

## VIII. STATISTICAL COMMUNICATION THEORY

Prof. Y. W. Lee  
 Prof. A. G. Bose (Absent)  
 M. B. Brilliant  
 D. A. Chesler

D. A. George  
 J. Y. Hayase  
 I. M. Jacobs

A. H. Nuttall  
 R. E. Wernikoff  
 H. E. White  
 G. D. Zames

### A. ANALYTIC NONLINEAR SYSTEMS

#### 1. Definitions, Norms, Transforms

An analytic system (1) is a device whose output  $g(t)$  can be expressed in terms of its input  $f(t)$  as

$$\begin{aligned} g(t) &= h_0 + \int_{-\infty}^{\infty} h_1(\tau) f(t-\tau) d\tau + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\tau_1, \tau_2) f(t-\tau_1) f(t-\tau_2) d\tau_1 d\tau_2 + \dots \\ &= \sum_{n=0}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_n(\tau_1, \dots, \tau_n) \prod_{i=1}^n f(t-\tau_i) d\tau_i \end{aligned} \quad (1)$$

The functions  $h_n(\tau_1, \dots, \tau_n)$  will be called the system functions. We shall also make use of the system transforms

$$H_n(\omega_1, \dots, \omega_n) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_n(\tau_1, \dots, \tau_n) \prod_{i=1}^n e^{-j\omega_i \tau_i} d\tau_i \quad (2)$$

The system of Eq. 1 will be denoted by  $\mathcal{H}$ ; thus, we write  $g = \mathcal{H}f$ , or  $g(t) = \mathcal{H}f(t)$ .

The norms of the system functions will be defined in the  $L_1$  sense:

$$\|h_n\| = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |h_n(\tau_1, \dots, \tau_n)| d\tau_1 \dots d\tau_n \quad (3)$$

The norm of the system  $\mathcal{H}$  is a function of a positive real variable  $x$ , and is defined by the power series

$$\|\mathcal{H}\|(x) = \sum_{n=0}^{\infty} \|h_n\| x^n \quad (4)$$

The radius of convergence of this series will be called the radius of convergence of the system, and will be denoted by  $\rho(\mathcal{H})$ . It is clear that if the input is suitably bounded,  $|f(t)| \leq M < \rho(\mathcal{H})$ , then the output will be defined and bounded,  $|\mathcal{H}f(t)| \leq \|\mathcal{H}\|(M)$ . A system will be called analytic only if its norm exists and its radius of convergence is not zero.

The notation of script letters for systems, lower-case letters for system functions,

and capital letters for transforms will be used consistently.

## 2. Simple Networks: Algebra of Systems

It will be convenient to have an algebra of systems in which the operations of addition and multiplication correspond to the elementary ways of combining systems. It is also convenient to have a method of computing the results of these operations. We shall find formulas, not only for the system functions of the resulting systems, but also for bounds on their norms and radii of convergence.

Figure VIII-1 illustrates the sum of two systems. The defining equation is

$$(\mathcal{H} + \mathcal{K})f = \mathcal{H}f + \mathcal{K}f \quad (5)$$

For its system functions, we immediately obtain

$$(h+k)_n(\tau_1, \dots, \tau_n) = h_n(\tau_1, \dots, \tau_n) + k_n(\tau_1, \dots, \tau_n) \quad (6)$$

and for its system transforms,

$$(H+K)_n(\omega_1, \dots, \omega_n) = H_n(\omega_1, \dots, \omega_n) + K_n(\omega_1, \dots, \omega_n) \quad (7)$$

We also immediately obtain for the norm,

$$\|\mathcal{H} + \mathcal{K}\|(x) \leq \|\mathcal{H}\|(x) + \|\mathcal{K}\|(x) \quad (8)$$

and, for the radius of convergence,

$$\rho(\mathcal{H} + \mathcal{K}) \geq \min\{\rho(\mathcal{H}), \rho(\mathcal{K})\} \quad (9)$$

Addition of systems is evidently commutative and associative.

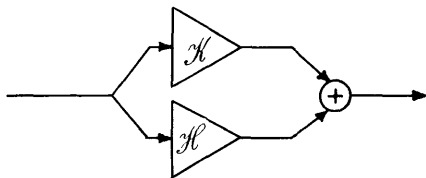


Fig. VIII-1. Parallel combination  
 $\mathcal{H} + \mathcal{K}$ .

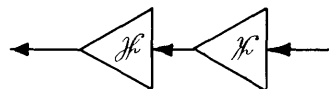


Fig. VIII-2. Cascade combination  
 $\mathcal{H} \mathcal{K}$ .

The product of two systems is illustrated in Fig. VIII-2. The defining equation is

$$(\mathcal{H} \mathcal{K})f = \mathcal{H}(\mathcal{K}f) \quad (10)$$

(VIII. STATISTICAL COMMUNICATION THEORY)

We have found (1) that

$$(\text{hk})_p(\sigma_1, \dots, \sigma_p) = \sum_{n=0}^{\infty} \sum_{\substack{n \\ \sum_{i=1}^n m_i = p}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_n(\tau_1, \dots, \tau_n) \prod_{i=1}^n k_{m_i}(\sigma_{r_i+1} - \tau_i, \dots, \sigma_{r_i+m_i} - \tau_i) d\tau_i \quad (11)$$

where  $r_i = r_{i-1} + m_{i-1}$  and  $r_1 = 0$ . In terms of transforms,

$$(\text{HK})_p(\omega_1, \dots, \omega_n) = \sum_{n=0}^{\infty} \sum_{\substack{n \\ \sum_{i=1}^n m_i = p}} H_n(\Omega_1, \dots, \Omega_n) \prod_{i=1}^n K_{m_i}(\omega_{r_i+1}, \dots, \omega_{r_i+m_i}) \quad (12a)$$

where

$$\Omega_i = \sum_{j=1}^{m_i} \omega_{r_i+j} \quad (12b)$$

To determine a bound on the norm of the product, we note that

$$\begin{aligned} \|\mathcal{H}\mathcal{K}\|(\|\mathcal{X}\|(\mathbf{x})) &= \sum_{n=0}^{\infty} \|h_n\| \left( \sum_{m=0}^{\infty} \|k_m\| x^m \right)^n \\ &= \sum_{p=0}^{\infty} x^p \sum_{n=0}^{\infty} \sum_{\substack{n \\ \sum_{i=1}^n m_i = p}} \|h_n\| \prod_{i=1}^n \|k_{m_i}\| \end{aligned} \quad (13)$$

and, comparing this equation with Eq. 11, using a generalization of the theorem (2) on the  $L_1$  norm of a convolution, we find that

$$\|\mathcal{H}\mathcal{K}\|(\mathbf{x}) \leq \|\mathcal{H}\|(\|\mathcal{X}\|(\mathbf{x})) \quad (14)$$

The series converges if  $\|\mathcal{X}\|(\mathbf{x}) < \rho(\mathcal{H})$ ; hence  $\rho(\mathcal{H}\mathcal{K})$  is not less than the least upper bound of all  $\mathbf{x} < \rho(\mathcal{X})$  that satisfy this condition.

$$\rho(\mathcal{H}\mathcal{K}) \geq \text{least upper bound } \{\mathbf{x} \mid \mathbf{x} < \rho(\mathcal{X}), \|\mathcal{X}\|(\mathbf{x}) < \rho(\mathcal{H})\} \quad (15)$$

Multiplication is associative, but not commutative,  $\mathcal{H}\mathcal{K} \neq \mathcal{K}\mathcal{H}$ . It is not left-distributive over addition,  $\mathcal{L}(\mathcal{H}+\mathcal{K}) \neq \mathcal{L}\mathcal{H} + \mathcal{L}\mathcal{K}$ , but the definition of addition (Eq. 5) shows that it is right-distributive,  $(\mathcal{H}+\mathcal{K})\mathcal{L} = \mathcal{H}\mathcal{L} + \mathcal{K}\mathcal{L}$ .

(VIII. STATISTICAL COMMUNICATION THEORY)

The product of two analytic systems does not always exist as an analytic system. The norm of  $\mathcal{X}$ , from Eq. 4, has  $|k_o|$  as its minimum value. If  $|k_o| \geq \rho(\mathcal{H})$ , then there is no  $x$  satisfying  $||\mathcal{X}|| (x) < \rho(\mathcal{H})$ , and Eq. 15 then shows that  $\rho(\mathcal{H}\mathcal{X})$  may be zero. There are two ways of eliminating this possibility in advance: arrange, by an appropriate choice of reference levels, to deal only with systems with no constant term; or, deal only with systems with infinite radii of convergence.

Table VIII-1. All Ordered Sets of  $n$  Numbers Whose Sum is  $p$ .

$n \backslash p$	0	1	2	3
0	empty set	0	00	000
1	---	1	10 01	100 010 001
2	---	2	11 02 20	110 101 011 200 020 002

In Eqs. 11 and 12a, the second  $\Sigma$  is a summation over all ordered sets of  $n$  numbers  $m_i$  whose sum is  $p$ . It is convenient to have a table of these sets; Table VIII-1 is a partial table. The most important feature of this table is not that the number of these sets becomes very large for large  $n$  and  $p$  [equal to the binomial coefficient  $\binom{n+p-1}{n-1}$ ], but that for a given  $p$  there are terms for every  $n$ . For example, the constant term in Eq. 12a is

$$(HK)_o = H_o + H_1(0) K_o + H_2(0, 0) K_o^2 + H_3(0, 0, 0) K_o^3 + \dots \quad (16)$$

Therefore, the system functions cannot be computed exactly. Two remedies are possible: one is to consider only systems with a finite number of terms (which, incidentally, have infinite radii of convergence); the other is to consider only systems with no constant terms. Table VIII-2 gives the ordered sets of  $m_i$  for systems without constant terms [here the number of sets for given  $n$  and  $p$  is  $\binom{p-1}{n-1}$ ]; for a given  $p$  we need consider only  $n \leq p$ . The first three terms then are

(VIII. STATISTICAL COMMUNICATION THEORY)

$$L_1(\omega) = H_1(\omega)K_1(\omega)$$

$$L_2(\omega_1, \omega_2) = H_1(\omega_1 + \omega_2)K_2(\omega_1, \omega_2) + H_2(\omega_1, \omega_2)K_1(\omega_1)K_1(\omega_2)$$

$$L_3(\omega_1, \omega_2, \omega_3) = H_1(\omega_1 + \omega_2 + \omega_3)K_3(\omega_1, \omega_2, \omega_3) + H_2(\omega_1 + \omega_2, \omega_3)K_2(\omega_1, \omega_2)K_1(\omega_3) \\ + H_2(\omega_1, \omega_2 + \omega_3)K_1(\omega_1)K_2(\omega_2, \omega_3) + H_3(\omega_1, \omega_2, \omega_3)K_1(\omega_1)K_1(\omega_2)K_1(\omega_3)$$

(17)

We complete our algebra with two more definitions. The identity operator  $\mathcal{I}$  is defined by  $\mathcal{I}f = f$  for all  $f$ ; its first system function is  $\delta(\tau)$  and all others are zero. The negative of a system is defined by  $(-\mathcal{H})f = -(\mathcal{H}f)$ ; its system functions are the negatives of those of the original system.

Table VIII-2. All Ordered Sets of  $n$  Nonzero Numbers Whose Sum is  $p$ .

$p \backslash n$	1	2	3	4
1	1	---	---	---
2	2	11	---	---
3	3	21 12	111	---
4	4	22 13 31	211 121 112	1111

3. Feedback Networks

The most general feedback loop involves two systems, one in the forward path and one in the feedback path, as in Fig. VIII-3. As shown, this general loop can be transformed into a simpler loop in cascade with another system, so that we lose no generality by considering the elementary loop of Fig. VIII-4. We further simplify the network

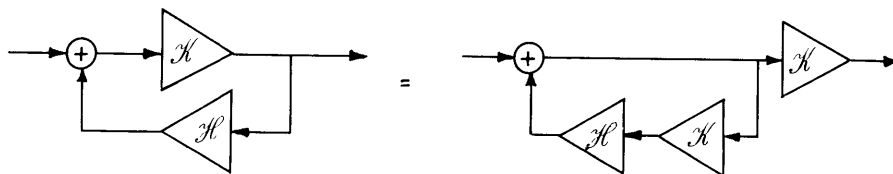


Fig. VIII-3. Reduction of the general feedback loop.

