adding the signals. Thus, what is really important in the phase synthesis is the difference in phase between the different channels. Making a sufficiently large just makes it possible to realize the phase networks. An a that is larger than actually needed only makes it harder to control the phase differences between the networks.

Since careful phasing is so important, it becomes expedient to combine the channels in pairs and then combine the pairs, pair-wise. In each combination a functional phase term $\exp\left[-j\beta_n(\omega)\right]$ is inserted instead of the linear phase $\exp(-ja\cdot\omega)$, the function $\beta_n(\omega)$ being used to facilitate the synthesis of the phase difference. Usually this kind of phase synthesis can be accomplished by simple cut-and-try (4). After the channels are all

combined another phase network $\exp\left[-j\alpha\cdot\omega+j\beta(\omega)\right]$ may be used to restore the uniform delay. If the delay distortion is of no consequence (as in speech and music) it may be omitted. A possible four-channel synthesis is shown in Fig. XI-4.

The practical aspects of this synthesis are the realization of the magnitude of the transfer function separate from the phase and the synthesis of the phase difference rather than the attainment of some absolute value of phase dependent upon the constant α. Only after all the networks are combined is it necessary to synthesize an absolute phase, this synthesis being only to correct delay distortion in the output.

Calculation of the error spectrum of a network of the type shown in Fig. XI-6 produces an equation identical to Eq. 7. Thus this synthesis produces a system equivalent to the optimum linear infinite-delay system.

One other practical technique is well worth consideration. Very often the desired signal requires essentially zero power in certain frequency regions. The networks may be carefully designed in frequency ranges of importance while paying little attention to the frequency regions where the desired signal has little power. In order to compensate for introduced errors in the synthesis, the network structure is followed by a filter that substantially attenuates these undesired frequency regions and passes the desired frequencies.

An excellent example of the use of this theory is in the design of the audio combining networks for synchronous detection. Two audio signals are obtained in the process of demodulation, one containing the audio signal and noise, the other containing essentially just noise. The two noise signals are correlated but not identical. By properly combining these two signals it is possible to obtain an increase in the signal-to-noise ratio at the output.

For a given spectrum of the noise (as might arise from adjacent channels or from jamming) it is possible to apply this theory and build combining networks that will handle different types of interference (5).

In summary, this report has put forth some practical aspects that arise in interference filtering in the design of the linear networks. The synthesis is formulated in such a way as to be performed in discrete steps, each step being performed in a simple, straightforward manner. The magnitudes of the transfer functions are first synthesized, then the phase differences are synthesized as the channels are combined in pairs. This procedure produces a system of networks equivalent to the optimum linear infinite-delay system.

C. S. Lorens

References

1. N. Wiener, Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications (John Wiley and Sons, Inc., New York, 1949).

2. J. P. Costas, Interference filtering, Technical Report 185, Research Laboratory of Electronics, M.I.T., March 1, 1951.

3. It is important to note that the mathematics and ideas are just as applicable to aperiodic signals as to statistically stationary signals. The only difference is that the total integrated error is used instead of the time-average error.

4. The cut-and-try process adds all-pass sections alternately to the pair of channels working out from zero frequencies.

5. In a forthcoming Technical Information Series report of the General Electric Company, C. S. Lorens deals with this specific problem at greater length. The networks were constructed and found to perform satisfactorily.

E. A PROOF OF THE NONINTERSECTING LOOP RULE FOR THE SOLUTION OF LINEAR EQUATIONS BY FLOW GRAPHS

In the past few years a new method that makes use of flow graphs for the solution of a linear set of equations has been evolved. With the use of flow graphs, solutions are obtained almost miraculously from a geometric relation of the coefficients of the linear set of equations.

In Technical Report 303 (to be published) S. J. Mason presents a rule for the solution of a flow graph in one reduction step and includes a proof of the general validity of the rule. His proof depends heavily upon previously derived properties of flow graphs.

The present report gives another proof of the general validity of the rule. The proof is based on Cramer's well-known rule for the solution of a linear set of equations.

In matrix notation a linear set of equations may be expressed as

$$AX + C = 0 \tag{1}$$

where A is an $N \times N$ coefficient matrix, X is the column matrix of unknown variables, and C is the column matrix of known variables.

