X. STATISTICAL COMMUNICATION THEORY

Prof. Y. W. Lee
Prof. A. G. Bose
Prof. I. M. Jacobs
R. W. Burton
D. A. Chesler
A. D. Hause
K. L. Jordan, Jr.
D. J. Sakrison
M. Schetzen
D. W. Tufts
H. L. Van Trees, Jr.
C. E. Wernlein, Jr.
G. D. Zames

A. WAVEFORM SIGNALS WITH MINIMUM SPECTRAL WIDTH

We consider in this report a type of signal that consists of successive waveforms periodically and independently chosen from an orthonormal set $s_1(t), s_2(t), \ldots, s_N(t)$ with probabilities $P_1, P_2, \ldots, P_N$. In addition, the amplitudes, a_i, of the waveforms are distributed according to a probability density $p_i(a_i)$. A typical signal of this form is shown in Fig. X-1. A single waveform is shown in Fig. X-2. It is defined to be zero

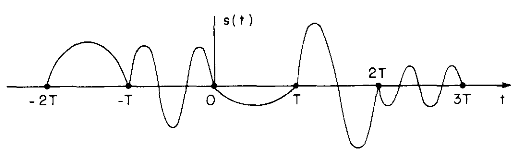

Fig. X-1. A typical waveform signal.

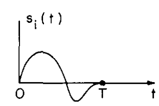

Fig. X-2. A typical waveform.

outside of the interval $[0, T]$. We shall now concern ourselves with the problem of picking the set of waveforms in such a way that the expression

$$\int_{-\infty}^{\infty} \omega^2 \, \Phi(\omega) \, d\omega \tag{1}$$

is minimized, where $\Phi(\omega)$ is the power density spectrum of the resulting stationary random process. In the following discussion we shall assume that the waveforms have continuous first derivatives and that $E[a_i] = 0$ (here, $i=1, 2, \ldots, N$) because, if $E[a_i]$ were nonzero, periodicities would occur, and $\Phi(\omega)$ would contain impulse functions. In this case it would not be apparent what set of waveforms minimizes expression 1.

We can find $\Phi(\omega)$ by assuming that the random process was derived by applying unit impulse functions to a bank of linear filters with impulse responses $s_1(t), \ldots, s_N(t)$ and adding the outputs, as shown in Fig. X-3. Impulses are successively applied to one of the N filters at a time, with probability P_i of being applied to the i^{th} filter. It is seen that the input processes are uncorrelated. Letting $\delta(\tau)$ be the unit impulse function, we obtain for the individual input correlation functions:

$$\phi_i(\tau) = \frac{P_i}{T} \, E\left[a_i^2\right] \delta(\tau) = c_i \delta(\tau)$$

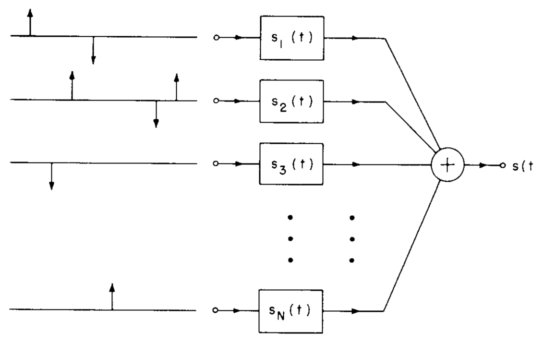

Fig. X-3. The system for deriving the signal.

where $c_i = \frac{P_i}{T}\int E\left[a_i^2\right]$. In accordance with the Wiener-Khinchin Theorem, the power density spectra are $\Phi_i(\omega) = c_i/2\pi$. It can be shown (1) that the resulting output process has a power density spectrum

$$\Phi(\omega) = \sum_{n=1}^{N} |S_n(\omega)|^2 \Phi_i(\omega) = \frac{1}{2\pi} \sum_{n=1}^{N} c_n |S_n(\omega)|^2 \tag{2}$$

where

$$S_n(\omega) = \int_{-\infty}^{\infty} s_n(t) \exp(-i\omega t)\, dt = \int_0^T s_n(t) \exp(-i\omega t)\, dt \tag{3}$$

The method used to find $\Phi(\omega)$ is similar to that used by Lee (2).

Equation 1 now takes the form

$$\begin{aligned} \int_{-\infty}^{\infty} \omega^2 \Phi(\omega) &= \frac{1}{2\pi} \sum_{n=1}^{N} c_n \int_{-\infty}^{\infty} \omega^2 |S_n(\omega)|^2\, d\omega \\ &= \frac{1}{2\pi} \sum_{n=1}^{N} c_n \int_{-\infty}^{\infty} \omega^2 S_n(\omega)\, S_n^*(\omega)\, d\omega \end{aligned} \tag{4}$$

where the asterisk denotes the complex conjugate.

In order for the integral of Eq. 1 to converge, it is necessary that

$$\omega^2 \Phi(\omega) = O(|\omega|^{-k}) \tag{5}$$

for large ω, with $k > 1$ (see refs. 3 and 4). Then,

$$\Phi(\omega) = O(|\omega|^{-k-2}) \tag{6}$$

and from Eq. 4

$$|S_n(\omega)|^2 = O(|\omega|^{-k-2}) \qquad n = 1, \ldots, N \tag{7}$$

or

$$|S_n(\omega)| = O(|\omega|^{(-k/2)-1}) \qquad n = 1, \ldots, N \tag{8}$$

with $k > 1$.