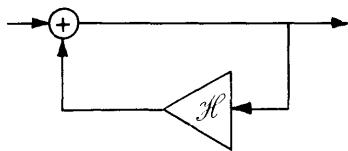


Fig. VIII-4. Elementary feedback loop.

by noting that if the system  $\mathcal{K}$  has a constant term, this is simply a constant added to the input, and we may suppose that this constant is added in another system in cascade with the feedback network.

We therefore consider the elementary feedback loop of Fig. VIII-4, in which  $h_0 = 0$ . Let the composite system be denoted by  $\mathcal{K}$ . Then, if the input is  $f$ , the output is  $\mathcal{K}f$ , and we have

$$\begin{aligned}\mathcal{K}f &= f + \mathcal{K}\mathcal{K}f \\ \mathcal{K} &= \mathcal{I} + \mathcal{K}\mathcal{K}\end{aligned}\tag{18}$$

A solution for  $\mathcal{K}$  can be obtained formally by noting that if two systems are equal their system functions are equal. For the constant terms in Eq. 18 we have (cf. Eq. 16)

$$K_0 = H_1(0)K_0 + H_2(0,0)K_0^2 + H_3(0,0,0)K_0^3 + \dots\tag{19}$$

Equation 19 may have many solutions, but it will always have the solution  $K_0 = 0$ , and this is the solution that must be chosen if zero initial conditions are assumed. With  $K_0 = 0$ , we can use Table VIII-2 and Eq. 17. For the first-degree terms, we have

$$\begin{aligned}K_1(\omega) &= 1 + H_1(\omega)K_1(\omega) \\ K_1(\omega) &= \frac{1}{1 - H_1(\omega)}\end{aligned}\tag{20}$$

For the second-degree terms, we have

$$\begin{aligned}K_2(\omega_1, \omega_2) &= H_1(\omega_1 + \omega_2)K_2(\omega_1, \omega_2) + H_2(\omega_1, \omega_2)K_1(\omega_1)K_1(\omega_2) \\ K_2(\omega_1, \omega_2) &= \frac{H_2(\omega_1, \omega_2)K_1(\omega_1)K_1(\omega_2)}{1 + H_1(\omega_1 + \omega_2)} \\ &= \frac{H_2(\omega_1, \omega_2)}{[1 + H_1(\omega_1 + \omega_2)][1 + H_1(\omega_1)][1 + H_1(\omega_2)]}\end{aligned}\tag{21}$$

and it is evident that we can continue this process to compute any desired number of system transforms of  $\mathcal{K}$ .

This formal solution is based on the assumption that there is an analytic  $\mathcal{K}$  that satisfies Eq. 18. It is, therefore, correct if and only if such a  $\mathcal{K}$  exists. It will be

(VIII. STATISTICAL COMMUNICATION THEORY)

shown that this assumption is correct if  $k_1(\tau)$ , which can be computed from its transform given in Eq. 20, is absolutely integrable – that is, if  $\|k_1\|$  is finite. This is, essentially, the condition that the linearized system be stable.

Bounds on  $\rho(\mathcal{K})$  and  $\|\mathcal{K}\|(x)$  can be obtained if we separate  $\mathcal{K}$  into two parts, the linear part  $\mathcal{K}'$  and a part  $\mathcal{K}'' = \mathcal{K} - \mathcal{K}'$  that has no linear term. Then we have

$$\rho(\mathcal{K}) \geq \text{least upper bound}_{x < \rho(\mathcal{K})} \left\{ \frac{x}{\|k_1\|} - \|\mathcal{K}''\|(x) \right\} \quad (22)$$

and, for all  $x$  less than the radius of convergence guaranteed by Eq. 22, we have

$$\|\mathcal{K}\|(x) \leq y(x) \quad (23a)$$

where  $y(x)$  is the smallest positive  $y$  that satisfies

$$x = \frac{y}{\|k_1\|} - \|\mathcal{K}''\|(y) \quad (23b)$$

Before presenting the proof of these conclusions, we note that every bound that has been given for the norm and radius of convergence of a composite system remains valid if every norm that appears on the right-hand side is replaced by its upper bound and every radius of convergence that appears on the right-hand side is replaced by its lower bound.

Now consider Eq. 18. Assume that  $k_1(\tau)$  has been computed and that  $\|k_1\|$  is finite. Define  $\mathcal{K}'$  by specifying its first system function as  $k_1(\tau)$  and its other system functions as zero. Then  $\mathcal{K}'$  is linear,  $\rho(\mathcal{K}') = \infty$ ,  $\|\mathcal{K}'\|(x) = \|k_1\| x$ , and

$$\mathcal{K}' = \mathcal{I} + \mathcal{K}'\mathcal{K}' \quad (24)$$

Define

$$\mathcal{Q} = \mathcal{K}''\mathcal{K}' \quad (25)$$

$\mathcal{Q}$  has no linear part, since  $\mathcal{K}''$  has none. Define  $\mathcal{P}$  as the solution of

$$\mathcal{P} = \mathcal{I} + \mathcal{Q}\mathcal{P} \quad (26)$$

Now

$$\begin{aligned} \mathcal{K}'\mathcal{P} &= (\mathcal{I} + \mathcal{K}'\mathcal{K}')\mathcal{P} \\ &= \mathcal{P} + \mathcal{K}'\mathcal{K}'\mathcal{P} \\ &= \mathcal{I} + \mathcal{Q}\mathcal{P} + \mathcal{K}'\mathcal{K}'\mathcal{P} \\ &= \mathcal{I} + \mathcal{K}''\mathcal{K}'\mathcal{P} + \mathcal{K}'\mathcal{K}'\mathcal{P} \\ &= \mathcal{I} + \mathcal{K}\mathcal{K}'\mathcal{P} \end{aligned} \quad (27)$$

hence  $\mathcal{K}'\mathcal{P}$  satisfies the equation defining  $\mathcal{K}$ . Therefore,

$$\mathcal{K} = \mathcal{K}'\mathcal{P} \quad (28)$$

We now prove that  $\mathcal{P}$  is analytic. First, note that Eq. 26 can be solved formally just as we solved Eq. 18. Next, consider the equation in the positive real variables  $x$  and  $y$

$$y = x + \|\mathcal{Q}\|(y) \quad (29)$$

where  $y$  is to be determined as a function of  $x$ . This can be considered as the equation of a no-memory feedback loop analogous to that of Eq. 26. Equation 13 can be used as a formula for cascading two no-memory analytic systems, and as such it is analogous to Eq. 12. Therefore, we can solve Eq. 29 formally and obtain a solution analogous to that of Eq. 26. Using, again, the generalized theorem on the  $L_1$  norm of a convolution, we conclude that the series for  $\mathcal{P}$  is dominated by the series for  $y(x)$ . Hence  $\mathcal{P}$  is analytic if  $y(x)$  is an analytic function,  $\rho(\mathcal{P})$  is at least equal to the radius of convergence of the Taylor expansion of  $y(x)$  about zero, and  $\|\mathcal{P}\|(x)$  is not greater than  $y(x)$ .

Equation 29 can be solved for  $x(y)$ :

$$\begin{aligned} x &= y - \|\mathcal{Q}\|(y) \\ &= y - \sum_{n=2}^{\infty} \|q_n\| y^n \end{aligned} \quad (30)$$

which is valid for  $y < \rho(\mathcal{Q})$ . We also have

$$\frac{dx}{dy} = 1 - \sum_{n=2}^{\infty} n \|q_n\| y^{n-1} \quad (31)$$

Then  $dx/dy = 1$  when  $y = 0$ , and decreases for increasing  $y$ . Let  $A$  be the least upper bound of all  $y$  for which  $dx/dy$  is positive; then  $x(y)$  is an increasing function if and only if  $y < A$ . Let  $B$  be the least upper bound of all  $x(y)$  for  $y < A$ ; then  $B$  is also the least upper bound of all  $x(y)$  for  $y < \rho(\mathcal{Q})$ . Then for all  $x < B$ ,  $y(x)$  is an increasing function,  $y(x) < A$ .

Now extend  $x(y)$  analytically into the complex plane and consider  $x$  and  $y$  as complex variables. For  $|y| < A$ , we find that  $|dx/dy| > 0$ ; hence  $dy/dx$ , the reciprocal of  $dx/dy$ , exists and  $y(x)$  is analytic. The circle  $|y| = r < A$  is mapped into a closed curve in the  $x$ -plane on which the minimum value of  $|x|$  is  $x(r)$ . Hence, for  $|x| < B$ ,  $y(x)$  is analytic. We conclude that  $y(x)$  can be expanded about zero in a Taylor series with  $B$  as its radius of convergence.

Therefore,



(VIII. STATISTICAL COMMUNICATION THEORY)

$$\begin{aligned}
 \rho(\mathcal{P}) &\geq B = \text{least upper bound } x(y) \\
 &\quad y < \rho(\mathcal{Q}) \\
 &\geq \text{least upper bound } \{y - \|\mathcal{Q}\|(y)\} \\
 &\quad y < \rho(\mathcal{Q})
 \end{aligned} \tag{32}$$

But, from Eqs. 25 and 14, we have

$$\begin{aligned}
 \|\mathcal{Q}\|(y) &\leq \|\mathcal{H}''\| \left( \|\mathcal{X}'\|(y) \right) \\
 &\leq \|\mathcal{H}''\| \left( \|\mathbf{k}_1\| y \right)
 \end{aligned} \tag{33}$$

and from Eq. 15, we have

$$\begin{aligned}
 \rho(\mathcal{Q}) &\geq \text{least upper bound } \{x \mid \|\mathbf{k}_1\| x < \rho(\mathcal{H}'')\} \\
 &\geq \frac{\rho(\mathcal{H}'')}{\|\mathbf{k}_1\|} = \frac{\rho(\mathcal{H})}{\|\mathbf{k}_1\|}
 \end{aligned} \tag{34}$$

Hence

$$\begin{aligned}
 \rho(\mathcal{P}) &\geq \text{least upper bound } \left\{ y - \|\mathcal{H}''\| \left( \|\mathbf{k}_1\| y \right) \right\} \\
 &\quad y < \frac{\rho(\mathcal{H})}{\|\mathbf{k}_1\|}
 \end{aligned} \tag{35}$$

Further, from Eqs. 28 and 15,

$$\begin{aligned}
 \rho(\mathcal{X}) &\geq \text{least upper bound } \{x \mid x < \rho(\mathcal{P})\} \\
 &\geq \rho(\mathcal{P})
 \end{aligned} \tag{36}$$

and Eq. 22 is obtained by a change of variable.

To obtain a bound on  $\|\mathcal{X}\|(x)$ , we have that for  $x < B$ ,

$$\|\mathcal{P}\|(x) \leq y(x) \tag{37}$$

where  $y(x)$  is defined by Eq. 29. Note that we want the value of  $y$  that is less than  $A$ ; there may also be one greater than  $A$ . Equations 28 and 14 give

$$\begin{aligned}
 \|\mathcal{X}\|(x) &\leq \|\mathbf{k}_1\| \|\mathcal{P}\|(x) \\
 &\leq \|\mathbf{k}_1\| y(x)
 \end{aligned} \tag{38}$$

Now, Eq. 30 defines  $x(y)$  as an increasing function for the range to which we are restricted. If in that equation we replace  $\|\mathcal{Q}\|(y)$  by  $\|\mathcal{H}''\| \left( \|\mathbf{k}_1\| y \right)$ , which is not only smaller but has a smaller derivative, the new  $x(y)$  will be smaller but will still be an increasing function; hence the new  $y(x)$  will be larger. Therefore, Eq. 37 will still hold if we define  $y(x)$  by

$$\mathbf{x} = \mathbf{y} - \|\mathcal{H}^n\| (\|k_1\| \mathbf{y}) \quad (39)$$

Then we obtain Eq. 23 by a change of variable.

M. B. Brilliant

#### References

1. M. B. Brilliant, Analytic nonlinear systems, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., Jan. 15, 1957, pp. 54-61.
2. S. Bochner and K. Chandrasekharan, Fourier Transforms (Annals of Mathematics Studies No. 19, Princeton University Press, Princeton, N.J., 1949), pp. 5-7.