By Cramer's rule, Eq. 1 has the following solution

$$x_j = \sum_{n=1}^{N} - C_n \frac{|A_{nj}|}{|A|} \tag{2}$$

where A_{nj} is the cofactor of the term a_{nj} of A.

In flow-graph notation, Eq. 1 can be written

$$(A+I)\,X + C = X \tag{3}$$

where I represents the unit matrix.

Nodes are associated with the unknown variables $(x_1, \ldots, x_N)$ and the known variables $(C_1, \ldots, C_N)$, while coefficients are associated with the branches connecting the nodes.

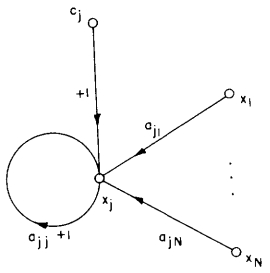

Fig. XI-7. Flow-graph representation of the equation:
$x_j = a_{j1}x_1 + \ldots + (a_{jj}+1)x_j + \ldots + a_{jN}x_N + C_j$.

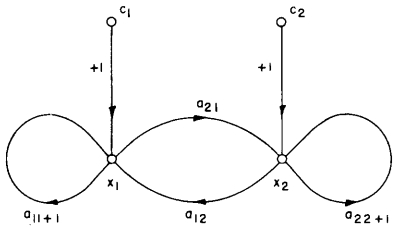

Fig. XI-8. Flow graph of the linear set of equations:
$a_{11}x_1 + a_{12}x_2 + C_1 = 0$
$a_{21}x_1 + a_{22}x_2 + C_2 = 0$.

Figure XI-7 represents a particular equation from Eq. 3. Figure XI-8 represents the flow graph for a second-order set of equations.

An important geometric property of a flow graph is the set of loops. A loop is defined as an ordered sequence of the nodes corresponding to either known or unknown variables, no variable appearing more than once in any one loop. The initial member is immaterial as long as the order is preserved. Two nonintersecting loops are sequences that have no common elements. A loop product is the product of the coefficients associated with the branches that go from one member of the sequence to the next as the loop $(x_3\ x_5\ x_8\ x_7)$ has the loop product $L_p = a_{53}\,a_{85}\,a_{78}\,a_{37}$. The loop product associated with the loop (x_i) is $L = a_{ii} + 1$.

The rule presented by Mason for the solution of the flow graph, and thus the linear set of equations, is

$$x_j = \sum_i C_i \frac{\sum_{k_i} G_{k_i}\,\Delta_{k_i}}{\Delta} \tag{4}$$

where G_{k_i} is the k^{th} sequence product from the node C_i to the unknown variable node x_j.

$$\Delta = 1 - \sum_w L_w + \sum_w P_w^2 - \sum_w P_w^3 + \ldots \tag{5}$$

where P_w^n is the product of the w^{th} possible combination of n-nonintersecting loops. Δ_{k_i} is a modified Δ where the loops are restricted to those that do not intersect the loop associated with the sequence product G_{k_i}.

To show that Mason's rule of nonintersecting loops is valid, it is sufficient to have only C_1 nonzero. The other C's may be added by superposition, as is shown by Eqs. 2 and 4.

In the evaluation of the determinant $|A|$ the numbering is immaterial, since a determinant permuted in the same order of column and row has the value of the original determinant. In like manner the loop products are independent of the numbering, since the product is only dependent upon the geometry of the flow graphs.

A general term of the number Δ is made up of products of factors of the form a_{ii} and loop products containing two or more coefficients. The coefficient of any term which does not have N coefficients will be zero. That is, if there are q loop products involving s variables ($s = s_a + \ldots + s_q$) and t factors of the form a_{ii} involved in the term, then the coefficient of the term $L_a \ldots L_q a_{ii} \ldots$ will be

$$(-1)^{q+t}\left\{1 - (N-s-t) + \binom{N-s-t}{2} - \ldots \pm 1\right\} = \begin{cases} 0 & s + t < N \\ (-1)^N (-1)^{s-q} & s + t = N \end{cases}$$

For this general term it is possible to renumber the flow graph so that the term appears in the following form:

$$(-1)^N (-1)^{s-q} \left(a_{1s_1}\, a_{21}\, a_{32}, \ldots, a_{s_1\, s_1-1}\right) \cdot \left(a_{s_1+1,\, s_1+s_2}, \ldots, a_{s_1+s_2\; s_1+s_2-1}\right) \ldots a_{s+1\; s+1}, \ldots, a_{NN} \qquad (6)$$

If the determinant $|A|$ is evaluated and renumbered, the same term appears as in Eq. 6, except for the factor $(-1)^N$. The sign $(-1)^{s-q}$ is correct, since there are $s_i - 1$ permutations of the second index number associated with each loop, or a total of $s - q$ permutations associated with the term. Then each term of Δ appears in the evaluation of $|A|$, being modified by the factor $(-1)^N$.

