DOCUMENT ROOM 36-412

Document Room, 36-412
Research Laboratory of Electronics
Massachusetts Institute of Technology

#5

COMMUNICATION IN THE PRESENCE OF ADDITIVE GAUSSIAN NOISE

F. A. MULLER

LOAN COPY

TECHNICAL REPORT NO. 244

MAY 27, 1953

RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

The Research Laboratory of Electronics is an interdepartmental laboratory of the Department of Electrical Engineering and the Department of Physics.

The research reported in this document was made possible in part by support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the Army Signal Corps, the Navy Department (Office of Naval Research), and the Air Force (Office of Scientific Research, Air Research and Development Command), under Signal Corps Contract DA36-039 sc-100, Project 8-102B-0; Department of the Army Project 3-99-10-022.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report No. 244 May 27, 1953

COMMUNICATION IN THE PRESENCE OF ADDITIVE GAUSSIAN NOISE

F. A. Muller

Abstract

This report presents an analysis of the properties of finite segments of noise taken from correlated gaussian noise. This analysis is applied to the problem of optimal detection of signals when a communication channel adds gaussian noise and introduces a linear distortion. Some specific examples are discussed briefly.

COMMUNICATION IN THE PRESENCE OF ADDITIVE GAUSSIAN NOISE

I. INTRODUCTION

This report will discuss systems of communication in which the symbols to be transmitted are represented by signal functions that are limited to consecutive intervals in time. These signals are disturbed by the addition of correlated gaussian noise. The problem is to compute a posteriori probabilities of transmitted symbols when the received signal and the a priori probabilities are known (1, 2). As correlated gaussian noise shows an autocorrelation over finite times, the a posteriori probabilities of different symbols in a sequence will be statistically dependent when the a priori probabilities of these symbols are statistically independent. Although this dependence is a part of the received information, it is difficult to make practical use of it. In this report, therefore, the condition will be imposed that the procedure of computing a posteriori probabilities for the symbol transmitted in a particular interval must be independent of the choice of all other symbols in a sequence. This condition may be further justified by the fact that the use of the statistical dependence of subsequent symbols as part of the received information would mean the use of a larger alphabet than had been agreed upon. A still sharper restriction will be placed upon the procedure of the detection; namely, that only the signal plus noise received in its own interval may be used, and that, therefore, no extrapolation of the noise from neighboring intervals is allowed. This condition is indicated when nothing is known about the signal outside the interval (this situation is often encountered when physical measurements are performed); or when, at least, the possible signals in neighboring intervals show a great variety, as they will do in communication systems using a large alphabet.

II. DESCRIPTION OF A SEGMENT OF CORRELATED GAUSSIAN NOISE

The procedure for analysis will be given here only in a rough physical outline. A sharp analysis of the necessary mathematics may be found in von Neumann's work (3). An exhaustive treatment of the mathematical statistics of very similar problems is given by Grenander (4).

It will be necessary to have an appropriate description of the segment of noise that the channel adds to the signal. One possibility would be to give the statistical properties of the coefficients n_μ of the Fourier series analysis. These coefficients may be seen as coordinates in a space with an infinite number of dimensions. Each function is represented by a point in this space. An ensemble of functions is represented by an ensemble of points. A gaussian ensemble is determined by a density function

$$P = \exp\left(-\frac{1}{2} \sum_{\mu, \nu} A_{\mu\nu}\, n_\mu\, n_\nu\right) \tag{1}$$

The coefficient $A_{\mu\mu}$ determines the variance of n_μ; the coefficient $A_{\mu\nu}$, $\mu \neq \nu$ determines the correlation of n_μ and n_ν. In general this correlation will not be zero.*

Hyperplanes of constant density are determined by

$$\sum_{\mu, \nu} A_{\mu\nu} n_\mu n_\nu = C \tag{2}$$

This is the equation of a multidimensional ellipsoid. When the axes of this ellipsoid are used as a new system of coordinates, the mixed terms will disappear from the density function, leaving

$$P = \exp\left(-\frac{1}{2} \sum_m \frac{n_m^2}{\sigma_m^2}\right) \tag{3}$$

The use of a new coordinate system indicates that we are no longer analyzing the noise in sines and cosines, but in a different, complete set of orthogonal functions.** By using this new set of functions, we simplify the problem greatly. We shall call these functions the eigenfunctions of the problem. They are determined by the statistical properties of the noise and the length of the time interval.