#### B. INVARIANCE OF CORRELATION FUNCTIONS UNDER NONLINEAR TRANSFORMATIONS

Previous work (1) on the invariance of correlation functions under nonlinear transformation was directed at obtaining sufficient conditions for the invariance property to hold for a class of nonlinear devices. It was shown that certain relations among cross-moments were sufficient, although not necessary.

A different method of attack on the problem has resulted in both a necessary and sufficient condition to be imposed upon the input statistics in order for the invariance property to hold. For completeness, we restate the invariance property: If, in the system of Fig. VIII-5, the crosscorrelation function of the two outputs is identical to the crosscorrelation function of the two inputs for every nonlinear no-memory device (for which the output crosscorrelation function exists), except for a scale factor  $C_f$  dependent on the particular nonlinear device, the invariance property is said to hold for that particular pair of inputs.

Let us define the input crosscorrelation function as

$$\phi(\tau) = \iint x_1 x_2 p(x_1, x_2; \tau) dx_1 dx_2 \quad (1)$$

and the output crosscorrelation function as

$$\Phi_f(\tau) = \iint x_1 f(x_2) p(x_1, x_2; \tau) dx_1 dx_2 \quad (2)$$

where  $p(x_1, x_2; \tau)$  is the joint probability density function of the system inputs, assumed stationary for the present. The nonlinear device is assumed time-invariant. (All integrals are over the whole range of the variables.)

Realizing that the output crosscorrelation function  $\Phi_f(\tau)$  exists, for all  $\tau$ , only for certain nonlinear devices, let us denote this class of allowable nonlinear devices by P. This class P depends only upon the joint probability density function  $p(x_1, x_2; \tau)$ .

(VIII. STATISTICAL COMMUNICATION THEORY)

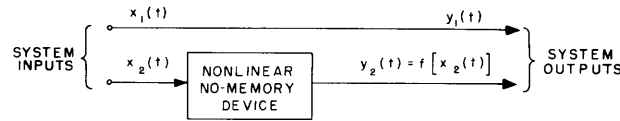


Fig. VIII-5. Nonlinear no-memory transformation of the input process.

Assuming that  $\phi(\tau)$  exists as a finite-valued Lebesgue double integral for all  $\tau$  in  $S$  (Eq. 1), where  $S$  is an arbitrary set of the real line, and defining the function  $g$  as

$$g(x_2, \tau) = \int x_1 p(x_1, x_2; \tau) dx_1 \quad (3)$$

we can now state the main result of this report:

$$g(x_2, \tau) = h(x_2) \phi(\tau) \text{ almost everywhere in } x_2, \text{ for } \tau \text{ in } S \quad (4)$$

is a necessary and sufficient condition for the invariance property to hold for  $\tau$  in  $S$ . The function  $h(x_2)$  is a function only of  $x_2$ , and not of  $\tau$ . Satisfaction of Eq. 4 will be called separability (of the  $g$  function). It is seen that the function  $g$  of two variables breaks into a product of two functions, each with one variable as its argument. Because of its length, the proof of this theorem is not presented.

Sufficiency of Eq. 4 has been demonstrated by Luce (2). However, necessity was shown only under very restrictive conditions on  $\phi(\tau)$  and  $g(x_2, \tau)$ .

It is to be noted from Eq. 3 that  $g$  depends only upon the joint probability density function of the system inputs. Since its determination requires only one integration, the satisfaction of Eq. 4 is easily determined. Equation 4 can be shown to generalize previous results (3, 4) on the invariance property.

For the special, but important, case in which  $x_1(t) = x_2(t)$  in Fig. VIII-5, it can be shown, that if  $g$  is separable, then

$$g(x_2, \tau) = \sigma^{-2} x_2 p(x_2) \phi(\tau) \quad (5)$$

where  $p(x_2)$  is the first-order probability density function of  $x_2(t)$ . That is,  $h(x_2)$  in Eq. 4 can be evaluated very simply. It then follows that the constant  $C_f$  that relates output crosscorrelation function to input crosscorrelation function can be determined as

$$C_f = \sigma^{-2} \int x_2 f(x_2) p(x_2) dx_2 = \overline{\sigma^{-2} x_2(t) f[x_2(t)]} \quad (6)$$

No such simple formula as Eq. 5 holds for  $g$  when  $x_1(t)$  and  $x_2(t)$  are different processes, even when  $g$  is separable. In this latter case,  $g$  must be found from Eq. 3.

For some purposes, computation of  $g$  from Eq. 3 is tedious. Accordingly, an

alternative method of determining separability can be demonstrated in terms of characteristic functions. We compute

$$G(u_2, \tau) = \overline{x_1(t) e^{ju_2 x_2(t+\tau)}} \quad (7)$$

$$= \iint x_1 e^{ju_2 x_2} p(x_1, x_2; \tau) dx_1 dx_2 \quad (8)$$

$$= \frac{1}{j} \left. \frac{\partial f(u_1, u_2; \tau)}{\partial u_1} \right|_{u_1=0} \quad (9)$$

where  $f(u_1, u_2; \tau)$  is the characteristic function of the processes:

$$f(u_1, u_2; \tau) = \iint e^{ju_1 x_1} e^{ju_2 x_2} p(x_1, x_2; \tau) dx_1 dx_2 \quad (10)$$

Now if  $G$  is separable, i. e. ,

$$G(u_2, \tau) = G_1(u_2) G_2(\tau) \quad (11)$$

then  $g$  may be shown to be separable, and conversely. Thus, our question of separability is answered by a differentiation of the characteristic function rather than by an integration of the joint probability density function. This differentiation operation is, in some cases, much simpler to work with. For the cases in which some physical properties of the time functions are apparent, Eq. 7 offers advantages in calculation and determination of separability.

In addition, to illustrate the connection of separability with more familiar notions, the following statement can be made for a separable  $g$  function:

$$\overline{x_1(t) x_2^n(t+\tau)} = b_n \phi(\tau) \quad b_n \text{ real} \quad (12)$$

for all  $n$  for which the left-hand side exists. Conversely, if Eq. 12 is satisfied for all  $n$ , then  $g$  is separable. Thus, we see that the question of separability is tied up with the question of whether or not the crosscorrelation of  $x_1(t)$  with any power of  $x_2(t)$  is the same except for scale factors. It is worth noting that a reservation is stated with Eq. 12 regarding the existence of  $\overline{x_1(t) x_2^n(t+\tau)}$ . Thus, Eq. 12 and separability of  $g$  are not equivalent. Separability is a much more lenient condition. The necessity of Eq. 12 could never be pointed out in the general case. Such was the trouble in the original method of attack on the problem (1).

The extension to nonstationary inputs and time-dependent devices will be stated briefly in the following paragraphs.

(VIII. STATISTICAL COMMUNICATION THEORY)

$$g(x_2; t_1, t_2) = h(x_2, t_2) \phi(t_1, t_2) \quad \text{for all } t_1, t_2, \text{ almost everywhere in } x_2 \quad (13)$$

is a necessary and sufficient condition for the invariance property to hold for a particular joint probability density function. If, in addition, we have the same input processes, and  $g$  is separable, we can show that

$$g(x_2; t_1, t_2) = \sigma^{-2}(t_2) x_2 p(x_2, t_2) \phi(t_1, t_2) \quad (14)$$

and we can evaluate

$$C_f(t_2) = \int f(x_2, t_2) \frac{x_2 - \mu(t_2)}{\sigma^2(t_2)} p(x_2, t_2) dx_2 \quad (15)$$

The parameter  $t_2$  in  $C_f$  must be kept because we are allowing time-varying networks. If we restrict ourselves to time-invariant networks, but allow nonstationary inputs, we get

$$g(x_2; t_1, t_2) = h(x_2) \phi(t_1, t_2) \quad \text{for all } t_1, t_2, \text{ almost everywhere in } x_2 \quad (16)$$

as a necessary and sufficient condition for the invariance property to hold. This relation is somewhat more restrictive than Eq. 13.

Our statement of the invariance property can be generalized to (the stationary case again)

$$\Phi_f(\tau) = C_{f_1} \phi(\tau) + C_{f_2} \quad \text{for all } \tau, \quad \text{for any } f \text{ in } P \quad (17)$$

in which case it may be shown that the necessary and sufficient condition on the input statistics is

$$g(x_2, \tau) = h_1(x_2) \phi(\tau) + h_2(x_2) \quad \text{almost everywhere in } x_2, \quad \text{for all } \tau \quad (18)$$

This is, of course, more general than Eq. 4. All of the previous results have analogous ones under this more general formulation of the invariance property. This will not be demonstrated here. Barrett and Lampard's formulation (4) is included in the formulation of Eq. 17.

Suppose we insert, in Fig. VIII-5, another nonlinear device  $f'$  in the top lead, and then ask, What (if any) is the necessary and sufficient condition that must be imposed on  $p(x_1, x_2; \tau)$  for the invariance property to hold for any pair  $ff'$ ? The answer will be stated here only for the simplest case: when we have stationary inputs, time-invariant devices, and no additive constant ( $C_{f_2}$ ), as in Eq. 17, the necessary and sufficient condition is

$$p(x_1, x_2; \tau) = h(x_1, x_2) \phi(\tau) \quad \text{for all } \tau, \text{ almost everywhere in } x_1, x_2 \quad (19)$$

## (VIII. STATISTICAL COMMUNICATION THEORY)

Luce (2) proved the sufficiency of this relation but was unable to prove its necessity except under very restrictive conditions. Extensions to nonstationary cases are straightforward.

A. H. Nuttall

### References

1. A. H. Nuttall, Invariance of correlation functions under nonlinear transformations, Quarterly Progress Report, Research Laboratory of Electronics, M. I. T., Jan. 15, 1957, p. 61.
2. R. D. Luce, Quarterly Progress Report, Research Laboratory of Electronics, M. I. T., April 15, 1953, p. 37.
3. J. J. Bussgang, Crosscorrelation functions of amplitude-distorted gaussian signals, Technical Report 216, Research Laboratory of Electronics, M. I. T., March 26, 1952.
4. J. F. Barrett and D. G. Lampard, An expansion for some second-order probability distributions and its applications to noise problems, Trans. IRE, vol. PGIT-1, pp. 10-15, March 1955.

### C. A THEORY OF SIGNALS

The object of this research is to study the possibilities of representing signals in ways other than the usual one of identifying them with functions. It is reasonable to suppose that other representations may exist and, in fact, be more economical than the functional one, because the function representation does not take into account – except, perhaps, as an afterthought – any of our limitations in performing measurements. Since all of our measuring instruments have only finite accuracy, it seemed that it might be profitable to try to make this feature an intrinsic part of our description, rather than to regard it as an undesirable complication to be neglected as often as possible. Various representations were considered and, while the analysis will not be reported here, the conclusion seems to be that the proper subject of study is the measurement process itself, that is, the detector. This is plausible when we consider that a signal is completely unknown to us until it is detected, and what it is after being detected depends very much on what we detect it with. For example, consider the difference in the nature (and hence in the most cogent or economical description) of signals as seen through a zero-crossing detector and as seen on a linear oscilloscope.

Our object, then, is to develop a description of signals that incorporates, as much as possible, the peculiarities and limitations of a given detection process. A good description should retain only signal information that is actually distinguishable to the detector. Clearly, no single general description will do: the appropriate description will be different for each significantly different detector. In the remainder of this

(VIII. STATISTICAL COMMUNICATION THEORY)

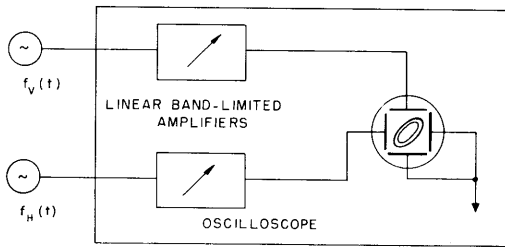


Fig. VIII-6. Comparison measurement of two signals on a band-limited, linear oscilloscope.

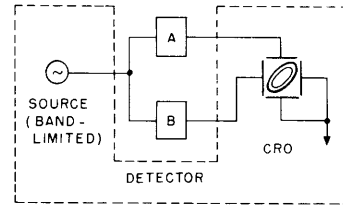


Fig. VIII-7. Measurement arrangement equivalent to that of Fig. VIII-6.

report, we present an algebra of signals appropriate to the linear, finite-accuracy measurement process.