For a general term in the expansion of $|A|$ the matrix can be renumbered so that it appears as Eq. 6 except for the factor $(-1)^N$. Thus each term of $|A|$ appears in Δ modified by the factor $(-1)^N$. Thus in general we have

$$\Delta = (-1)^N |A| \qquad (7)$$

As stated before, it is sufficient to show the validity of Mason's rule for a single C_1, the rest being added by superposition.

By Cramer's rule, consider the solution of the following set of equations corresponding to Eq. 1.

$$
\begin{aligned}
&a_{11}\,x_1 + \ldots + a_{1j}\,x_j + \ldots + a_{1N}\,x_N + C_1 = 0 \\
&\ldots \\
&a_{N1}\,x_1 + \ldots + a_{Nj}\,x_j + \ldots + a_{NN}\,x_N \quad\;\; = 0 \\
&\qquad\qquad\qquad\quad x_j \qquad\qquad\qquad\quad = x_j
\end{aligned}
\tag{8}
$$

In using Cramer's rule the known variable C_1 is considered unknown, and the unknown variable x_j is considered known.

Solving for C_1 we obtain the following:

$$
C_1 = \frac{\begin{vmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1N} & 0 \\ \cdots & & & & & \\ a_{N1} & \cdots & a_{Nj} & \cdots & a_{NN} & 0 \\ 0 & \ldots & 1 & \ldots & 0 & x_j \end{vmatrix}}{\begin{vmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1N} & 1 \\ \cdots & & & & & \\ a_{N1} & \cdots & a_{Nj} & \cdots & a_{NN} & 0 \\ 0 & \ldots & 1 & \ldots & 0 & 0 \end{vmatrix}}
$$

This equation can then be expanded into the following equation and then solved for x_j in terms of C_1.

$$
C_1 = \frac{x_j \cdot \begin{vmatrix} a_{11} & \cdots & a_{1N} \\ \cdots & & \\ a_{N1} & \cdots & a_{NN} \end{vmatrix}}{\begin{vmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1N} & 1 \\ \cdots & & & & & \\ a_{N1} & \cdots & a_{Nj} & \cdots & a_{NN} & 0 \\ 0 & \ldots & 1 & \ldots & 0 & -1 \end{vmatrix} + \begin{vmatrix} a_{11} & \cdots & a_{1N} \\ \cdots & & \\ a_{N1} & \cdots & a_{NN} \end{vmatrix}} = \frac{x_j \cdot |A|}{|A'| + |A|}
$$

$$
x_j = C_1 \frac{|A'| + |A|}{|A|}
$$

The matrix $|A|$ is determined from the flow-graph loops by Eq. 7. In like manner $|A'|$ may be determined from a flow graph. This flow graph for the determination of $|A'|$ may be obtained by adding a branch from x_j to C_1 with a coefficient +1. In this case Δ' contains all the terms of Δ plus those terms added because of the formation

of the new loops L_k through C_1. Then

$$x_j = C_1 \frac{-\Delta' + \Delta}{\Delta} \tag{9}$$

The term $-\Delta' + \Delta$ contains only those terms added by the formation of the new loops through C_1. Collecting all the terms that have L_k as a factor forms a term of the form $-L_k\Delta_k$, where Δ_k is calculated on the basis of those loops that do not intersect the loop L_k. The loop product L_k is equal to the k^{th} sequence product G_k from C_1 to the unknown variable x_j.

Thus, from Eq. 9 and the above discussion, the following equation is valid:

$$x_j = C_1 \frac{\sum_k L_k \Delta_k}{\Delta} \tag{10}$$

Equation 10 is based on Cramer's rule. It is also Mason's rule. Thus the general validity of the rule is established.