The first problem is to find these eigenfunctions. They are orthogonal and assumed to be normalized:

$$\int_{-1/2T}^{+1/2T} \psi_i \psi_j \, dt = \delta_{ij} \tag{4}$$

The analysis of a function F(t) is given by

$$F(t) = \sum f_j \psi_j(t) \tag{5}$$

---

*For clarification of this statement, consider, for example, a short segment of noise that contains only very low frequencies. The segment will consist of a nearly straight line. Fourier analysis of this segment gives sine terms, beside a constant term. Insofar as the approximation as a straight line is correct, the coefficients of all of these sine terms are proportional to only one stochastic variable. Therefore, the sine terms in this case will show a strong correlation.

**Each orthogonal coordinate system corresponds to a complete set of orthogonal functions. One of these complete sets is the set of δ-functions at all times. The analysis in terms of these δ-functions is trivial, and equal to defining the noise function in the time domain. The coefficients A_{mn} correspond in this case to the autocorrelation function.

Therefore

$$\int_{-T/2}^{+T/2} \psi_i(t)\, F(t)\, dt = \int_{-T/2}^{+T/2} \psi_i(t) \sum f_j\, \psi_j(t)\, dt = \sum f_j \int_{-T/2}^{+T/2} \psi_i(t)\, \psi_j(t)\, dt$$

$$= \sum f_j\, \delta_{ij} = f_i = \int_{-T/2}^{+T/2} \psi_i(t)\, F(t)\, dt \tag{6}$$

and of course with Eq. 5

$$\sum \psi_i(\theta) \int_{-T/2}^{+T/2} \psi_i(t)\, F(t)\, dt = F(\theta) \tag{7}$$

Let N(t) be a section of duration T from a noise function from an ensemble of gaussian noise, characterized by the autocorrelation function Φ. The set ψ_i must be chosen so that all correlations $\overline{n_i\, n_j}$ will be zero for $i \neq j$. Calculation of n_i and n_j with Eq. 6 gives

$$\overline{n_i\, n_j} = \overline{\int_{-T/2}^{+T/2} \psi_i(t)\, N(t)\, dt \int_{-T/2}^{+T/2} \psi_j(\tau)\, N(\tau)\, d\tau} \tag{8}$$

or, by interchanging integrations and averaging,

$$\overline{n_i\, n_j} = \iint_{-T/2}^{+T/2} \overline{N(t)\, N(\tau)}\, \psi_i(t)\, \psi_j(\tau)\, dt\, d\tau \tag{9}$$

and with the definition of Φ

$$\overline{n_i\, n_j} = \int_{-T/2}^{+T/2} \psi_i(t)\, dt \int_{-T/2}^{+T/2} \Phi(t-\tau)\, \psi_j(\tau)\, d\tau \tag{10}$$

with the condition $\overline{n_i\, n_j} = \delta_{ij}\, \sigma_i^2$ (the axis length of the ellipsoid)

$$\delta_{ij}\, \sigma_i^2 = \int_{-T/2}^{+T/2} \psi_i(t)\, dt \int_{-T/2}^{+T/2} \Phi(t-\tau)\, \psi_j\, \tau\, d\tau \tag{11}$$

$$\sum_i \psi_i(\theta)\, \delta_{ij}\, \sigma_i^2 = \sum \psi_i(\theta) \int_{-T/2}^{+T/2} \psi_i(t)\, dt \int_{-T/2}^{+T/2} \Phi(t-\tau)\, \psi_j\, \tau\, d\tau \tag{12}$$

and with Eq. 4

$$\sigma_j^2\, \psi_j(\theta) = \int_{-T/2}^{+T/2} \Phi(\theta-\tau)\, \psi_j(\tau)\, d\tau \tag{13}$$

indicating that the set of functions ψ_i is a set of eigenfunctions of the integral equation, Eq. 10, and that the variances σ_i^2 are the corresponding eigenvalues.