We shall now show that in order for Eq. 8 to hold, it is necessary that

$$s_n(0) = s_n(T) = 0 \qquad n = 1, \ldots, N \tag{9}$$

Integrating Eq. 3 by parts, we get

$$S_n(\omega) = \frac{s_n(0) - s_n(T) \exp(-i\omega T)}{i\omega} + \frac{1}{i\omega} \int_0^T s_n'(t) \exp(-i\omega t)\, dt \tag{10}$$

Since the $s_n'(t)$ are continuous on [0, T], they are bounded (5), and $|s_n'(t)| \leq K$ for $0 \leq t \leq T$, for some number K. It follows that

$$\left| \frac{1}{i\omega} \int_0^T s_n'(t) \exp(-i\omega t)\, dt \right| \leq \frac{K}{|\omega|} \frac{|1-\exp(-i\omega T)|}{|\omega|} = O(|\omega|^{-2}) \tag{11}$$

Unless the conditions of Eq. 9 hold, it is seen that

$$|S_n(\omega)| = O(|\omega|^{-1}) \tag{12}$$

which violates Eq. 8.

Since the $s_n'(t)$ exist on [0, T] and the conditions of Eq. 9 hold, the Fourier transforms of $s_n'(t)$ exist and are $(-i\omega)\, S_n(\omega)$. Parseval's Formula then holds, and

$$\int_{-\infty}^{\infty} (-i\omega)\, S_n(\omega)(i\omega)\, S_n^*(\omega)\, d\omega = 2\pi \int_{-\infty}^{\infty} [s_n'(t)]^2\, dt = \int_{-\infty}^{\infty} \omega^2 |S_n(\omega)|^2\, d\omega \tag{13}$$

The minimization problem is then reduced to the minimization of

$$\sum_{n=1}^{N} c_n \int_0^T [s_n'(t)]^2\, dt \tag{14}$$

under the constraints that $\{s_n(t)\}$ be an orthonormal set and $s_n(0) = s_n(T) = 0$ for all $n = 1, \ldots, N$.

We now assume that the c_i are arranged in the following manner: $c_1 \geq c_2 \geq \ldots \geq c_N$. It can then be shown by the calculus of variations (6) that the minimization of expression 14 is achieved by the first N solutions of the differential equation

$$s''(t) + \lambda s(t) = 0 \tag{15}$$

with the boundary conditions

$$s(0) = s(T) = 0$$

These solutions are

$$s_n(t) = \left(\frac{2}{T}\right)^{1/2} \sin \frac{n\pi}{T} t \qquad n = 1, \ldots, N \tag{16}$$

for which

$$|S_n(\omega)|^2 = \frac{8\pi^2 n^2 T}{(n^2\pi^2-\omega^2T^2)^2} \cos^2 \frac{\omega T}{2} \qquad n \text{ odd}$$

$$= \frac{8\pi^2 n^2 T}{(n^2\pi^2-\omega^2T^2)^2} \sin^2 \frac{\omega T}{2} \qquad n \text{ even} \tag{17}$$

From Eqs. 2 and 17 the power density spectrum $\Phi(\omega)$ then becomes

$$\Phi(\omega) = 4\pi\left(\cos^2 \frac{\omega T}{2}\right) \sum_{n \text{ odd}} \frac{P_n E\left[a_n^2\right] n^2}{(n^2\pi^2-\omega^2T^2)^2} + 4\pi\left(\sin^2 \frac{\omega T}{2}\right) \sum_{n \text{ even}} \frac{P_n E\left[a_n^2\right] n^2}{(n^2\pi^2-\omega^2T^2)^2} \tag{18}$$

Several examples of power density spectra for signals using the waveforms of Eq. 16 with T = 1 are shown in Fig. X-4. In these examples it was assumed that the waveforms

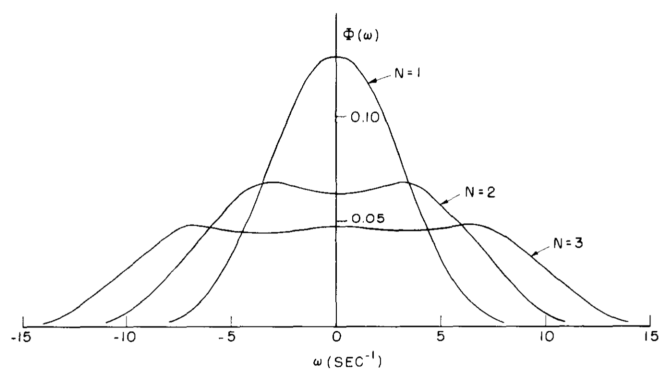

Fig. X-4. Power density spectrum for the optimum waveforms (T=1).

were equally likely with $E\left[a_i^2\right] = 1$, $i = 1, 2, \ldots, N$. It is interesting that the bandwidths of the spectra shown in Fig. X-4 are approximately equal to those that might be predicted from the Sampling Theorem; i.e.,

$$\omega_o \approx \frac{N\pi}{T} \tag{19}$$

where N is the number of degrees of freedom of the signal, and ω_o is the bandwidth in radians per second. This gives 3.14, 6.28, and 9.42 for n = 1, 2, and 3, respectively.

These results can be further generalized to the minimization of the integral

$$\int_{-\infty}^{\infty} \omega^{2n} \Phi(\omega)\, d\omega \tag{20}$$

where n is any integer. In this case the waveforms will be the first N solutions to the differential equation

$$s^{(2n)}(t) + \lambda s(t) = 0$$

where (n) denotes differentiation n times, with the boundary conditions

$$\begin{aligned} s^{(n-1)}(0) &= s^{(n-1)}(T) = 0 \\ s^{(n-2)}(0) &= s^{(n-2)}(T) = 0 \\ &\vdots \\ s(0) &= s(T) = 0 \end{aligned}$$

K. L. Jordan, Jr.

References

1. W. B. Davenport and W. L. Root, An Introduction to the Theory of Random Signals and Noise (McGraw-Hill Publishing Co., New York, 1958), p. 184.

2. Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., Jan. 15, 1955, pp. 56-62.

3. R. Courant, Differential and Integral Calculus (Interscience Publishers, Inc., New York, 1937, 2nd Edition), Volume 1, p. 250.

4. For a definition of O(f(x)), see H. Cramér, Methods of Mathematical Statistics (Princeton University Press, 1946), p. 122.