An interesting aspect of this proof is the general solution of a linear set of equations by a flow graph. Starting with the set of equations it is possible to proceed to the flow graph and then immediately to the solution.

In setting up the flow graph it is advisable to number the variables so that any (-1) coefficients fall on the diagonal of the coefficient matrix. (In general, a (-1) can always be made to appear on the diagonal by the use of division.) Each (-1) on the diagonal reduces the number of loops and so reduces the amount of labor in evaluating the graph.

With Eq. 7 it is possible to calculate directly from a flow graph the characteristic equation corresponding to the linear set of equations that the graph represents. A self loop having the value $-\lambda$ is attached to each node corresponding to an unknown variable. Solution of this equation for the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$ permits the calculation of the right eigenvectors λ_j: $x_{1j}, x_{2j}, \ldots, x_{Nj}$.

A particular member x_{ij} of the eigenvector is obtained by evaluating $\sum_k G_k^i \Delta_k^{ij}$ with $\lambda = \lambda_j$ from a fixed unknown variable to the particular member in question for each member of the vector. The normalized member is then

$$x_{ij} = \frac{\sum_k G_k^i \Delta_k^{ij}}{\left[\sum_{i=1}^{N} \left(\sum_k G_k^i \Delta_k^{ij}\right)^2\right]^{1/2}} \tag{11}$$

Comparison of Eqs. 2, 4, and 7 shows that we are able to calculate generally a determinant and any cofactor of the determinant directly from a flow graph.

From this proof of Mason's rule we are able to formulate a simpler rule for evaluating a flow graph containing N unknown variables. The coefficients of the flow graph are represented by b_{ji}. The value of any particular unknown variable is given by

$$x_j = \sum_{r=1}^{N} T_{jr} C_r \tag{12}$$

where

$$T_{jr} = \frac{\sum_k L_{k_{jr}} \Delta_{k_{jr}}}{\Delta} \tag{13}$$

and $L_{k_{jr}}$ is the k^{th} sequence product from the node C_r to the node x_j.

$$\Delta = \sum (-1)^q L_1 L_2 \ldots L_q (1 - b_{ii}) \ldots (1 - b_{jj}) \tag{14}$$

where each term is composed of nonintersecting loops and contains N and only N coefficients as factors. The summation is over all possible distinct combinations. L refers to loops involving two or more coefficients; q is the number of nonintersecting loops of two or more coefficients and $\Delta_{k_{jr}}$ is modified Δ where the loops are restricted to those which do not intersect the loop associated with the sequence product $L_{k_{jr}}$.

The proof of this equation follows immediately from

$$BX + C = X \qquad B = \left[b_{ji}\right] \tag{15}$$

which corresponds to Eq. 3, and from Eqs. 6 and 10.

The main advantage of this rule is that it substantially reduces the number of terms in the calculation of Δ and Δ_k when there are self loops in the flow-graph system.

C. S. Lorens

F. PROPERTIES OF SECOND-ORDER AUTOCORRELATION FUNCTIONS

In the Quarterly Progress Report of October 15, 1955, page 52, the Fourier transform of the second-order autocorrelation function of an aperiodic function, $f_1(t)$, was found to be

$$\Phi^*(\omega_1, \omega_2) = 2\pi F(\omega_1) F(\omega_2) \overline{F(\omega_1 + \omega_2)} \tag{1}$$

If $\Phi^*(\omega_1, \omega_2)$ is integrated with respect to one of the variables, let us say ω_2, this integral is the cross-spectral-density function of the function $f_1(t)$ and the square of the

original function $f_1^2(t)$. This can be shown by considering

$$\int_{-\infty}^{\infty} \Phi^*(\omega_1, \omega_2)\, d\omega_2 = 2\pi\, F(\omega_1) \int_{-\infty}^{\infty} F(\omega_2)\, F(-\omega_1 - \omega_2)\, d\omega_2 \tag{2}$$

If we denote the transform of $f_1^2(t)$ by $G(\omega_1)$, we obtain

$$\overline{G(\omega_1)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} f_1^2(t)\, e^{j\omega_1 t}\, dt = \int_{-\infty}^{\infty} F(\omega_2)\, F(-\omega_1 - \omega_2)\, d\omega_2 \tag{3}$$

then Eq. 2 becomes

$$\int_{-\infty}^{\infty} \Phi^*(\omega_1, \omega_2)\, d\omega_2 = 2\pi\, F(\omega_1)\, \overline{G(\omega_1)} \tag{4}$$