We begin with two signals  $f_V(t)$  and  $f_H(t)$  which are unrestricted in any way except that they must be of finite duration  $T$ , and attempt to compare them by observing them on an oscilloscope, as in Fig. VIII-6. As for the oscilloscope, we assume that it has identical horizontal and vertical band-limited, linear amplifiers, a screen of finite size, and a trace of finite width. In the laboratory, we say that two signals are equal (within equipment accuracy) if the trace on the CRO screen is a straight line at  $45^\circ$ . Let  $f_V(t)$  and  $f_H(t)$  be the signals applied to the vertical and horizontal deflection plates, respectively, and suppose that the screen shows a line of width  $\sqrt{2} \epsilon$  at  $45^\circ$ . If the equally adjusted CRO amplifiers were flat (with linear phase) out to infinite frequencies, we would then know that  $|f_V(t) - f_H(t)| < \epsilon$  for all values of  $t$ . But how similar do  $f_V$  and  $f_H$  have to be to produce a line when the CRO has band-limited amplifiers? What features of the signals are distinguishable to the oscilloscope, and how can these features be summarized concisely, disregarding those that are indistinguishable? These questions are the subject of the present analysis.

To achieve a neater presentation, we shall think of our signals as impulse responses of networks. This can always be done, since the signals are of finite duration. We shall excite two networks with signals from a band-limited source (having the same spectrum as the CRO amplifiers) and compare the network outputs on an oscilloscope of finite trace length and width, but equipped with ideal amplifiers. The diagram is shown in Fig. VIII-7. Clearly, the results of measuring with this arrangement are the same as the results obtained with the arrangement of Fig. VIII-6. We regard the source-CRO combination as the detector to be studied. Besides linearity, only three features of the detector are relevant: (a) the source, band-limited to the radian-frequency interval  $(-W, W)$ , but with an arbitrary spectrum in that interval; (b) the finite trace width of the oscilloscope; (c) the finite diameter of the oscilloscope screen, and consequent finite trace length. Each one of these characteristics yields an

important part of the analysis.

The source, as long as it is band-limited, may be periodic, almost periodic, or aperiodic. The present analysis will be limited to the aperiodic case; in fact, to aperiodic functions whose square is integrable. Since we are assuming our source to be band-limited, we could use the Shannon sampling theorem to characterize the allowable source functions, by requiring that they all be expressible in the form

$$\sum_{n=-\infty}^{\infty} f(t_n) \frac{\sin W(t - t_n)}{W(t - t_n)}$$

with the sampling points  $t_n$  suitably chosen. This is inconvenient because it is more awkward to work with infinite series than with integrals. A different way of characterizing band-limited source functions is made possible by the following result.

**THEOREM 1:**  $f(t)$  is a source function [i. e., a function whose Fourier transform is zero outside of the band  $-W \leq \omega \leq W$ ] if and only if it is a solution of the equation

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\tau) \phi_W(t-\tau) d\tau \quad (1)$$

where

$$\phi_W(t) = 2W \frac{\sin Wt}{Wt} \quad (2)$$

The proof is very easy. Note first that the Fourier transform  $\Phi_W(\omega)$  of  $\phi_W(t)$  is a rectangular pulse, of unit height for  $\omega \in (-W, W)$  and zero outside this interval. Then, taking Fourier transforms of both sides of Eq. 1 [and writing  $F(\omega)$  for the transform of  $f(t)$ ], we obtain

$$F(\omega) = F(\omega) \Phi_W(\omega)$$

Clearly, this equation is true if and only if  $F(\omega)$  is limited to a band smaller than or equal to  $(-W, W)$ .

The second important feature of the detector is that the CRO trace has finite width. This immediately implies a finite measurement error, which in turn allows us to substitute, for the impulse response  $h(t)$  of a network, a singular impulse response of the form

$$\sum_{n=1}^N a_n \delta(t - t_n) \quad (3)$$



(VIII. STATISTICAL COMMUNICATION THEORY)

(where  $\delta(t)$  denotes a unit impulse) without a distinguishable difference in the outputs resulting from band-limited inputs. A possible set of specific conditions under which the substitution is possible is given in reference 1. Here we need only notice that this result allows us to represent networks (and therefore signals) by numerical operators

$$\Omega = \sum_{n=1}^N a_n E^{-t_n} \quad (4)$$

where  $E^{-t_n}$  is the shift operator, defined by  $E^{-t_n} f(t) = f(t - t_n)$ . Thus, as soon as the trace has finite width, however small, networks can be represented by finite linear combinations of shift operators.

Before passing to a study of the third property of the detector, we pause to notice that, if we combine Eqs. 1 and 4, we can express the effect of any network on our source functions in the form

$$\Omega f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\tau) [\Omega \phi_W(t-\tau)] d\tau \quad (5)$$

The interchange of operation and integration follows from the fact that  $\Omega$  operates on functions of time only, and is just a finite linear combination of shifts. Of course, by making the change of variable  $\mu = t - \tau$ , Eq. 1 can be rewritten as

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t-\mu) \phi_W(\mu) d\mu \quad (6)$$

in which case

$$\Omega f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\Omega f(t-\mu)] \phi_W(\mu) d\mu \quad (7)$$

In view of Theorem 1, Eq. 7 proves the obvious fact that if the input of a linear, time-invariant network is band-limited, then so is its output. Equation 5 is more interesting: it states that the effect of an operator  $\Omega$  on any input  $f(t)$  can be determined simply by knowing how  $\Omega$  affects  $\phi_W(t)$ .

The third important feature of the detector is the finite CRO trace length. In the laboratory, this makes it necessary to adjust the source and the CRO amplifiers so that the display will fit inside the screen. In our analysis, it makes it necessary to normalize the outputs of our networks. This is accomplished in two steps: first, we normalize the inputs, that is, the set of allowable source functions; then we normalize the set of allowable operators.

(VIII. STATISTICAL COMMUNICATION THEORY)

Let  $\mathcal{S}$  be the set of aperiodic, band-limited functions of integrable square. Our first step is made possible by the following result.

**THEOREM 2:** If  $f \in \mathcal{S}$  and has an energy (into a one-ohm load) equal to that of  $\phi_W(t)$ , then  $|f(t)| \leq \phi_W(0)$ . (Roughly,  $\phi_W(t)$  is the tallest function in  $\mathcal{S}$ , energies being equal.)

The proof, which is simple, will be omitted.

Let us now modify every  $f \in \mathcal{S}$  by multiplying it by a constant so chosen that the total energy of  $f$  (into one ohm) is  $\pi/W$ . That is, choose  $K$  so that

$$K^2 \int_{-\infty}^{\infty} f^2(t) dt = \frac{\pi}{W} \quad (8)$$

Since  $f$  is of integrable square, an appropriate finite  $K$  always exists. Call the set of modified source functions  $\mathcal{S}^*$ . Then, if  $f \in \mathcal{S}^*$ , we can apply the Schwartz inequality to Eq. 1 to obtain

$$|f(t)| \leq \frac{1}{2\pi} \left[ \int_{-\infty}^{\infty} f^2 dt \right]^{1/2} \left[ \int_{-\infty}^{\infty} \phi_W^2 dt \right]^{1/2} = \frac{1}{2\pi} \left( \frac{\pi}{W} \right)^{1/2} (4\pi W)^{1/2} = 1 = \frac{\phi_W(0)}{2W} \quad (9)$$

We proceed now to the second step in the normalization process. From Eq. 5 and the Schwartz inequality, we have that

$$|\Omega f(t)| \leq \frac{1}{2\pi} \left[ \int_{-\infty}^{\infty} f^2(t) dt \right]^{1/2} \left[ \int_{-\infty}^{\infty} [\Omega \phi_W(t)]^2 dt \right]^{1/2}$$

If now  $f \in \mathcal{S}^*$ ,

$$|\Omega f(t)| \leq \frac{1}{2\pi} \left( \frac{\pi}{W} \right)^{1/2} \left[ \int_{-\infty}^{\infty} [\Omega \phi_W(t)]^2 dt \right]^{1/2} = \frac{1}{(4\pi W)^{1/2}} \left[ \int_{-\infty}^{\infty} [\Omega \phi_W(t)]^2 dt \right]^{1/2} \quad (10)$$

Therefore, we can place a bound on  $|\Omega f(t)|$  simply by knowing how the operator  $\Omega$  acts on  $\phi_W(t)$ , thus making it possible to normalize operators independently of inputs. Let  $\mathcal{O}$  be the set of all finite operators, that is, of all operators  $\Omega$  of the form of Eq. 4. These operators themselves have an arbitrary multiplicative constant, corresponding to the gains of the CRO vertical and horizontal amplifiers. Let us modify each  $\Omega \in \mathcal{O}$  by multiplying it by the constant  $K$  that makes

$$K^2 \int_{-\infty}^{\infty} [\Omega \phi_W(t)]^2 dt = 4\pi W$$

(VIII. STATISTICAL COMMUNICATION THEORY)

and call the new set of modified  $\Omega$ 's  $\mathcal{O}^*$ . Since  $\Omega$  is a finite operator and  $\phi_W$  is integrable square, an appropriate finite  $K$  always exists. It now follows from Eq. 10 that, for any  $f \in \mathcal{S}^*$  and any  $\Omega \in \mathcal{O}^*$ ,  $|\Omega f(t)| \leq 1$ . The normalization process is complete.

Note, by the way, that Eq. 10 is useful for a far better purpose than just normalization. For if  $\Omega$  happens to be an error operator, i. e.,  $\Omega = \Omega_1 - \Omega_2$ , where  $\Omega_1$  and  $\Omega_2$  are two operators that are being compared, then Eq. 10 tells us that we can get a conservative estimate of the maximum error which will be incurred with any  $f \in \mathcal{S}^*$  simply by knowing what the error is for  $\phi_W(t)$ , and then evaluating

$$\int_{-\infty}^{\infty} [\Omega \phi_W(t)]^2 dt \quad (11)$$

We might guess, from the special nature of the function  $\phi_W$  and the finiteness of  $\Omega$ , that Eq. 11 could be expressed in numerical or algebraic form. It turns out that this can actually be done, as is shown by the following theorem.

THEOREM 3:† If  $\Omega$  is any finite operator (i. e.,  $\Omega \in \mathcal{O}$ ) then

$$\int_{-\infty}^{\infty} [\Omega \phi_W(t)]^2 dt = 2\pi \bar{\Omega} \Omega \phi_W(0) \quad (12)$$

where

$$\Omega = \sum_{n=1}^N a_n E^{-t_n} \quad \text{and} \quad \bar{\Omega} = \sum_{n=1}^N a_n E^{+t_n}$$

(That is,  $\bar{\Omega}$  is just  $\Omega$  folded over, so as to sample forward instead of backward.)

PROOF: Since  $\Omega \phi_W(t)$  is a band-limited function, it can be written in the form

$$\Omega \phi_W(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\Omega \phi_W(\tau)] \phi_W(t-\tau) d\tau \quad (13)$$

If  $\Omega$  is given by

$$\Omega = \sum_{n=1}^N a_n E^{-t_n}$$

then Eq. 13 can be written more explicitly as

---

† Besides its usefulness in our problem, and its being a source of curious identities and inequalities, Eq. 12 is interesting in itself for the numerical evaluation of certain types of definite integrals.

$$\Omega \phi_W(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\sum_n a_n \phi_W(\tau - t_n)] \phi_W(t-\tau) d\tau$$

If we operate on  $\Omega \phi_W(t)$  with  $\bar{\Omega}$ , we obtain

$$\begin{aligned} \bar{\Omega} \Omega \phi_W(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} [\sum_n a_n \phi_W(\tau - t_n)] [\bar{\Omega} \phi_W(t-\tau)] d\tau \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} [\sum_n a_n \phi_W(\tau - t_n)] [\sum_m a_m \phi_W(t + t_m - \tau)] d\tau \end{aligned}$$

At  $t = 0$ ,

$$\bar{\Omega} \Omega \phi_W(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\sum_n a_n \phi_W(\tau - t_n)] [\sum_m a_m \phi_W(t_m - \tau)] d\tau$$

Because of the evenness of  $\phi_W(t)$ ,  $\phi_W(t_m - \tau) = \phi_W(\tau - t_m)$ . If we make this change, the two brackets in the integrand become equal and Eq. 12 is obtained.