5. W. Rudin, Principles of Mathematical Analysis (McGraw-Hill Publishing Co., New York, 1953), p. 67, Theorem 4.12.

6. The methods used to prove this are given in Chap. VI of R. Courant and D. Hilbert, Methods of Mathematical Physics (Interscience Publishers, Inc., New York, 1953). Note especially the theorems of § 1, sec. 4 and § 7, sec. 3.
See also, K. Fan, "On A Theorem of Weyl concerning Eigenvalues of Linear Transformations," Proceedings of the National Academy of Sciences, vol. 35, pp. 652-655, 1949.

B. MEASUREMENT OF CORRELATION FUNCTIONS

A limitation to the use of higher-order correlation functions is the difficulty of their determination. A practical and relatively simple method of measuring them without the use of delay lines is presented in this report. To explain the method, the measurement of a second-order crosscorrelation function will be discussed.

Let us consider the second-order crosscorrelation function

$$R_{abc}(\tau_1, \tau_2) = \lim_{T\to\infty} \frac{1}{T} \int_0^T f_a(t)\, f_b(t+\tau_1)\, f_c(t+\tau_2)\, dt \tag{1}$$

Under some general conditions, it can be shown (1) that this function is absolutely integrable and continuous for all values of τ_1 and τ_2. Thus it generally may be represented as

$$R_{abc}(\tau_1, \tau_2) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} A_{ij}\, \phi_i(\tau_1)\, \phi_j(\tau_2) \tag{2}$$

where $\{\phi_n(x)\}$ is a complete set of orthonormal functions:

$$\int_{-\infty}^{\infty} \phi_i(x)\, \phi_j(x)\, dx = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \tag{3}$$

The coefficients, A_{ij}, are given by

$$A_{ij} = \int_{-\infty}^{\infty} \phi_i(\tau_1)\, d\tau_1 \int_{-\infty}^{\infty} \phi_j(\tau_2)\, d\tau_2\, R_{abc}(\tau_1, \tau_2) \tag{4}$$

We now restrict our attention to those sets that are orthonormal over the interval $[0, \infty]$ and realizable as the impulse responses of linear networks. The Laguerre functions are one such set. This set and others with their network realizations are given by Lee (2). For these sets, Eq. 2 is valid only for the region $\tau_1 \geq 0$, $\tau_2 \geq 0$. We shall find, however, that this does not restrict our determination of $R_{abc}(\tau_1, \tau_2)$ over the whole τ_1-τ_2 plane.

We shall now present an experimental procedure for determining the coefficients, A_{ij}. Consider the circuit shown in Fig. X-5. For this circuit, the network with the impulse response $\phi_j(t)$ has the input $f_a(t)$ and the output $g_a(t)$; the network with the impulse response $\phi_i(t)$ has the input $f_b(t)$ and the output $g_b(t)$. We shall denote the average product of $g_a(t)\, g_b(t)\, f_c(t)$ by B_{ij}. Then,

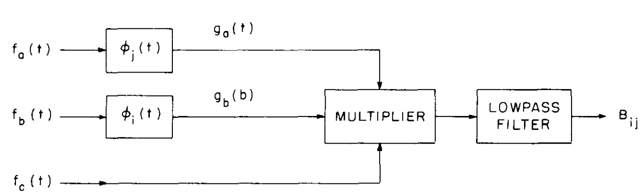

Fig. X-5. Circuit for the experimental determination of B_{ij}.

$$B_{ij} = \lim_{T\to\infty} \frac{1}{T} \int_0^T g_a(t)\, g_b(t)\, f_c(t)\, dt$$

$$= \lim_{T\to\infty} \frac{1}{T} \int_0^T f_c(t)\, dt \int_{-\infty}^{\infty} \phi_j(x)\, f_a(t-x)\, dx \int_{-\infty}^{\infty} \phi_i(y)\, f_b(t-y)\, dy$$

$$= \int_{-\infty}^{\infty} \phi_j(x)\, dx \int_{-\infty}^{\infty} \phi_i(y)\, dy \lim_{T\to\infty} \frac{1}{T} \int_0^T f_a(t-x)\, f_b(t-y)\, f_c(t)\, dt$$

$$= \int_{-\infty}^{\infty} \phi_j(x)\, dx \int_{-\infty}^{\infty} \phi_i(y)\, dy\, R_{abc}(x-y, x) \tag{5}$$

The interpretation of Eq. 5 is simpler if we let $x = \tau_2$ and let $x - y = \tau_1$. Equation 5 then becomes

$$B_{ij} = \int_{-\infty}^{\infty} \phi_j(\tau_2)\, d\tau_2 \int_{-\infty}^{\infty} \phi_i(\tau_2 - \tau_1)\, d\tau_1\, R_{abc}(\tau_1, \tau_2) \tag{6}$$

By comparing Eq. 6 with Eq. 4, we observe that the average products, B_{ij}, are the coefficients in the expansion

$$R_{abc}(\tau_1, \tau_2) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} B_{ij}\, \phi_i(\tau_2 - \tau_1)\, \phi_j(\tau_2) \qquad \tau_2 \geq \tau_1;\ \tau_2 \geq 0 \tag{7}$$

Since the orthonormal functions are realizable as network functions, Eq. 7 may be realized as the double-impulse response of a network. Consider the element network shown in Fig. X-6. The unit impulse, $\mu_o(t)$, applied to the network with the impulse response $\phi_j(t)$ occurs δ seconds ahead of the unit impulse applied to the network with the impulse response $\phi_i(t)$. The responses are multiplied and amplified by an amplifier with a gain $A = B_{ij}$. The double-impulse response, $h_{ij}(t, \delta)$, is

$$h_{ij}(t, \delta) = B_{ij}\, \phi_i(t-\delta)\, \phi_j(t) \qquad t \geq \delta;\ t \geq 0 \tag{8}$$

By summing the responses of such element networks, we obtain

$$H_1(t, \delta) = \sum_i \sum_j B_{ij}\, \phi_i(t-\delta)\, \phi_j(t) \qquad t \geq \delta;\ t \geq 0 \tag{9}$$

This is Eq. 7 in which $\delta = \tau_1$ and $t = \tau_2$. Thus, for each δ, the double-impulse response, $H_1(t, \delta)$, yields the second-order crosscorrelation function along a line in region 1 of the τ_1-τ_2 plane as shown in Fig. X-7.