In the case of a random function, $f_1(t)$, the transform of the second-order autocorrelation function, $\phi_{111}^*(\tau_1, \tau_2)$, of $f_1(t)$ is

$$\Phi^*(\omega_1, \omega_2) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_{111}^*(\tau_1, \tau_2)\, e^{-j(\omega_1\tau_1 + \omega_2\tau_2)}\, d\tau_1\, d\tau_2$$

$$= \lim_{T\to\infty} \frac{\pi}{T} \left[F_T(\omega_1)\, F_T(\omega_2)\, \overline{F_T(\omega_1 + \omega_2)} \right] \tag{5}$$

where $F_T(\omega)$ is the transform of $f_{1T}(t)$, and $f_{1T}(t) = f_1(t)$ when $-T \leq t \leq T$ and zero elsewhere. We can proceed as in the aperiodic case and obtain the same result, but we take the product of $\Phi^*(\omega_1, \omega_2)$ and $e^{j\omega_2\tau_2}$ and integrate it with respect to ω_2:

$$\int_{-\infty}^{\infty} \Phi^*(\omega_1, \omega_2)\, e^{j\omega_2\tau_2}\, d\omega_2 = \lim_{T\to\infty} \frac{\pi}{T} \left[F_T(\omega_1) \int_{-\infty}^{\infty} F_T(\omega_2)\, F_T(-\omega_1 - \omega_2)\, e^{j\omega_2\tau_2}\, d\omega_2 \right] \tag{6}$$

If the transform of $f_{1T}(t)\, f_{1T}(t + \tau_2)$ is denoted by $G_T(\omega_1, \tau_2)$ for a fixed τ_2, then we have

$$\overline{G_T(\omega_1, \tau_2)} = \int_{-\infty}^{\infty} F_T(\omega_2)\, F_T(-\omega_1 - \omega_2)\, e^{j\omega_2\tau_2}\, d\omega_2 \tag{7}$$

Equation 6 now becomes

$$\int_{-\infty}^{\infty} \Phi^*(\omega_1, \omega_2)\, e^{j\omega_2\tau_2}\, d\omega_2 = \lim_{T\to\infty} \frac{\pi}{T}\left[F_T(\omega_1)\, \overline{G_T(\omega_1, \tau_2)}\right] \tag{8}$$

Equation 8 shows that when $\tau_2 = 0$, the integral of $\Phi^*(\omega_1, \omega_2)$ with respect to ω_2 is the cross-spectral-density function of $f_1^2(t)$ and $f_1(t)$.

If the crosscorrelation function of $g_1(t, \tau_2) = f_1(t)\, f_1(t + \tau_2)$, where τ_2 is fixed, and $f_1(t)$ is denoted by $\phi_{g_1(t,\tau_2), f_1(t)}(\tau_1)$, then, since Eq. 8 gives the cross-spectral-density function of $g_1(t, \tau_2)$ and $f_1(t)$, we have,

$$\int_{-\infty}^{\infty} \Phi^*(\omega_1, \omega_2)\, e^{j\omega_2\tau_2}\, d\omega_2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_{g_1(t,\tau_2), f_1(t)}(\tau_1)\, e^{-j\omega_1\tau_1}\, d\tau_1$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi^*_{111}(\tau_1, \tau_2)\, e^{-j\omega_1\tau_1}\, d\tau_1 \tag{9}$$

This relation shows that it may be convenient to interpret the second-order autocorrelation function of $f_1(t)$ as a crosscorrelation function of $g_1(t, \tau_2) = f_1(t)\, f_1(t + \tau_2)$ and $f_1(t)$.

J. Y. Hayase

G. A COMPENSATOR FOR THE DIGITAL ELECTRONIC CORRELATOR

The pulse-comparing circuit used in the compensator scheme described in the Quarterly Progress Report of July 15, 1955, pages 48-49, has been tested; constant-width pulses were used. The circuit that will be used in the correlator, however, will have random-width pulses as inputs. Figure XI-9 gives a plot of $e_{c_{final}}$ as a function of the pulse difference (P-M) with (P+M) held constant; this will be the case in the correlator. The slight difference between the experimental and theoretical curves can probably be accounted for by leakage resistance in the capacitor and the input resistance of the measuring instrument.