When the operator Φ_T is defined by

$$\Phi_T\, f(t) = \int_{-T/2}^{+T/2} \Phi(t-\tau)\, f(\tau)\, d\tau \tag{14}$$

Eq. 13 may be written

$$\sigma_i^2\, \psi_i(t) = \Phi_T\, \psi_i(t) \tag{15}$$

showing that the set of functions ψ_i is a set of eigenfunctions of the operator ϕ_T and that the variances σ_i^2 are the corresponding eigenvalues.

So far, $\psi_i(t)$ has only been defined for $-T/2 < t < +T/2$. When we define $\psi_i(t) = 0$ for $t \leqslant -T/2$ and $+T/2 \leqslant t$, Eq. 13 may be written

$$\sigma_j^2\, \psi_j(t) = I_T(t) \int_{-\infty}^{+\infty} \Phi(t-\tau)\, \psi_j(\tau)\, d\tau \tag{16}$$

where

$$\left.\begin{aligned} I_T(t) &= 1 \quad \text{for} \quad -T/2 < t < +T/2 \\ I_T(t) &= 0 \quad \text{for} \quad t \leqslant -T/2 \text{ and } +T/2 \leqslant t \end{aligned}\right\} \tag{17}$$

or

$$\sigma_j^2\, \psi_j(t) = I_T(t)\, \Phi_\infty\, \psi_j(t) \tag{18}$$

With the operator Φ_∞ defined by

$$\Phi_\infty\, f(t) = \int_{-\infty}^{+\infty} \Phi(t-\tau)\, f(\tau)\, d\tau \tag{19}$$

The operator Φ_T works on functions inside the interval $-T/2 < t < +T/2$; the operator Φ_∞ on functions in the entire time domain. Since $\Phi(\tau)$ is an even function, both Φ_T and Φ_∞ are hermitian and their eigenvalues are, therefore, real. A complete set of eigenfunctions of Φ_∞ are the sines and cosines. The eigenvalue as a function of the frequency therefore characterizes Φ_∞. This eigenvalue equals the power density $2\pi\, N(\omega)$ of the noise spectrum.

$$\Phi_\infty \cos \omega t = \int_{-\infty}^{+\infty} \Phi(t-\theta) \cos \omega\theta\, d\theta = \int_{-\infty}^{+\infty} \phi(\tau) \cos \omega(\tau+t)\, d\tau$$

$$= \int_{-\infty}^{+\infty} \Phi(\tau)\, (\cos \omega\tau \cos \omega t - \sin \omega\tau \sin \omega t)\, d\tau = \cos \omega t \int_{-\infty}^{+\infty} \phi\tau \cos \omega\tau\, d\tau$$

and

$$\Phi_\infty \cos \omega t = 2\pi \, N(\omega) \cos \omega t \tag{20}$$

(The Wiener-Khinchine relation, the symmetry of $\Phi(\tau)$, and the assumption that $\Phi(\tau) \to 0$ for $\tau \to \pm \infty$ are used.)

The Heaviside operator with transfer function $2\pi \, N(js)$, which is constructed by substituting $-s^2 = -(d/dt)^2$ for ω^2 in $2\pi \, N(\omega)$, has the same set of eigenfunctions and eigenvalues as ϕ_∞ and may therefore be identified with it.

$$\Phi_\infty \equiv 2\pi \, N(js) \tag{21}$$

If $N(\omega)$ is (or may be sufficiently approximated by) a rational function of ω^2, the eigenfunctions and eigenvalues can be found by a fairly simple process. Equation 18 may be written

$$I_T(t) \left(\Phi_\infty - \sigma_i^2 \right) \psi_i(t) = 0 \tag{22}$$

or

$$\left.\begin{aligned} &\left(\Phi_\infty - \sigma_i^2 \right) \psi_i(t) = \eta_i(t) \\ &\text{with} \\ &\psi_i(t) = 0 \quad \text{for} \quad t < -T/2 \text{ and } +T/2 < t \\ &\text{and} \\ &\eta_i(t) = 0 \quad \text{for} \quad -T/2 < t < +T/2 \end{aligned}\right\} \tag{23}$$

The behavior at the points $\pm T/2$ will be specified later.