Now consider the average product, C_{ij}, for the circuit shown in Fig. X-8. In a manner analogous to the derivation of Eq. 6, it can be shown that

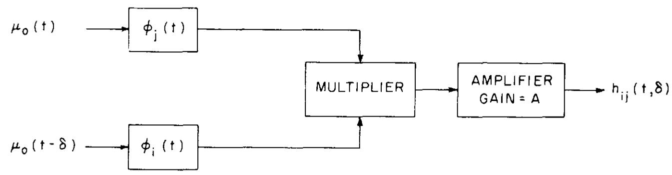

Fig. X-6. Basic circuit element for the synthesis of $R_{abc}(\tau_1, \tau_2)$.

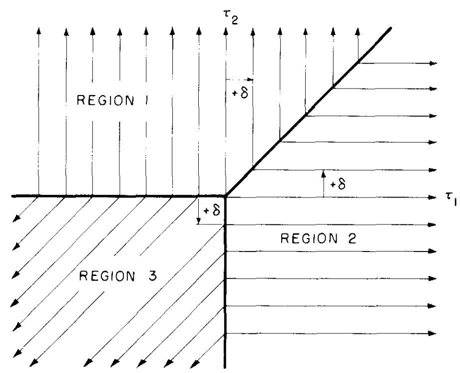

Fig. X-7. Lines of double-impulse responses in the τ_1-τ_2 plane (arrows point in the direction of increasing t).

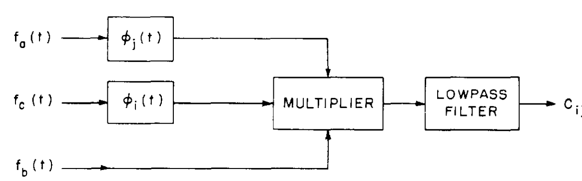

Fig. X-8. Circuit for the experimental determination of C_{ij}.

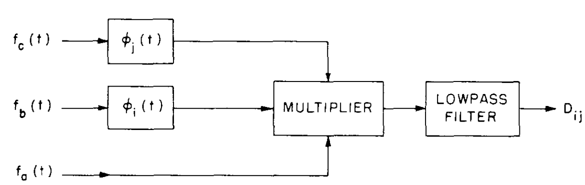

Fig. X-9. Circuit for the experimental determination of D_{ij}.

$$C_{ij} = \int_{-\infty}^{\infty} \phi_j(\tau_1)\, d\tau_1 \int_{-\infty}^{\infty} \phi_i(\tau_1-\tau_2)\, d\tau_2\, R_{abc}(\tau_1, \tau_2) \tag{10}$$

By comparing Eq. 10 with Eq. 4, we observe that the average products, C_{ij}, are the coefficients in the expansion

$$R_{abc}(\tau_1, \tau_2) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} C_{ij}\, \phi_i(\tau_1-\tau_2)\, \phi_j(\tau_1) \qquad \tau_1 \geq \tau_2;\ \tau_1 \geq 0 \tag{11}$$

Thus, by adjusting the amplifier of the circuit shown in Fig. X-6 to be $A = C_{ij}$, the summed double-impulse response becomes

$$H_2(t, \delta) = \sum_i \sum_j C_{ij}\, \phi_i(t-\delta)\, \phi_j(t) \qquad t \geq \delta;\ t \geq 0 \tag{12}$$

This is Eq. 11 in which $\delta = \tau_2$, and $t \doteq \tau_1$. Thus, for each δ, the double-impulse response, $H_2(t, \delta)$, yields the second-order crosscorrelation function along a line in region 2 of the τ_1-τ_2 plane as shown in Fig. X-7.

Finally, consider the average product, D_{ij}, for the circuit shown in Fig. X-9. In a manner analogous to the derivation of Eq. 6, it can be shown that

$$D_{ij} = \int_{-\infty}^{\infty} \phi_j(-\tau_2)\, d\tau_2 \int_{-\infty}^{\infty} \phi_i(-\tau_1)\, d\tau_1\, R_{abc}(\tau_1, \tau_2) \tag{13}$$

By comparing Eq. 13 with Eq. 4, we observe that the average products, D_{ij}, are the coefficients in the expansion

$$R_{abc}(\tau_1, \tau_2) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} D_{ij}\, \phi_i(-\tau_1)\, \phi_j(-\tau_2) \qquad \tau_1 \leq 0;\ \tau_2 \leq 0 \tag{14}$$

Thus, by adjusting the amplifier of the circuit shown in Fig. X-6 to be $A = D_{ij}$, the summed double-impulse response becomes

$$H_3(t, \delta) = \sum_i \sum_j D_{ij}\, \phi_i(t-\delta)\, \phi_j(t) \qquad t \geq \delta;\ t \geq 0 \tag{15}$$

This is Eq. 14 in which $\delta = \tau_1 - \tau_2$, and $t = -\tau_2$. Thus, for each δ, the double-impulse response, $H_3(t, \delta)$, yields the second-order crosscorrelation function along a line in region 3 of the τ_1-τ_2 plane as shown in Fig. X-7.

In this manner, the second-order crosscorrelation function over the whole τ_1-τ_2 plane may be obtained experimentally. Note that if $f_b(t) = f_c(t)$, then $B_{ij} = C_{ij}$, and only two sets of measurements are required. For the measurement of the second-order autocorrelation function $f_a(t) = f_b(t) = f_c(t)$, and thus only one set of measurements is required. The second-order correlation function can be approximated, with minimum mean-square error, to any degree of accuracy by using N members of the orthonormal

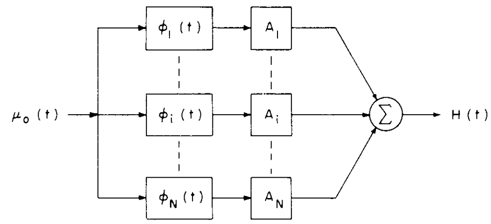

Fig. X-10. The network synthesis of $R_{ab}(\tau)$.