Theorem 3 is one of the main results of this work, as will be seen presently, when we discuss the algebra of signals. Meanwhile, we note that, using Theorem 3 in Eq. 10, we obtain, for any  $f \in \delta^*$ ,

$$|\Omega f(t)| \leq \frac{1}{(4\pi W)^{1/2}} [2\pi \bar{\Omega} \Omega \phi_W(0)]^{1/2} = \frac{1}{(2W)^{1/2}} [\bar{\Omega} \Omega \phi_W(0)]^{1/2} \quad (14)$$

We have all the results necessary to establish that the operators  $\Omega$ , together with the ordinary linear-circuit laws of combination, constitute an algebra. Remembering that the  $\Omega$ 's represent networks or signals, this is our desired signal algebra. We notice first that the set  $\mathcal{O}$  of all finite operators  $\Omega$ , together with addition and multiplication by a scalar (linear amplification) defined as usual, satisfies the axioms of a linear vector space (2). (Observe that the definition of  $\mathcal{O}$  does not make a distinction between forward and backward operators, so that whenever  $\Omega \in \mathcal{O}$ ,  $\bar{\Omega}$  belongs also. Physical realizability, of course, establishes a distinction between forward and backward sampling, but this fact is irrelevant here.) The importance of Theorem 3 is that it provides us with a reasonable definition for the norm  $||\Omega||$  of any operator  $\Omega \in \mathcal{O}$ . That the norm (or length) of an operator might be meaningful is suggested by Eq. 14. We define the norm of  $\Omega$  by

$$||\Omega|| = [\bar{\Omega} \Omega \phi_W(0)]^{1/2} = \left[ \frac{1}{2\pi} \int_{-\infty}^{\infty} [\Omega \phi_W(t)]^2 dt \right]^{1/2} \quad (15)$$

(VIII. STATISTICAL COMMUNICATION THEORY)

It is easy to check that  $||a\Omega|| = |a| \cdot ||\Omega||$  ( $a$  real), and that  $||\Omega|| > 0$  for  $\Omega \neq 0$ , since the integrand in Eq. 15 is always positive and the  $\phi_W$ 's are linearly independent with respect to shifts. The third requirement that  $||\Omega||$  has to satisfy to be a norm (3) is the triangle inequality

$$||\Omega_1 + \Omega_2|| \leq ||\Omega_1|| + ||\Omega_2||$$

To prove this, we start from the definition of norm and note that

$$\int [(\Omega_1 + \Omega_2)\phi_W]^2 = \int [\Omega_1\phi_W]^2 + \int [\Omega_2\phi_W]^2 + 2 \int [\Omega_1\phi_W][\Omega_2\phi_W]$$

or

$$||\Omega_1 + \Omega_2||^2 = ||\Omega_1||^2 + ||\Omega_2||^2 + 2 \cdot \frac{1}{2\pi} \int [\Omega_1\phi_W][\Omega_2\phi_W]$$

Using the Schwartz inequality, we have

$$\frac{1}{2\pi} \left| \int [\Omega_1\phi_W][\Omega_2\phi_W] \right| \leq \left[ \frac{1}{2\pi} \int [\Omega_1\phi_W]^2 \right]^{1/2} \left[ \frac{1}{2\pi} \int [\Omega_2\phi_W]^2 \right]^{1/2} = ||\Omega_1|| \cdot ||\Omega_2||$$

Therefore,

$$||\Omega_1 + \Omega_2||^2 \leq ||\Omega_1||^2 + ||\Omega_2||^2 + 2||\Omega_1|| \cdot ||\Omega_2|| = [||\Omega_1|| + ||\Omega_2||]^2$$

Taking square roots of both sides establishes the desired result.

Therefore,  $\mathcal{O}$  is a normed vector space. But it is more: the necessary and sufficient condition (3) for the existence of an inner product giving rise to a preassigned norm  $||\Omega||$  is the general validity of

$$||\Omega_1 + \Omega_2||^2 + ||\Omega_1 - \Omega_2||^2 = 2[||\Omega_1||^2 + ||\Omega_2||^2]$$

Straightforward substitution of  $||\Omega|| = [\overline{\Omega\Omega\phi_W(0)}]^{1/2}$  immediately establishes this result. Therefore, the normed vector space based on  $\mathcal{O}$  is, in fact, a unitary space. The advantage of a unitary space is that it gives us a natural definition of orthogonality (between networks or signals) in addition to a natural definition of distance.

One more operation is physically meaningful, and that is multiplication of the elements of  $\mathcal{O}$ , corresponding to the cascading of two networks or the passage of a signal through a network. It is easy to see that multiplication should be defined by

$$\begin{aligned}\Omega_1\Omega_2 &= \left[ \sum_i a_{1i} E^{t_{1i}} \right] \left[ \sum_j a_{2j} E^{t_{2j}} \right] \\ &= \sum_{i,j} a_{1i} a_{2j} E^{(t_{1i}+t_{2j})}\end{aligned}$$

just as it is for polynomials. In terms of the impulse responses associated with the  $\Omega$ 's, this definition corresponds to the ordinary convolution process. Clearly, the result of multiplication is again an element of  $\mathcal{O}$ , and the operation is associative and distributive (in fact, for time-invariant systems it is even commutative). The unitary space based on  $\mathcal{O}$ , plus multiplication, constitutes an algebra of signals appropriate to our detector.

To make the algebra useful for practical problems, we have to be able to go back from the elements of  $\mathcal{O}$  to smooth responses or signals. This can be done; the reconstruction process is simple and the error is controlled at every step. The derivation of these results will be published in the Quarterly Progress Report, July 15, 1957.

R. E. Wernikoff

#### References

1. R. E. Wernikoff, Quarterly Progress Report, Research Laboratory of Electronics, July 15, 1956, p. 44.
2. P. R. Halmos, Finite Dimensional Vector Spaces (Princeton University Press, Princeton, N. J., 1942).
3. J. von Neumann, Functional Operators, Vol. II (Princeton University Press, Princeton, N. J., 1950).

#### D. AN ANALOG PROBABILITY DENSITY ANALYZER

The analog probability density analyzer described in previous issues of the Quarterly Progress Report was experimentally tested to evaluate its performance. The analyzer was initially checked by comparing the experimental probability density of a sine wave to calculated values. Further testing was concerned with determining the frequency response and drift stability of the analyzer. A summary of these test results and examples of probability density functions are included in this report.

An experimental probability density function of a 1-kc sine wave analyzed by 50 amplitude intervals is shown in Fig. VIII-8a. Calculated values of sinusoidal probability density for 50 intervals are shown on the same figure by crosses. Comparison of the experimental curve and the calculated values reveals that error is present in regions

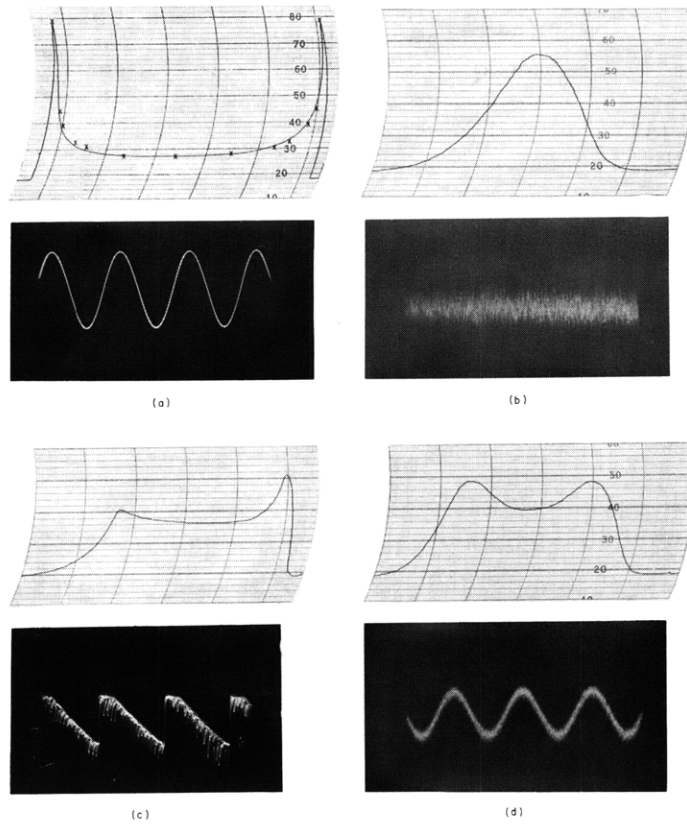


Fig. VIII-8. Amplitude probability density functions. (a) Sine wave. (b) Gaussian noise. (c) Clipped saw tooth and envelope of gaussian noise. (d) Sine wave and gaussian noise.

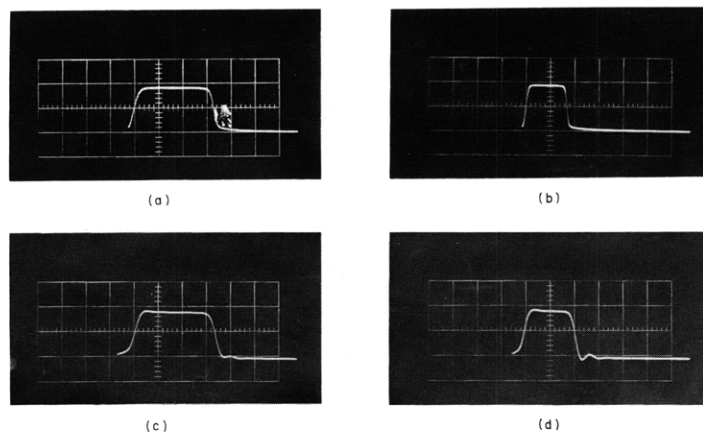


Fig. VIII-9. Pulses resulting from the analysis of a sine wave. (a) 1-kc sine wave. (b) 2-kc sine wave. (c) 10-kc sine wave. (d) 20-kc sine wave.

## (VIII. STATISTICAL COMMUNICATION THEORY)

where the slope of the probability density is large. Verification that this error is produced by the recorder was obtained by replacing the recorder with a voltmeter. Values of probability density read from the voltmeter showed no deviation from the calculated values. Reversing the sequence of scanning the amplitude intervals, scanning from negative amplitudes to positive rather than from positive amplitudes to negative, resulted in no change of the recorded sine-wave density function. The analyzer output is therefore symmetrical, and the dissymmetry of the recorded curve is introduced by the recorder. Since this error is small, useful results have been obtained with this recorder; however, it is useful to know that the error is not an inherent part of the system.

The analyzer frequency response may be investigated by either comparing sine-wave density functions for a range of frequencies or by examining the shape of the pulses which are averaged by the integrator to produce the probability density function. The pulses that are averaged when analyzing the center amplitude interval of a 1-kc, 2-kc, 10-kc, and a 20-kc sine wave are shown in Fig. VIII-9. The time base of 2  $\mu$ sec/cm used for the 1-kc and 2-kc pulses is changed to 0.2  $\mu$ sec/cm for the 10-kc and 20-kc pulses to facilitate comparison of the pulse shape. Examination of Fig. VIII-9 shows that the 1-kc and 10-kc pulses enclose almost equal areas, but that the 20-kc pulse encloses more area than the 2-kc pulse. The stretching of the 20-kc pulse is a result of rise-time limitations of the diode level selector and pulse amplifier and results in approximately a 10 per cent positive error in the probability density for amplitudes near the axis of a 20-kc sine wave. Since the probability density function is a minimum at the axis, a 10 per cent error in this region is 1 per cent of the maximum value of the sinusoidal density function. The lower bound of the analyzer frequency response is equal to the frequency response of the amplifier used to amplify the input signal and is 30 cps for the analog probability density analyzer. This limit, however, could be extended to direct current if a stable amplifier was available, since the other parts of the analyzer have direct-current response.

Experimental determination of the analyzer drift stability is readily accomplished by comparing repeated probability density functions. This process is facilitated, since the analog analyzer employs a periodic scanning system which automatically repeats the probability functions. Various functions have been repeated for periods of 6 and 8 hours with no deviations between the repeated curves greater than 2 per cent of the maximum value of the function. Since the analyzation time for a density function may be varied from 10 minutes to 5 hours, no difficulty has been encountered with analyzer drift.