This circuit may be useful in many other applications for measuring pulse differences. For example, it might well be used as a pulse-length measuring device in which a pulse can be compared with a standard pulse obtained from a laboratory instrument.

One arrangement for such a device is shown in Fig. XI-10. The input pulse A is fed into the pulse comparator activating switch S_1. The signal B from the standard frequency source is consecutively clipped and amplified into pulses C, which trigger a flip-flop FF_2 to create the standard pulse D, which activates switch S_2. To avoid error, the average repetition rate of the two pulses (A and D) must remain equal; in other words, there must be one standard pulse for each input pulse. The gate pulse E, therefore,

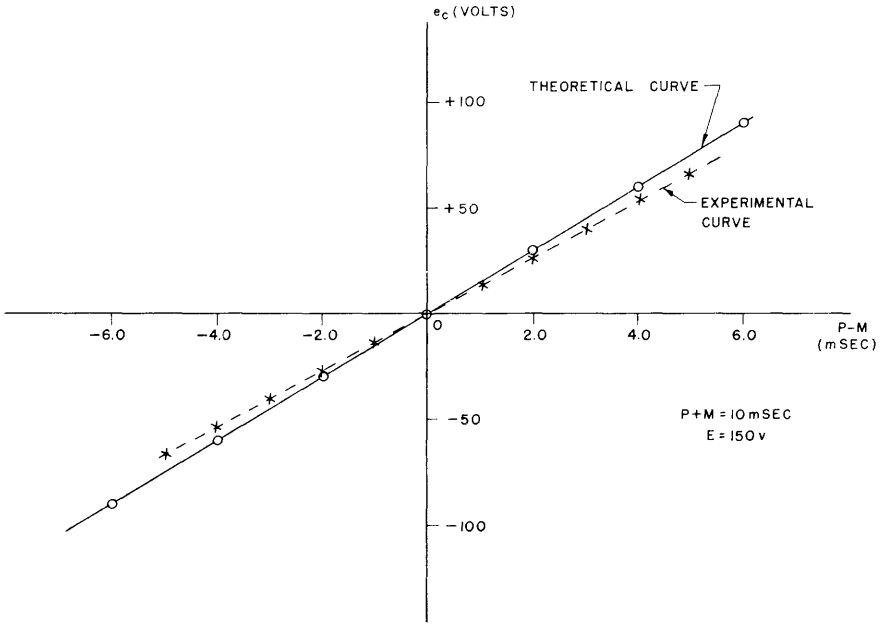

Fig. XI-9. Voltage output versus pulse-length difference.

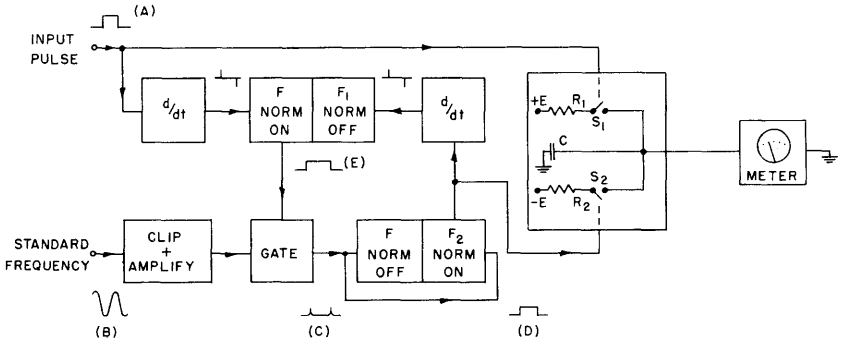

Fig. XI-10. Block diagram for measuring pulse lengths.

allows only two pulses C per period of the input pulse to reach the flip-flop. The normal states shown on the flip-flops are the states that they will assume with no input pulse.

When this method is used there are two measurement procedures that can be followed. One is to calibrate the voltmeter in time units for specified standard pulse lengths. Another is to vary the standard pulse length for a null indication, thus finding the pulse length from the period standard frequency source. In either case the standard pulse may be effectively multiplied by a constant factor by changing the ratio of the resistors in the pulse-comparing circuit.