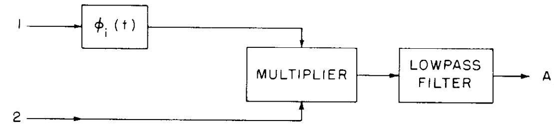

Fig. X-11. Circuit for the experimental determination of A_i.

set. For such a representation, the second-order crosscorrelation function can be experimentally determined over the whole τ_1-τ_2 plane by, at most, $3N^2$ measurements.

It is now clear that this procedure can be extended to all orders of correlation functions. Thus, if the third-order correlation function exists and is absolutely integrable, it can be experimentally obtained as the triple-impulse response of a network. For such a determination, $4N^3$ measurements, at most, would be required if N members of the orthonormal set are used. A special case of this general procedure, which has been described by Lampard (3), is the determination of the first-order crosscorrelation function. For this case, the single-impulse responses of two networks of the form shown in Fig. X-10 would be required – that is, for $t \geq 0$, $H_1(t) = R_{ab}(t)$ and $H_2(t) = R_{ab}(-t)$. The coefficients, A_i, are experimentally determined by means of the circuit shown in Fig. X-11. However, for this special case, we note that since

$$R_{ab}(t) = H_1(t) + H_2(-t) \qquad -\infty < t < \infty \tag{16}$$

the cross power density spectrum is

$$\Phi_{ab}(\omega) = F_1(\omega) + F_2^*(\omega) \tag{17}$$

where $F(\omega)$ is the Fourier transform of $H(t)$, and the star indicates the complex conjugate. Thus the cross power density spectrum may be obtained directly by measuring the transfer function of each of the two networks. For an autocorrelation function, the power density spectrum is just twice the real part of $F_1(\omega)$. Experimentally, this is obtained by measuring the inphase component of the sine-wave response of the network.

M. Schetzen

References

1. J. Y. Hayase, Properties of Second-Order Correlation Functions, Technical Report 330, Research Laboratory of Electronics, M.I.T. (May 17, 1957).

2. Y. W. Lee, The Statistical Theory of Communication (John Wiley and Sons, Inc., New York, 1960).

3. D. G. Lampard, A New Method of Determining Correlation Functions of Stationary Time Series, Proc. I.E.E. (London), Monograph No. 104R, August 1954 (102C, pp. 35-41).

C. DESIGN OF BINARY, SYNCHRONOUS PULSE-TRANSMISSION LINKS

1. Model for a Synchronous Pulse-Transmission Link

A diagram of our model for one link of a synchronous pulse-transmission system is shown in Fig. X-12. A synchronous pulse-transmission system (e.g., a pulse code modulation (PCM) system) would usually include a long chain of links. In each link the pulse signals are sent through some transmission line (e.g., a length of waveguide or a distance in air). At the receiving end the message is recovered from the received signal, and new pulse signals are formed for the next link of the system. The adjective "synchronous" is intended to imply that, ideally, the transmitter and receiver operate periodically with a constant difference in phase.

In the model of Fig. X-12 the message generator produces a positive integer from the set (1, 2, ..., M) every T seconds. Each integer identifies which of M possible messages is to be transmitted. In a binary link, M = 2.

The coder, C, instantaneously substitutes a fixed real number, a_i (to be determined later) for each message i; i = 1, 2, ..., M. We assume that any more complicated coding (such as block coding or error-detection coding) has been included in the message generator. The output of C is a discrete time series, ..., A_{-2}, A_{-1}, A_0, A_1, A_2, ..., where A_i is the particular output that occurs during the time interval (iT, (i+1)T); i = 0, ±1, ±2, ...

The error noise generator, N_1, represents the changes that have taken place in the

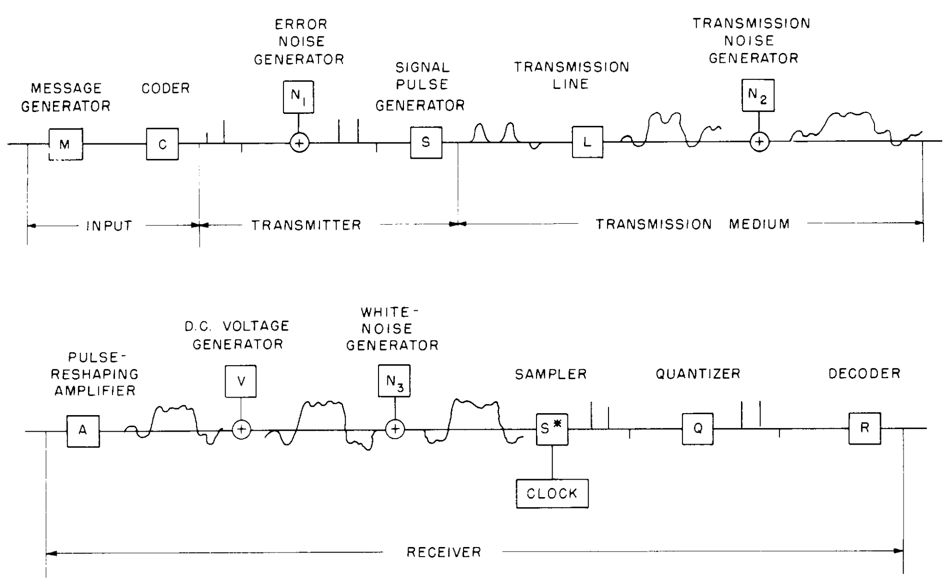

Fig. X-12. Model for one link of a synchronous pulse-transmission system.

original message time series as a result of its passage through all of the previous links of the system. The outputs of C and N_1 occur simultaneously. The noise values are picked from the set $\{a_i - a_j\}$, where i and j independently assume the values 1, 2, ..., M. The output of N_1 is a discrete time series, ..., X_{-2}, X_{-1}, X_o, X_1, X_2, ..., where X_i is the output during the time interval $(iT, (i+1)T)$; $i = 0, \pm 1, \pm 2, \ldots$.