Some other examples of probability density functions determined by the analog probability density analyzer and the corresponding time functions are shown in Fig. VIII-8b, c, and d. Future experimental work will be concerned with determining the probability density of functions that require longer integration times than those shown in Fig. VIII-8.

H. E. White



## (VIII. STATISTICAL COMMUNICATION THEORY)

### E. ON THE SYNTHESIS OF LINEAR SYSTEMS FOR PURE TRANSMISSION, DELAYED TRANSMISSION, AND LINEAR PREDICTION OF SIGNALS<sup>†</sup>

#### VI. DELAYED TRANSMISSION

##### 6.0 OBJECTIVE OF THIS CHAPTER

We now propose to give in condensed form some results associated with the delayed transmission of a signal  $\phi(t)$ . This solution is represented by the expression

$$\gamma(t) = \phi(t - T_o)$$

where  $T_o$  is the delay time.

The measures of the windows that form the distribution are given by expression 4(II-2.5), which is repeated for convenience

$$a_j = \prod_{p=0}^m \left( \frac{T_o - \mu_p}{\mu_j - \mu_p} \right) \quad 1(\text{VI-6.0})$$

For simplicity in the use of this formula, the delay time will be measured by taking as the unit of time the aperture  $\delta$  of a window. We shall then set

$$T_o = t_o \delta \quad 2(\text{VI-6.0})$$

When the window distribution has a constant aperture, we have

$$\mu_p = (p+1) \delta \quad p = 0, 1, \dots, m$$

Then

$$a_j = \prod_{p=0}^m \left( \frac{t_o - p - 1}{j - p} \right) \quad 3(\text{VI-6.0})$$

##### 6.1 INTRINSIC DELAY

In the transmission of signals with delay we consider two cases similar to those found in pure transmission: (a) intrinsic transmission and (b) weighted transmission. In this section we refer to intrinsic transmission. To simplify the discussion, we

---

<sup>†</sup>(Continued from the Quarterly Progress Report of Jan. 15, 1957.) Translated by R. E. Wernikoff from the Spanish — with some corrections and additions that were made, particularly in the sections dealing with error analysis, by Mr. Wernikoff in cooperation with Dr. Cerrillo.

[Editor's note: This material, which was published under the title "Sobre la Sintesis de Sistemas Lineales para la Transmision sin Retraso, Retrasada, y Prediccion Lineal de Señales," in Revista Mexicana de Fisica, is an application of the theory given in Technical Report 270, by Dr. Cerrillo, "On Basic Existence Theorems. Part V. Window Function Distributions and the Theory of Signal Transmission" (to be published). The direct connection of the present paper with the work of the Statistical Communication Theory group and other groups in the Laboratory led to its translation, by Mr. Wernikoff, and its presentation here.]

(VIII. STATISTICAL COMMUNICATION THEORY)

consider again the polynomial  $G_{(j)}(\mu)$  defined by Eq. 1(II-2.22) and we recall its basic property

$$G_{(j)}(\mu_p) \begin{cases} =0 & \text{for } p = 0, 1, \dots, j-1, j+1, \dots, m \\ \neq 0 & \text{for } p = j \end{cases} \quad 1(\text{VI-6.1})$$

Expression 1(VI-6.0) may be written in terms of this polynomial as follows:

$$a_j = \frac{G_{(j)}(t_o)}{G_{(j)}(\mu_j)} \quad 2(\text{VI-6.1})$$

Having stated this, we can define intrinsic delay as that which is equal to one of the values  $\mu$ . Let us say

$$t_o = \mu_j \quad 3(\text{VI-6.1})$$

Introducing this condition in Eq. 2(VI-6.1), we have

$$a_k = \begin{cases} 0 & \text{for } k = 0, 1, \dots, j-1, j+1, \dots, m \\ 1 & \text{for } k = j \end{cases}$$

which indicates that the window distribution contains only the  $j^{\text{th}}$  window, independent of the number  $m$ . Figure 1(VI-6.1) shows the graphical interpretation of the convolution integral. We see that the contribution to the response comes from just one window

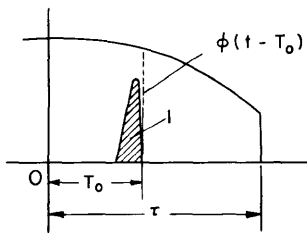


Fig. 1(VI-6.1). Operation of intrinsic delayed transmission.

operating at time  $t = T_o = \mu_j$ . Hence the term "intrinsic." The linear system behaves like a simple delay system, the response function being precisely the input function at time  $t - T_o$ .

It is worth while to interpret this result in connection with the ideas of present, past, and future, mentioned before. Comparing, at the same time, the excitation  $\phi(t)$  and the response  $\phi(t - T_o)$  is equivalent to thinking that exciting the system at time  $t$

generates a "memory" of the system of a past  $\phi(t - T_o)$  which occurred  $T_o$  units of time before. This memory is formed by the behavior of the excitation function at the precise instant  $t - T_o$  when the transmission is intrinsic. This memory is also called "intrinsic."

(VIII. STATISTICAL COMMUNICATION THEORY)

6.2 WEIGHTED TRANSMISSION

When  $T_o \neq \mu_j$ ,  $j = 0, 1, \dots, m$ , then all  $m + 1$  measures  $a_k$ ,  $k = 0, 1, \dots, m$ , are non-zero. The response is

$$\gamma(t) = \phi(t - T_o) = \sum_{p=0}^m a_p \phi(t - \mu_p) \quad 1(\text{VI-6.2})$$

which is formed by the weighted sampling of the function  $\phi(t)$  not at the time  $t - T_o$ , but at the times  $t - \mu_p$ . Comparing the function  $\phi(t)$  and  $\phi(t - T_o)$ , we now say that the system produces a weighted memory.

Table 1(VI-6.2)

	$m = 0$	$m = 1$	$m = 2$	$m = 3$	$m = 4$
$a_0$	---	$-(t_o - 2)$	$\frac{t_o^2 - 5t_o + 6}{2}$	$-\frac{t_o^3 - 9t_o^2 + 26t_o - 24}{6}$	$\frac{t_o^4 - 14t_o^3 + 71t_o^2 - 154t_o + 120}{24}$
$a_1$		$t_o - 1$	$-(t_o^2 - 4t_o + 3)$	$\frac{t_o^3 - 8t_o^2 + 19t_o - 12}{2}$	$-\frac{t_o^4 - 13t_o^3 + 59t_o^2 - 107t_o + 60}{6}$
$a_2$			$\frac{t_o^2 - 3t_o + 2}{2}$	$-\frac{t_o^3 - 7t_o^2 + 14t_o - 8}{2}$	$\frac{t_o^4 - 12t_o^3 + 49t_o^2 - 78t_o + 40}{4}$
$a_3$				$\frac{t_o^3 - 6t_o^2 + 11t_o - 6}{6}$	$-\frac{t_o^4 - 11t_o^3 + 41t_o^2 - 61t_o + 30}{6}$
$a_4$					$\frac{t_o^4 - 10t_o^3 + 35t_o^2 - 50t_o + 24}{24}$

Use of formula 3(VI-6.0) produces the results given in Table 1(VI-6.2). As a numerical example, we give the distributions corresponding to the delays  $T_o = 10$  and  $T_o = 20$ , with  $m = 3$ . See Fig. 1(VI-6.2), which is not to scale. Note the strong growth of the measures in going from  $T_o = 1$  to  $T_o = 10$  to  $T_o = 20$ .

In delayed transmission the measures  $a$  have an important property that follows from expression 3(II-2.4), which is

$$\sum_{p=0}^m (p+1)^k a_p = t_o^k \quad k = 0, 1, \dots, m \quad 1(\text{VI-6.2})$$

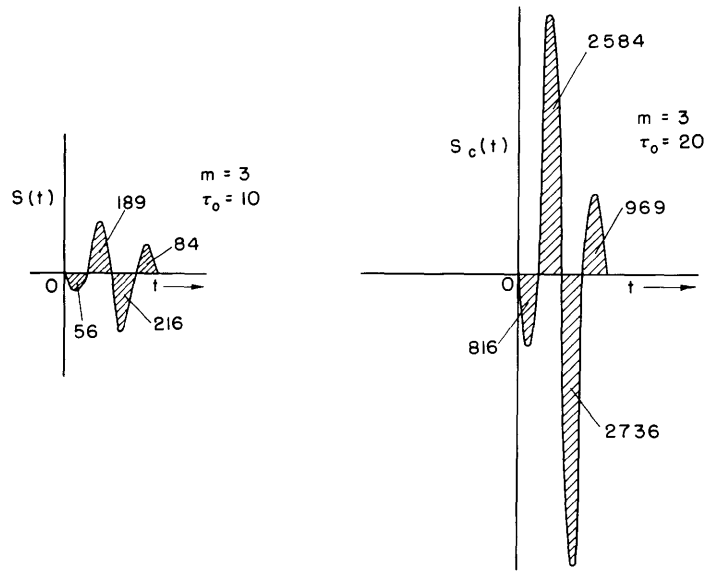


Fig. 1(VI-6.2). Distributions corresponding to the delays  $T_0 = 10$  and  $T_0 = 20$ , with  $m = 3$ .

### 6.3 DETERMINATION OF THE ERROR

We start from the general expression for the error, Eq. 5(II-2.7), which we repeat for convenience

$$|\xi_{(m)}| \approx \frac{\Omega^{(m)}(\delta) \delta^m}{(m+1)!} \left| \sum_{k=0}^m (1+k)^{m+1} a_k \right| \quad 1(\text{VI-6.3})$$

It is difficult to gain an idea of the magnitude of the error by using Eq. 1(VI-6.3) as it stands. To make the equation more useful, we must find a simpler expression for the sum, or at least a simple upper bound for the magnitude of the sum.

Let

$$\sigma = \sum_{k=0}^m (1+k)^{m+1} a_k$$

Expanding the summand by the binomial theorem, using Eq. 1(VI-6.2), and after some manipulating, we obtain

$$\sigma = \sum_{k=0}^m (1+k)^{m+1} a_k = t_0^{m+1} - (t_0 - 1)^{m+1} + \sum_{k=0}^m k^{m+1} a_k \quad 2(\text{VI-6.3})$$

(VIII. STATISTICAL COMMUNICATION THEORY)

In the sum

$$\sigma_1 = \sum_{k=0}^m k^{m+1} a_k \quad 3(\text{VI-6.3})$$

the polynomial  $k^{m+1}$  takes on  $m + 1$  values. But it is well known that we can find a polynomial  $P_m(x)$  of order  $m$  which, at the  $m + 1$  points  $x = k$  ( $k = 0, 1, \dots, m$ ), takes on the same values as  $k^{m+1}$ . Let this polynomial be

$$P_m(x) = \beta_m x^m + \beta_{m-1} x^{m-1} + \dots + \beta_0 = \sum_{j=0}^m \beta_j x^j \quad 4(\text{VI-6.3})$$

Then  $\sigma_1$  is also given by

$$\sigma_1 = \sum_{k=0}^m \left( \sum_{j=0}^m \beta_j k^j \right) a_k$$

If we invert the order of summation, this becomes

$$\sigma_1 = \sum_{j=0}^m \beta_j \left( \sum_{k=0}^m k^j a_k \right) \quad 5(\text{VI-6.3})$$

Now, we can easily derive from Eq. 1(VI-6.2) that

$$\sum_{k=0}^m k^j a_k = (t_0 - 1)^j \quad 6(\text{VI-6.3})$$

so that Eq. 5(VI-6.3) becomes

$$\sigma_1 = \sum_{j=0}^m \beta_j (t_0 - 1)^j \quad 7(\text{VI-6.3})$$

But this is just our polynomial  $P_m(x)$  evaluated at  $x = (t_0 - 1)$ . Thus we have

$$\sigma_1 = P_m(t_0 - 1) \quad 8(\text{VI-6.3})$$

This answer is exact, and holds for all values of  $t_0$ .