Suppose

$$\Phi_\infty = 2\pi \, N(js) = \frac{\sum_{\ell=0}^{n} C_\ell \, s^{2\ell}}{\sum_{\ell=0}^{n} D_\ell \, s^{2\ell}} = \frac{C(s^2)}{D(s^2)} \tag{24}$$

$$\left[\frac{C(s^2)}{D(s^2)} - \sigma_i^2 \right] \psi_i(t) = \frac{C(s^2) - \sigma_i^2 \, D(s^2)}{D(s^2)} \, \psi_i(t) = \eta_i(t) \tag{25}$$

$$\left[C(s^2) - \sigma_i^2 \, D(s^2) \right] \psi_i(t) = D(s^2) \, \eta_i(t) = \xi_i(t) \tag{26}$$

The operators working in Eq. 26 on ψ_i and η_i are pure differential operators. Therefore ξ_i must fulfill the conditions for both ψ_i and η_i in Eq. 23

$$\xi_i(t) = 0 \text{ for every } t \text{ except for } t = \pm T/2$$

At the points $\pm T/2$, δ-functions and derivatives may exist. It may be shown easily that there is always a complete set of eigenfunctions that are either even or odd. We must therefore be able to find a complete set by assuming

$$\xi_i(t) = A_i(s)\,\delta(t + T/2) \pm A_i(-s)\,\delta(t - T/2) \tag{27}$$

$$A_i(s) = \sum_{\ell=0}^{2} A_{i\ell}\, s^{\ell} \tag{28}$$

The value of r must be chosen so that it will be consistent with Eq. 11 which may be written, for i = j,

$$\sigma_i^2 = \int_{-\infty}^{+\infty} \psi_i(t)\,\eta_i(t)\,dt = \int_{-\infty}^{+\infty} \frac{1}{C - \sigma^2 D}\,\xi_i(t) \cdot \frac{1}{D}\,\xi_i(t)\,dt \tag{29}$$

Whenever $r \geqslant 2n$, this integral will diverge at the points $t = \pm T/2$. Therefore

$$r = 2n - 1 \tag{30}$$

We may now try to determine the coefficients $A_{i\ell}$ from the conditions of Eq. 23 for η_i and ψ_i

$$\eta_i(t) = \frac{A_i(s)}{D(s^2)}\,\delta(t + T/2) \pm \frac{A_i(-s)}{D(s^2)}\,\delta(t - T/2) \tag{31}$$

The pulse response corresponding to the operator $A_i(s)/D(s^2)$ may be found with Heaviside's expansion theorem. This theorem must be used in a form that gives an even pulse response (for example $\phi(t)$) for a symmetrical operator (for instance ϕ_∞), that is, in a form giving the pulse response that is valid in the strip of convergence containing $s = 0$. Denoting the roots of $D(s^2) = 0$ by s_h and assuming that there are no double roots, Heaviside's theorem may be written

$$\frac{A_i(s)}{D(s^2)} \cdot \delta(t) = \sum_h \left\{\epsilon(t) - \epsilon\left[\mathrm{Re}(s_h)\right]\right\} \cdot \frac{A_i(s_h)}{\left\{\frac{d}{ds}\left[D(s^2)\right]\right\}_{s=s_h}} \exp(s_h t) \tag{32}$$

with

$$\epsilon(x) = 0,\ x < 0;\ \epsilon(0) = \tfrac{1}{2};\ \epsilon(x) = 1,\ x > 0 \tag{33}$$

Equation 31 may be written in the interval $-T/2 < t < +T/2$

$$0 = \sum_h \left\{\left[1 - \epsilon(\mathrm{Re}(s))\right] \frac{A_i(s)}{\frac{d}{ds}\left[D(s^2)\right]} \exp s(t + T/2) \mp \epsilon(\mathrm{Re}(s)) \frac{A_i(-s)}{\frac{d}{ds}\left[D(s^2)\right]} \exp s(t - T/2)\right\}_{s=s_h} \tag{34}$$