Thus, the input to the signal pulse generator, S, is the sequence of real numbers (B_i), $i = 0, \pm 1, \pm 2, \ldots$, where $B_i = A_i + X_i$. The output of S is $\sum_{n=-\infty}^{\infty} B_n\, s(t-nT)$, where s(t) is a standard pulse that is integrable square.

The linear, time-invariant transmission line, L, has a unit impulse response $\ell(t)$. It operates on the output of the pulse generator to form

$$\sum_{n=-\infty}^{\infty} B_n\, r_1(t-nT)$$

where $r_1(t) = \int_{-\infty}^{\infty} s(u)\, \ell(t-u)\, du$. To this waveform is added $n_2'(t)$, a member of the ergodic noise ensemble $[n_2'(t)]$ with correlation function $n_2(u)$. The transmission noise generator, N_2, represents the effects of noise in the line. The resulting waveform at the input to A corresponds to the received signal in a pulse transmission system.

The linear, time-invariant, pulse-reshaping amplifier, A, has the unit impulse response a(t). It operates on this combination of signal pulses and noise. Its output is

$$\sum_{n=-\infty}^{\infty} B_n\, r_2(t-nT) + n_2^*(t)$$

where $r_2(t) = \int_{-\infty}^{\infty} r_1(u)\, a(t-u)\, du$ and $n_2^*(t) = \int_{-\infty}^{\infty} n_2'(u)\, a(t-u)\, du$. To this output is added a dc voltage that will offset the average noise level.

Noise generator N_3 adds white noise that represents the thermal noise in the amplifier and sampler.

The time axis is divided into intervals of width T, called measuring intervals, and S^* samples its input once and only once during each measuring interval. The exact sampling times are determined by impulses from the clock. [More detailed assumptions concerning the clock operation will be made later.]

The quantizer, Q, sorts the samples sequentially into M voltage intervals, each corresponding to one unique message. From the quantizer output, the decoder, R, produces a new message integer sequence according to this correspondence. (This sequence may be used directly or it may become the input for another link of the system.) The sampling, quantizing, and decoding will be considered to be instantaneous operations.

It is important to note that, in our synchronous pulse transmission system, time, as well as voltage, is quantized. There are only M possible messages, and each message is alloted T seconds at the transmitter and T seconds at the receiver. We shall say that the link operates perfectly if the decoded message sequence is the same as the message generator sequence (except for a fixed time delay). The probability that these sequences will differ in any one place is called the error probability.

Let us now assume that, in our model of Fig. X-12, the messages represented by the output integers of the message generator are coded, quantized samples of a band-limited signal. The signal energy is assumed to be essentially confined to the frequency interval $\left(\frac{-1}{2T_o}, \frac{1}{2T_o}\right)$, and samples of the signal are instantaneously taken every T_o seconds. As shown in Fig. X-13, the samples are sorted into L voltage intervals by the quantizer. Each quantized sample is then coded into a sequence of g integers from the set $(1, 2, \ldots, M)$. An example of such a code is given in Fig. X-13 for M = 2. In order that there shall be a one-to-one correspondence between the possible sequences of length g and the possible quantized voltages that occur, the following equation must hold:

$$L = M^g \tag{1}$$

It follows that $gT = T_o$.

If we assume that the message integers of Fig. X-12 are produced as outlined in the two preceding paragraphs (that is, M of Fig. X-12 is composed of the blocks depicted in Fig. X-13), then the model of Fig. X-12 becomes a model for one link of a PCM system.

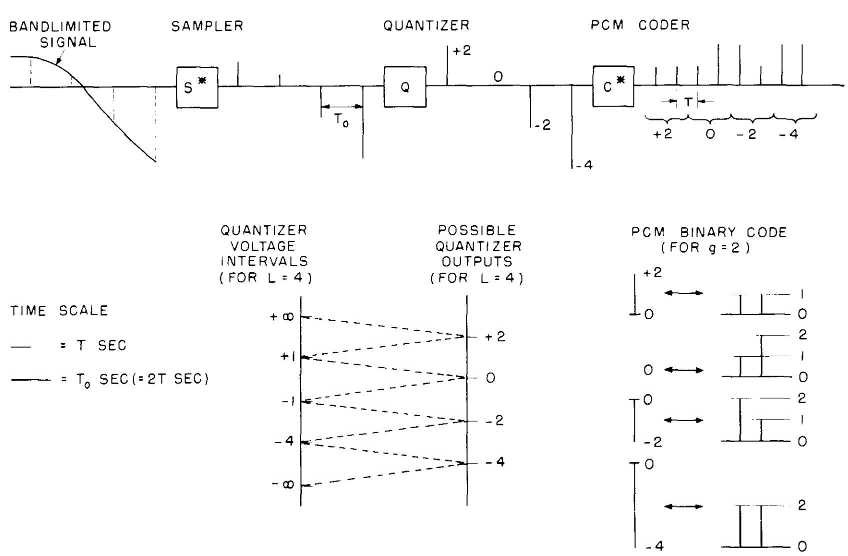

Fig. X-13. Message source in a PCM link.

2. An Upper Bound to the Error Probability

Let us denote the actual waveform at the input to the sampler of Fig. X-12 by $f_a(t)$. We define a desired waveform at the same point in the link by

$$f_d(t) = \sum_{j=-\infty}^{\infty} A_j d(t-jT) \tag{2}$$

where $\{A_j\}$ is the coded message time series;

$$d(t) = \begin{cases} d & \text{when } p(t-bT) > 0 \\ 0 & \text{for all other time instants} \end{cases}; \tag{3}$$

the positive real number d is a measure of the voltage gain of the link; and the positive integer b is the time delay of the link in units of T seconds. (The actual time delay from message generator to quantizer is a random variable with possible values between bT and (b+1)T.)