Since we constructed  $P_m(x)$  in such a way that  $P_m(x) = x^{m+1}$  when  $x = k$  ( $k = 0, 1, \dots, m$ ), for integer values of  $t_0$  with the property that  $0 \leq (t_0 - 1) \leq m$ , we have

$$\sigma_1 = P_m(t_0 - 1) = (t_0 - 1)^{m+1} \quad 9(\text{VI-6.3})$$

(VIII. STATISTICAL COMMUNICATION THEORY)

For other values of  $t_o$ , this simple form cannot be used, and the whole polynomial 4(VI-6.3) must be evaluated. This is no simpler than the original sum, Eq. 3(VI-6.3); thus nothing has been gained. However, for very large values of  $t_o$ , only the highest term of the polynomial 4(VI-6.3) is important, and so, for  $t_o \rightarrow \infty$  (or actually,  $t_o \gg m$ ), we have

$$\sigma_1 = P_m(t_o - 1) \approx \beta_m (t_o - 1)^m \quad 10(\text{VI-6.3})$$

Substituting Eqs. 9(VI-6.3) and 10(VI-6.3) in Eq. 2(VI-6.3), we have

$$\sigma = t_o^{m+1} \quad \text{for } 0 \leq t_o \leq m + 1 \quad 11(\text{VI-6.3})$$

$$\sigma \rightarrow \beta_m t_o^m \quad \text{for } t_o \gg m \quad 12(\text{VI-6.3})^*$$

Putting these results in the error expression 1(VI-6.3), we find that, for intrinsic transmission ( $0 \leq t_o \leq m$ ) and also for  $t_o = m + 1$ , the error is given by

$$|\xi_{(m)}| = \frac{\Omega^{(m)}(\delta) \delta^m}{(m+1)!} |t_o^{m+1}| \quad 13(\text{VI-6.3})$$

while for large delays,  $t_o \rightarrow \infty$ , the error is given by

$$|\xi_{(m)}| \rightarrow \frac{\Omega^{(m)}(\delta) \delta^m}{(m+1)!} |\beta_m t_o^m| \quad 14(\text{VI-6.3})$$

For intermediate values of  $t_o$ , Eq. 14(VI-6.3) can be used to obtain an upper bound on the error, since the error grows monotonically with  $t_o$ .

---

\* Nothing has been said about how to determine  $\beta_m$ . However, if we recall that only one polynomial of order  $m$  can be passed through  $m + 1$  points,  $\beta_m$  is uniquely determined and can be obtained from any polynomial that satisfies the stated requirements. In particular, if we represent  $x^{m+1}$  by the Lagrange polynomial of order  $m$ , we immediately obtain

$$\beta_m = \sum_{j=0}^m \frac{(-1)^{m-j} (j)^{m+1}}{j! (m-j)!}$$

(VIII. STATISTICAL COMMUNICATION THEORY)

6.4 TRANSFER FUNCTION ASSOCIATED WITH THE CASE OF DELAYED TRANSMISSION

Taking into account expression 1(V-5.7), which is repeated for convenience, we have, in general,

$$\mathcal{L} [S_c(t)] = H(s) e^{s\delta} \sum_{p=0}^m a_p e^{-(p+1)\delta s} \quad 1(\text{VI-6.4})$$

and, for delayed transmission, we have

$$\mathcal{L} [S_c(t)] = H(s) e^{s\delta} \sum_{p=0}^m \prod_{j=0}^m \left( \frac{t_0 - j - 1}{p - j} \right) e^{-(p+1)\delta s} \quad 2(\text{VI-6.4})$$

Equation 2(VI-6.4) is difficult to discuss and interpret and in direct application it is practically useless in this form. To clarify the situation somewhat, we proceed as follows. Let us take the general formula 1(VI-6.4) and let us write

$$\begin{aligned} \mathcal{L} [S_c(t)] &= H(s) e^{s\delta} \sum_{p=0}^m a_p \sum_{k=0}^{\infty} \frac{[-(p+1)\delta]^k}{k!} s^k \\ &= H(s) e^{s\delta} \sum_{k=0}^{\infty} \frac{(-\delta s)^k}{k!} \sum_{p=0}^m a_p (p+1)^k \end{aligned} \quad 3(\text{VI-6.4})$$

If we use expression 1(VI-6.2) as though it held for all  $k$ , we have

$$\mathcal{L} [S_c(t)] = H(s) e^{s\delta} \sum_{k=0}^{\infty} \frac{(-s\delta)^k}{k!} t_0^k = H(s) e^{s\delta} e^{-s\delta t_0} \quad 4(\text{VI-6.4})$$

Thus we see that the transfer function which is obtained has the expected form. For example, for a singular distribution  $S_c(t)$ , we have  $H(s) e^{s\delta} = 1$ . (See Chapter V, section 5.7.) In this way we obtain the ideal case

$$\mathcal{L} [S_c(t)] = e^{-sT_0} \quad 5(\text{VI-6.4})$$

When we substitute the function  $\gamma^*(t)$  for  $\gamma(t)$  it is equivalent to using only the first  $m$  terms of the Taylor series of  $\phi[t - (k+1)\delta]$ . The relation

$$\sum_{p=0}^m (p+1)^k a_p = t_0^k$$

obtains only for the values  $k = 0, 1, \dots, m$ , but not for  $k \geq m + 1$ . Consequently, expression 4(VI-6.4) does not follow exactly from expression 3(VI-6.4) and does not represent the transfer function formed with  $m$  windows, except as  $m \rightarrow \infty$ . To understand the transfer function associated with  $m$  windows, let us write

$$\sum_{p=0}^m (p+1)^k a_p = t_o^k (1 - \beta_k) \quad k = m+1, \dots \quad 6(\text{VI-6.4})$$

where  $\beta_k$  is defined by the equation itself. Substituting this expression in Eq. 3(VI-6.4), we finally obtain

$$\mathcal{L}[S_c(t)] = H(s) e^{\delta s} \left[ e^{-sT_o} - \sum_{k=m+1}^{\infty} \frac{(-s\delta t_o)^k}{k!} \beta_k \right] \quad 7(\text{VI-6.4})$$

which is the desired formula.

Expression 7(VI-6.4) gives immediately the complementary filter. The transfer function of the complementary filter is

$$F(s) = H(s) e^{s\delta} \sum_{k=m+1}^{\infty} \frac{(-s\delta t_o)^k}{k!} \beta_k \quad 8(\text{VI-6.4})$$

For a singular distribution,

$$F(s) = \sum_{k=m+1}^{\infty} \frac{(-s\delta t_o)^k}{k!} \beta_k \quad 9(\text{VI-6.4})$$

Finally, let us consider the form, analogous to the form of Eq. 9(VI-6.4), of the complementary filter in the case of pure transmission. Let us also consider the general formula 3(VI-6.4). The case of pure transmission is characterized by the condition

$$\sum_{p=0}^m (p+1)^k a_p = \begin{cases} 1 & \text{for } k = 0 \\ 0 & \text{for } k = 1, 2, \dots, m \end{cases}$$

Thus, the transfer function associated with transmission without delay is

$$\mathcal{L}[S_c(t)] = H(s) e^{s\delta} \left[ 1 - \sum_{k=m+1}^{\infty} \frac{(-s\delta)^k}{k!} \lambda_k \right] \quad 10(\text{VI-6.4})$$

where



(VIII. STATISTICAL COMMUNICATION THEORY)

$$-\lambda_k = \sum_{p=0}^m (p+1)^k a_p \quad k = m+1, \dots$$

and the complementary filter has the transfer function

$$F(s) = H(s) e^{s\delta} \sum_{k=m+1}^{\infty} \frac{(-s\delta)^k}{k!} \lambda_k \quad 11(\text{VI-6.4})$$

VII. ADVANCED TRANSMISSION – LINEAR PREDICTION

7.0 OBJECTIVE OF THIS CHAPTER

We shall give in condensed form some results associated with the advanced transmission of a signal  $\phi(t)$ . Some important points which are obscure will be clarified by the accompanying discussion. The mathematical similarity between the results for delayed and advanced transmission, when  $T_0$  is exchanged for  $-T_0$ , facilitates this concise presentation.

Since the effect cannot precede the cause in the physical systems considered here, the solution of the problem of prediction does not exist in an absolute sense. However, solutions of a relative character can be found, as the discussion will show.

Let  $t_\ell$  be the life of the distribution of windows  $S_C(t)$ . Let us suppose initially that the excitation  $\phi(t)$  is continuous, and is sufficiently differentiable to validate the methods used for the determination of the measures  $a$  of the window distribution, which are given by the fundamental expression 2(II-2.22). We consider two cases:

1.  $-\infty < t < t_\ell$                       ( $0 < t_\ell = \text{constant}$ )
2.  $t_\ell < t < \infty$

In the first case the response of the system,  $\gamma^*(t)$ , does not represent the function  $\gamma(t)$ , because the pertinent interval belongs to the aperture set  $[a_j]$ . In the second case the transition situation is finished. Then the window distribution begins to produce a weighted sampling, which, in turn, produces a weighted extrapolation of the function  $\phi(t)$ , taking elements from the interval  $t - t_1, t$ . The response is expressed by means of Eq. 1(III-3.32) as follows:

$$\phi(t + T_0) = \sum_{p=0}^m a_p \phi[t - (p+1)\delta]$$

which indicates that the predicted function is formed from a weighted sampling of the

present, corresponding to the window  $\alpha_o$ , and of the past of  $\phi(t)$  in the interval  $t - t_1, t$ . This is the relative sense of prediction considered here.

### 7.1 MEASURES OF THE WINDOWS

The problem of advanced transmission is characterized by the expression

$$\left. \begin{aligned} \gamma^*(t) &= \phi(t + T_o) && (t > t_\ell) \\ y_k &= \sum_{p=0}^m (p+1)^k \alpha_p = T_o^k && (k = 0, 1, \dots, m) \\ a_j &= (-1)^m \prod_{p=0}^m \left( \frac{T_o + \mu_p}{\mu_j - \mu_p} \right) = \frac{G(-T_o)}{G(\mu_j)} \end{aligned} \right\} \quad 1(\text{VII-7.1})$$

taken from section 2.5. Let us add here, for future convenience, the expression

$$\sum_{p=0}^m (p+1)^k \alpha_p = T_o^k - \beta_k \quad k = m+1, \dots \quad 2(\text{VII-7.1})$$

The third equation in Eqs. 1(VII-7.1), which gives  $a_j$ , shows that there is no intrinsic future, since for all values of  $T_o$  ( $T_o > 0$ ) all the measures  $a_j$  exist and are nonzero.

### 7.2 EXPLICIT EXPRESSION OF THE $a$ 's FOR $m = 1, 2, 3, 4$

For windows of equal apertures, if we measure the advance time in units of  $\delta$ , we have the expression

$$a_j = (-1)^m \prod_{p=0}^m \left( \frac{t_o + p + 1}{j - p} \right)$$

The  $a$ 's calculated for  $m = 1, 2, 3, 4$  yield the values given in Table 1(VII-7.1).

Figure 1(VII-7.2) shows the window distributions corresponding to the advance times  $T_o = 10$  and  $T_o = 20$ , with  $m = 3$ . Note the fast increase of the measures  $\alpha$  with  $T_o$ . For this reason, it is necessary to include linear amplifiers in the passive circuits that represent the system. For example, the practical arrangements which synthesize the system corresponding to  $T_o = 20$ , with  $m = 3$ , is shown schematically in Fig. 2(VII-7.2). The numbers indicate the gain that each amplifier must have.

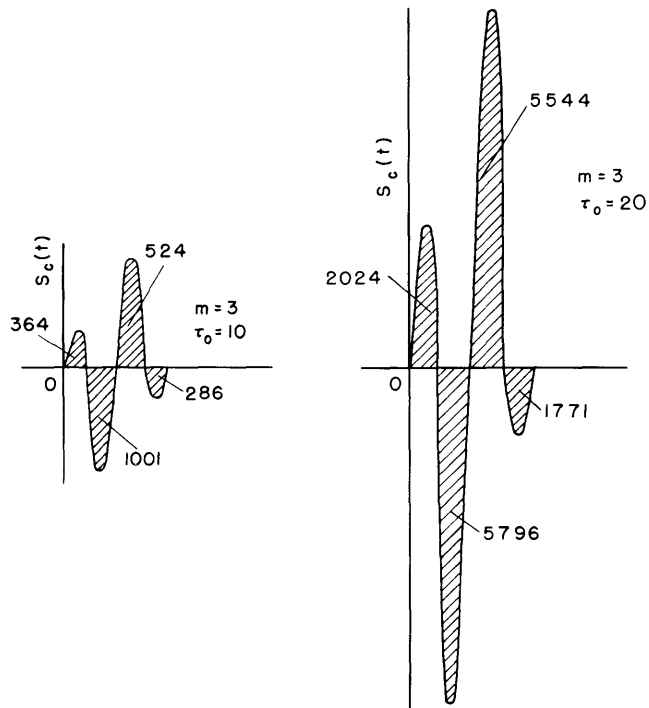


Fig. 1(VII-7.2). Window distributions corresponding to the prediction times  $T_0 = 10$  and  $T_0 = 20$ , with  $m = 3$ .