K. L. Jordan

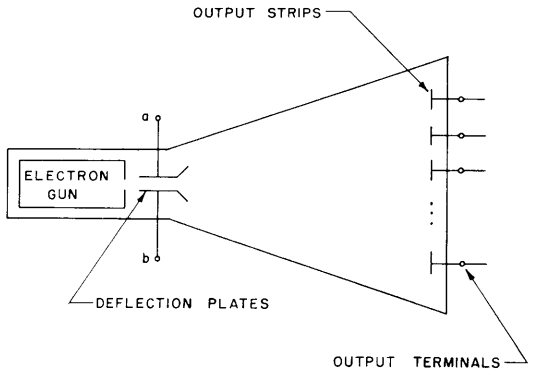

Fig. XI-11. A level selector tube.

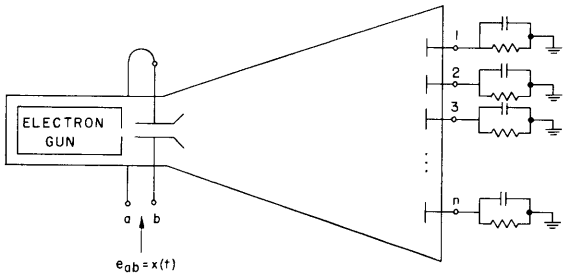

Fig. XI-12. Application of the level selector tube to the measurement of the first probability density of a time function x(t).

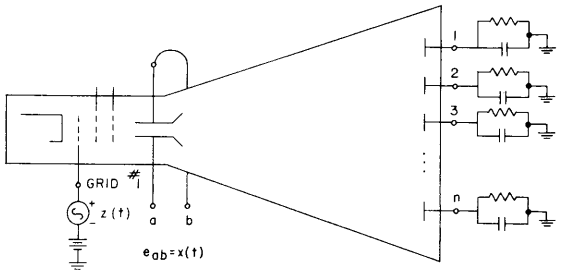

Fig. XI-13. Application of the level selector tube to the determination of optimum no-memory filters.

H. A LEVEL SELECTOR TUBE

Figure XI-11 shows a level selector tube which, among its many uses, is of interest in our work as a device for the measurement of probability densities and as a circuit element in the apparatus for the determination and synthesis of optimum nonlinear filters in accordance with the theory stated in the Quarterly Progress Report of October 15, 1955.

The input time function is applied across the terminals a-b. The value of the input at any instant of time determines which of the output terminals will collect the beam current. The use of the tube for probability density measurements is indicated in Fig. XI-12. The time function x(t) to be analyzed is fed to the deflection plates. The amplitude probability density is proportional to the voltages at the output terminals 1 through n.

In Fig. XI-13 the level selector tube is used to determine the coefficients for an optimum no-memory filter according to the theory previously discussed. In Fig. XI-13 x(t) is an ensemble member of the filter input and z(t) is the corresponding ensemble member of the desired filter output. The voltages at the n output terminals are measured with the tube inputs as shown. Then the voltage source z(t) is set to zero and the n output voltages are again measured. The j^{th} optimum filter coefficient is proportional to the difference of the two measured voltages at output terminal j divided by the second measured voltage.

With the cooperation of Mr. P. Youtz the first model of the level selector tube has been constructed in the tube laboratory of the Barta Building. This model had only four output terminals and enabled us to study the optimum output strip separation and structure. A study of the secondary emission and the maximum current attainable from the output strips has led to the design of the second model, which will make use of the secondary emission properties of beryllium-copper output strips.

The two suggestions for the use of the tube, illustrated by Figs. XI-12 and -13, do not require a multiple-strip tube. The operations indicated could be carried out sequentially with a single-strip tube. However, the multiple-strip tube is useful in general nonlinear filter synthesis and, of course, it is more convenient in the examples cited above. With the addition of another set of deflection plates and a masking aperture between the two sets the tube can be conveniently used for the measurement of second-probability densities.

We wish to acknowledge the helpful suggestions of Prof. L. D. Smullin, of this Laboratory, and of Dr. C. W. Mueller, of the David Sarnoff Research Center, concerning some problems associated with the tube.

A. G. Bose