The probability density functions $\{p(t-jT)\}$, $j = 0, \pm 1, \pm 2, \ldots,$ characterize the operation of the clock of Fig. X-12. That is,

$$\int_{-\infty}^{t} p(u-jT)\, du$$

is the probability that the sampling instant associated with the measuring interval $(jT, (j+1)T)$ occurs before time t.

We call $f_d(t)$ a desired waveform because we shall always choose the pertinent design parameters (the set of coded message values $\{a_i\}$, the "voltage gain" d, and the set of quantizer voltage intervals) in such a way that

(a) the only possible values for $f_d(t)$ are zero and the center voltages (so that we may use Tchebychef's theorem as shown below) of each voltage quantization interval; and

(b) if $f_d(t)$ were the actual sampler input, then the decoded output message sequence will be the same as the input message sequence (except for a finite time delay of bT seconds).

Using this notation, we can say that an error occurs whenever

$$|f_d(t_s)-f_a(t_s)| > \frac{1}{2} W_i \tag{4}$$

where t_s is a time instant at which the sampler operates; W_i is the "voltage width" of the quantizer interval in which $f_d(t_s)$ lies; and $f_d(t_s)$ decodes into message i.

The error probability can be written as

$$P_E = \sum_{i=1}^{M} P_r(|y_i(t_s)| > k_i\sigma_i)\, p'(i) \tag{5}$$

where $p'(i)$ is the probability of occurrence of the i^{th} message; $P_r(|y_i(t_s)|>k_i\sigma_i)$ is the probability that $|y_i(t_s)|$ exceeds $k_i\sigma_i$; the quantity σ_i is the standard deviation of the random variable $y_i(t_s) = f_d(t_s) - f_a(t_s)$, given that $f_d(t_s)$ decodes into message i; and the real numbers k_i, $i = 1, 2, \ldots, M$, are selected in such a way that $k_i\sigma_i = W_i/2$.

We obtain our upper bound, B, by applying Tchebychef's theorem to each term on the right-hand side of Eq. 5. This theorem states that if x is a random variable with mean m and variance σ^2, and if P is the probability that $|x-m|$ exceeds $k\sigma$, then P does not exceed $1/k^2$.

Thus, if the conditional means $\{m_i=E[y_i(t_s)]\}$ (with $E[x]$ denoting the ensemble average of x) are zero for $i = 1, 2, \ldots, M$, then we may say that

$$P_E \leqslant \sum_{i=1}^{M} \frac{p'(i)}{k_i^2} = B \tag{6}$$

3. Design of a Binary Link

To design a link with an upper bound to the error probability as expressed by Eq. 6, the following assumptions are made:

(a) The transmission line of Fig. X-12 is bandlimited to frequencies of less than W cps. Its input impedance is a constant resistance, R.

(b) The message rate, 1/T, is not less than the Nyquist rate, 2W.

(c) Every member of the coded message time series $\{A_n\}$, $n = 0, \pm1, \pm2, \ldots$ is statistically independent of all other members.

(d) The coder operates in such a way that

$$E\left[A_n^2\right] = 1 \qquad \text{for } n = 0, \pm1, \pm2, \ldots \tag{7}$$

and

$$E[A_n] = 0 \qquad \text{for } n = 0, \pm1, \pm2, \ldots \tag{8}$$

(e) There are only two messages, M = 2. The occurrences of these messages are equally likely.

(f) The error noise generator is removed from the model of Fig. X-12.

(g) The average power entering the transmission line is E watts.

(h) The voltage range of the quantizer is $2V_2$ volts.

(i) The two noise random variables are statistically independent of each other and of the message random variable.

Assumptions (a) and (e) imply that $a_1 = +1$ and $a_2 = -1$. This defines the coder.

If we now reduce our upper bound B as much as possible by varying the widths of the receiver quantization intervals (subject to the constraint of assumption (h) above), we find, by elementary calculus of variations, that each quantization interval should be V_2 volts wide. That is,

$$W_1 = W_2 = V_2 \tag{8}$$

Our definition of $f_d(t)$ (or the linearity of the transmission line and amplifier) shows that the two quantization intervals must be positioned symmetrically about zero volts. Hence our voltage gain d (cf. Eq. 3) must be equal to $V_2/2$. Thus, we have defined the quantizer of Fig. X-12.

We now wish to choose the input pulse shape, s(t), and the amplifier impulse response, a(t). This must be done in such a way that the conditional means, m_1 and m_2, are each zero. As mentioned prior to Eq. 6, each conditional mean must be zero if B is to be a valid upper bound to the error probability.

If we calculate m_1 and m_2, using assumptions (c), (d), and (f) of this section, we find that sufficient conditions for $m_1 = m_2 = 0$ are

$$V = -E\left[n_2^*(t)\right] - E[n_3(t)] \tag{9}$$

and

$$\int_0^T p(t)\, h(t+bT)\, dt = d \tag{10}$$

where

$$h(t) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} s(x)\, \ell(y-x)\, dx\right] a(t-y)\, dy \tag{11}$$

Equation 9 is satisfied by the correct choice of our dc bias voltage, V. Equation 10 must be used as a constraint when we optimize s(t) and a(t) by minimizing B.