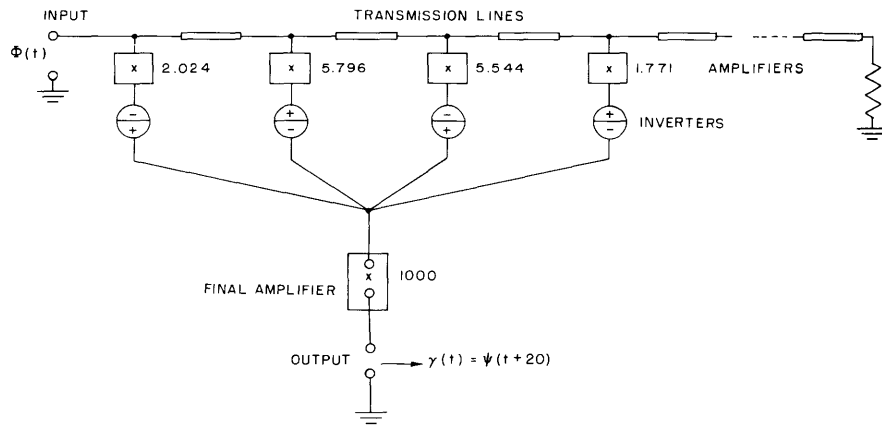


Fig. 2(VII-7.2). Synthesis of the prediction system for  $T_0 = 20$ , with  $m = 3$ .

Table 1(VII-7.1)

	m = 1	m = 2	m = 3	m = 4
$a_0$	$t_0 + 2$	$\frac{t_0^2 + 5t_0 + 6}{2}$	$\frac{t_0^3 + 9t_0^2 + 26t_0 + 24}{6}$	$\frac{t_0^4 + 14t_0^3 + 71t_0^2 + 154t_0 + 120}{24}$
$a_1$	$-1 - t_0$	$-t_0^2 + 4t_0 + 3$	$-\frac{t_0^3 + 8t_0^2 + 19t_0 + 12}{2}$	$-\frac{t_0^4 + 13t_0^3 + 59t_0^2 + 107t_0 + 60}{6}$
$a_2$		$\frac{t_0^2 + 3t_0 + 2}{2}$	$\frac{t_0^3 + 7t_0^2 + 14t_0 + 8}{2}$	$\frac{t_0^4 + 12t_0^3 + 49t_0^2 + 78t_0 + 40}{4}$
$a_3$			$-\frac{t_0^3 + 6t_0^2 + 11t_0 + 6}{6}$	$-\frac{t_0^4 + 11t_0^3 + 41t_0^2 + 61t_0 + 30}{6}$
$a_4$				$\frac{t_0^4 + 10t_0^3 + 35t_0^2 + 50t_0 + 24}{24}$

### 7.3 DETERMINATION OF THE ERROR

The procedure is identical with that for delayed transmission, with  $T_0$  exchanged for  $-T_0$ . Then the expression for tolerance 1(VI-6.3) is valid here also.

### 7.4 ASSOCIATED TRANSFER FUNCTION

The transfer function for prediction is

$$\mathcal{L}[S_c(t)] = H(s) e^{s\delta} \left[ e^{sT_0} - \sum_{k=m+1}^{\infty} \frac{(s\delta t_0)^k}{k!} \beta_k \right] \quad 1(\text{VII-7.4})$$

If the distribution is singular, we have

$$H(s) e^{s\delta} = 1$$

### 7.5 COMPLEMENTARY FILTER

The transfer function of the complementary filter is

$$F(s) = H(s) e^{s\delta} \sum_{k=m+1}^{\infty} \frac{(s\delta t_0)^k}{k!} \beta_k \quad 1(\text{VII-7.5})$$

(VIII. STATISTICAL COMMUNICATION THEORY)

If the distribution is singular, we have

$$H(s) e^{s\delta} = 1$$

VIII. DIFFERENTIATION OF SIGNALS – COMPLEMENTARY FILTERS

8.0 OBJECTIVE OF THIS CHAPTER

The theory presented in this report allows us to obtain easily a four-terminal system whose response is the  $k^{th}$  derivative of the excitation  $\phi(t)$ , when  $\phi(t)$  possesses that derivative. We limit ourselves here to two objectives:

1. Producing the measures  $a$  that characterize the appropriate distributions.
2. Applying them to complementary filters.

8.1 MEASURES OF THE DISTRIBUTION

The problem of differentiation is expressed by

$$\gamma^*(t) = \frac{d^k}{dt^k} \phi(t) \tag{VIII-8.1}$$

Formula 2(II-2.6) gives the  $a$ 's. It is repeated here for convenience

$$a_j = \frac{(-1)^k y_k s_k^{(j)}}{\prod_{p=0}^m (\mu_j - \mu_p)} \tag{VIII-8.1}$$

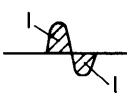
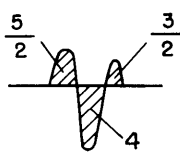
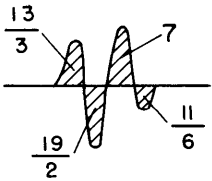
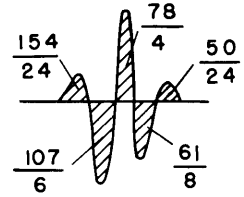
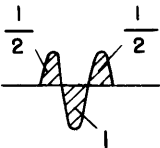
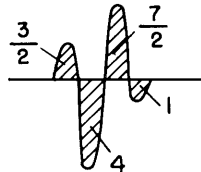
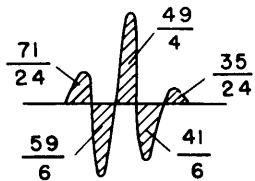
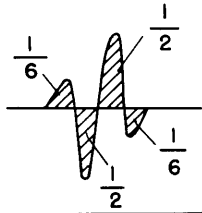
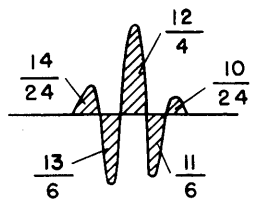
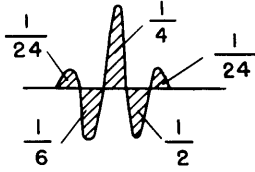
Table 1(VIII-8.1) shows the distributions for  $m = 1, 2, 3, 4$  for the first four derivatives.

8.2 APPLICATION TO THE SYNTHESIS OF COMPLEMENTARY TRANSMISSION FILTERS

The equations that produce the transfer functions of the complementary filters in the three cases of transmission studied here may be written:

$$\begin{array}{ll}
\text{Pure transmission} & F(s) = s^{m+1} \left[ H(s) e^{s\delta} M_P(s) \right] \\
\text{Delayed transmission} & F(s) = s^{m+1} \left[ H(s) e^{s\delta} M_D(s) \right] \\
\text{Advanced transmission} & F(s) = s^{m+1} \left[ H(s) e^{s\delta} M_A(s) \right]
\end{array} \tag{VIII-8.2}$$

wherein

TABLE I (VIII - 8.1)				
$S_c(t)$				
DERIVATIVE	$m = 1$	$m = 2$	$m = 3$	$m = 4$
1				
2	—			
3	—	—		
4	—	—	—	

Note: Schematic representation not drawn to scale.

(VIII. STATISTICAL COMMUNICATION THEORY)

$$M_P(s) = \sum_{j=0}^{\infty} \frac{(-\delta)^{m+1+j}}{(m+1+j)!} \lambda_{(m+1+j)} s^j$$

$$M_D(s) = \sum_{j=0}^{\infty} \frac{(-\delta t_0)^{m+1+j}}{(m+1+j)!} \beta_{(m+1+j)} s^j$$

$$M_A(s) = \sum_{j=0}^{\infty} \frac{(\delta t_0)^{m+1+j}}{(m+1+j)!} \beta_{(m+1+j)} s^j$$

We propose to synthesize, in schematic form, the systems whose transfer functions are given in Eq. 1(VIII-8.2). Because of the analytical resemblance between the functions  $M_P(s)$ ,  $M_D(s)$ , and  $M_A(s)$ , we shall consider only one of them. In the complementary filters  $H(s) e^{s\delta}$  may be considered almost equal to one. If not, we expand it in a power series in  $s$  and form the products with the series  $M_P$ ,  $M_D$ ,  $M_A$ , in order to construct a new power series, the synthesis procedure for the brackets in Eqs. 1(VIII-8.2) being similar to that for  $M_P$ ,  $M_D$ ,  $M_A$ . The problem is divided into two parts:

1. Synthesize the system corresponding to the power series. These synthesis methods are well known (1). The corresponding network is indicated by the symbol  $M(s)$  in Fig. 1(VIII-8.2).

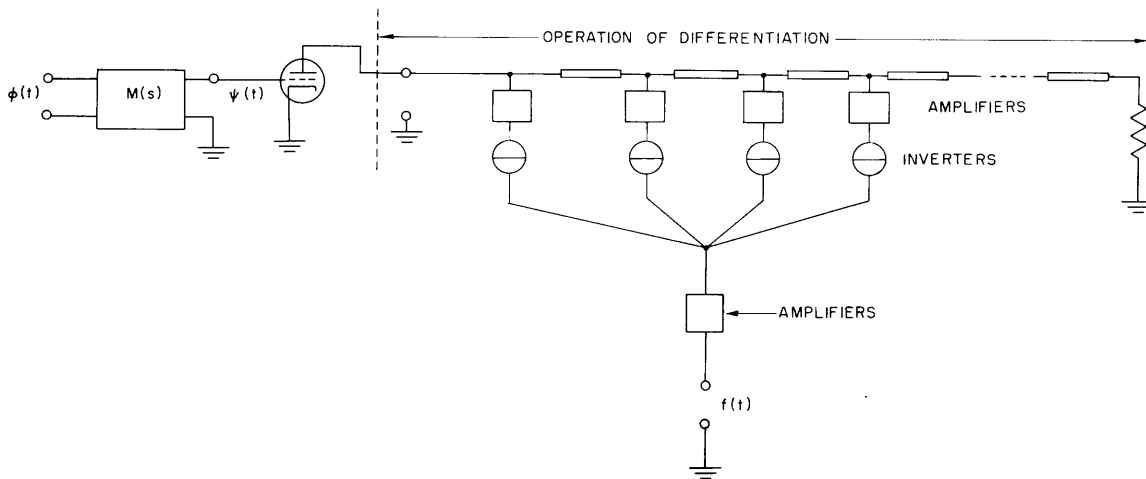


Fig. 1(VIII-8.2). Method of synthesizing the complementary filter described in section 8.2.

2. Let

$$x(t) = \mathcal{L}^{-1} M(s)$$

$$y(t) = \mathcal{L}^{-1} F(s)$$

If we assume sufficient continuity in  $x(t)$ , we have

$$y(t) = \frac{d^{m+1}}{dt^{m+1}} x(t)$$

This operation is performed by the arrangement shown in Fig. 2(I-1.51). The sketch of the method of synthesis of the complementary filters is shown in Fig. 1(VIII-8.2). The operation of differentiation is performed by using the distribution whose measures are given by expression 1(VIII-8.1).

#### 8.9 NOTE

The transmission filters described here are fundamentally low-pass filters. Their complementary systems are high-pass filters in the complementary frequency band. They are complementary with respect to frequency. With respect to time, one filter acts as a transmitter of functions that possess a modulus of oscillation which is small in the interval  $\delta$ . The complementary filter acts as an annihilator of such functions. The annihilator property allows the use of complementary filters in noise reduction.

M. V. Cerrillo

#### References

1. M. V. Cerrillo and E. F. Bolinder, On basic existence theorems in network synthesis. IV. Transmission of pulses, Technical Report 246, Research Laboratory of Electronics, M.I.T., Aug. 15, 1952.