Our next step is to rewrite Eq. 6, using assumption (e) and Eq. 8:

$$B = \frac{1}{2} \sum_{i=1}^{2} \frac{1}{k_i^2} = \frac{1}{2} \sum_{i=1}^{2} \frac{\sigma_i^2}{(k_i\sigma_i)^2} = 2 \sum_{i=1}^{2} \frac{\sigma_i^2}{W_i^2} = \frac{2}{V_2^2} \sum_{i=1}^{2} \sigma_i^2 \tag{12}$$

If we calculate σ_1^2 and σ_2^2, we find that

$$\sigma_1^2 = \sigma_2^2 = \sigma^2 = \int_{-\infty}^{\infty} \left[\sum_{n=-\infty}^{\infty} p(t-nT)\right] [d(t)-h(t)]^2\, dt + \int_{-\infty}^{\infty} a(\sigma)\, d\sigma \int_{-\infty}^{\infty} a(\tau)\, n_2(\sigma-\tau)\, d\tau$$

$$+ E\left([n_3(t)]^2\right) - (E[n_3(t)])^2 - \left(E\left[n_2^*(t)\right]\right)^2 \tag{13}$$

where σ^2 is the variance of the random variable $y(t_s) = f_d(t_s) - f_a(t_s)$. Hence,

$$B = \frac{4\sigma^2}{V_2^2} \tag{14}$$

Since V_2, the maximum permissible quantizer voltage, is held constant, minimization of σ^2 is equivalent to minimization of B. We now minimize σ^2 by varying s(t) and a(t) subject to the constraints of Eq. 10 and assumption (g).

If B does attain a minimum value as we vary s(t) and a(t), then, by Fourier and variational analysis, their Fourier transforms, S(f) and A(f), must satisfy the following phase and magnitude conditions:

$$\theta_S(f) + \theta_L(f) + \theta_A(f) = \theta_P(f) - 2\pi bTf \tag{15}$$

$$|S(f)|^2 = \frac{T(d+\mu)|P(f)|}{|L(f)|}\left(\frac{N_2(f)}{\lambda}\right)^{1/2} - \frac{TN_2(f)}{|L(f)|^2} \tag{16}$$

$$|A(f)|^2 = \frac{T(d+\mu)|P(f)|}{|L(f)|}\left(\frac{\lambda}{N_2(f)}\right)^{1/2} - \frac{T\lambda}{|L(f)|^2} \tag{17}$$

These equations hold only for values of f that are such that

$$(d+\mu)|P(f)| \geq \frac{(\lambda N_2(f))^{1/2}}{|L(f)|} \tag{18}$$

Otherwise,

$$S(f) = A(f) = 0 \tag{19}$$

In Eqs. 15 to 19, the frequency functions denoted by capital letters are the Fourier transforms of the time functions denoted by the corresponding lower-case letters and $S(f) = |S(f)| \exp(j\theta_S(f))$, $P(f) = |P(f)| \exp(j\theta_P(f))$, and so on. The parameters λ and μ are Lagrange multipliers that must be chosen in such a way that our constraints are satisfied. In the frequency domain our constraint equations are

$$\int_{-\infty}^{\infty} P(f)\,\overline{H(f)}\,\exp(-j2\pi bTf)\,df = d \tag{20}$$

and

$$\frac{1}{T}\int_{-\infty}^{\infty} R|S(f)|^2\,df = E \tag{21}$$

where

$$H(f) = S(f)\,L(f)\,A(f) \tag{22}$$

(Our expression for the average power, E, is simple because of assumptions (c), (d), and (f).)

The transfer function from coder to quantizer, H(f), is found from Eqs. 15, 16, 17, 18, and 22. We have

$$\theta_H(f) = \theta_P(f) - 2\pi bTf \tag{23}$$

and

$$|H(f)| = T\left[(d+\mu)|P(f)| - \frac{(\lambda N_2(f))^{1/2}}{|L(f)|}\right] \tag{24}$$

provided that inequality 18 is satisfied. Otherwise

$$H(f) = 0 \tag{25}$$

To gain some insight into what behavior our criterion demands of our link and what compromises must be made because of our constraints, let us consider a particular example. We now assume that $|P(f)|$ is greater than some positive number N throughout the frequency range $-W < f < W$. (For example, the probability density function p(t) might be a narrow pulse.)

If λ is small (i.e., we are willing to "spend" average power), if $N_2(f)$ is small (i.e., the transmission noise power density spectrum is small at all frequencies), and if $|L(f)|$ is large in the band (-W, W) (i.e., the transmission medium does not attenuate greatly in this band), then, for all frequencies in the band,

$$H(f) \approx T(d+\mu)\, P(f) \exp(-j2\pi bTf) \tag{26}$$

and otherwise,

$$H(f) = 0 \tag{27}$$

If we specialize further by taking p(t) to be a unit impulse and

$$\frac{1}{T} = 2W \tag{28}$$

then Eqs. 10, 26, and 27 may be used to show that the Fourier transform of H(f) is

$$h(t) = \frac{d \sin 2\pi W\left(t-\frac{b}{2W}\right)}{2\pi W\left(t-\frac{b}{2W}\right)} \tag{29}$$

which is consistent with Shannon's (1) sampling theorem.

Equations 24 and 25 and inequality 18 show that at frequencies at which conditions are poor (i.e., if λ is large or $N_2(f)$ is large or $|L(f)|$ is small) we should not "try very hard" to increase signal power. And, if conditions are so poor that inequality 18 is not satisfied, then we should not try to put in any signal power or attempt any amplification at that frequency. Positive, quantitative results concerning "how hard we should try" to signal and amplify (according to our criterion) are contained in Eqs. 16 and 17.

Extension of our results to M-ary, dependent, not equally likely messages is direct if we are willing to use a quantizer with equally spaced levels (separation, W volts). In this case, $(4\sigma^2/W^2)$ is an upper bound to the error probability. If the quantizer is also to be optimized, we must minimize the upper bound, B, directly. This leads to more complicated notation and results.

Signaling below the Nyquist rate introduces dependencies that increase the number of equations to be solved. The mathematical problems are then similar to those in reference 2.

D. W. Tufts

References

1. C. E. Shannon, Communication in the presence of noise, Proc. IRE 37, 10-21 (1949).

2. D. W. Tufts, An optimum received pulse shape for pulse code modulation, Quarterly Progress Report No. 55, Research Laboratory of Electronics, M.I.T., Oct. 15, 1959, p. 96.