This requires

$$A_i(s_h) = 0 \quad (\mathrm{Re}(s_h) < 0) \tag{35}$$

and the identical condition

$$A_i(-s_h) = 0 \quad (\mathrm{Re}(s_h) > 0)$$

Pure imaginary roots do not exist as $N(\omega)$ is bounded. Double roots require that

$$\left.\begin{aligned} A_i(s_h) &= 0 \\ \frac{d}{ds} A_i(s_h) &= 0 \end{aligned}\right\} \quad \left[\operatorname{Re}(s_h) < 0\right] \tag{36}$$

In this way we obtain n conditions for the $A_{i\ell}$

$$A_{io} + A_{i1} s_h + A_{i2} s_h^2 + \ldots = 0 \quad \left[\operatorname{Re}(s_h) \leqslant 0\right] \tag{37}$$

and eventually derivatives for multiple roots.

The same procedure gives, when applied to $\psi_i(t)$,

$$\begin{aligned} \psi_i(t) = \sum_k \Bigg\{ & \left[\epsilon(t + T/2) - \epsilon(\operatorname{Re}(s))\right] \frac{A_i(s)}{\frac{d}{ds}\left[C(s^2) - \sigma_i^2\, D(s^2)\right]} \exp s(t + T/2) \\ & \pm \left[\epsilon(t - T/2) - \epsilon(\operatorname{Re}(s))\right] \frac{A_i(-s)}{\frac{d}{ds}\left[C(s^2) - \sigma_i^2\, D(s^2)\right]} \exp s(t - T/2) \Bigg\}_{s=s_k(\sigma_i)} \end{aligned} \tag{38}$$

with $s_k(\sigma_i)$, the roots of $C(s^2) - \sigma_i^2\, D(s^2) = 0$.

The condition that $\psi_i(t)$ be zero for $t < -T/2$ and $+T/2 < t$ requires

$$A_i(s_k) \exp(s_k T/2) \pm A_i(-s_k) \exp(-s_k T/2) = 0 \tag{39}$$

or

$$\begin{aligned} A_o \cosh(s_k T/2) &+ A_1 s_k \sinh(s_k T/2) \\ &+ A_2 s_k^2 \cosh(s_k T/2) + \ldots = 0 \quad (\text{for even } \psi_i) \end{aligned} \tag{40}$$

and

$$\begin{aligned} A_o \sinh(s_k T/2) &+ A_1 s_k \cosh(s_k T/2) \\ &+ A_2 s_k^2 \sinh(s_k T/2) + \ldots = 0 \quad (\text{for odd } \psi_i) \end{aligned} \tag{41}$$

For a double root, one of these has to be replaced by its derivative as shown in Eq. 36. In the set of Eq. 38, each equation is double; therefore, Eq. 38 represents 2n conditions. In total we have just as many equations as coefficients $A_{i\ell}$. However, as these equations are homogeneous, in order to have a solution $\not\equiv 0$, the determinant of these equations must be zero. This determinant is a function of σ_i^2 and determines the eigenvalues σ_i^2. It may be written easily for each application.

Double roots will occur only at discrete values of σ_i^2. The determinant, set up in the assumption of exclusive single roots, will indicate these values erroneously as eigenvalues. It will not be difficult to sift them out afterwards.

When an eigenvalue σ_i^2 is known, Eqs. 37 and 40, or 41, may be solved to find a set

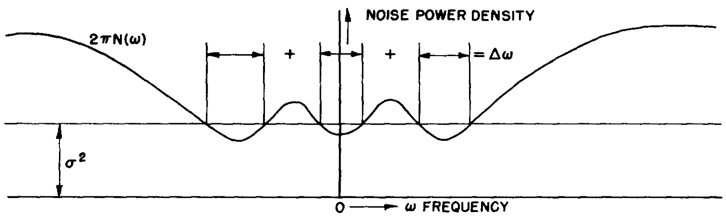

Fig. 1
An estimation of the number of eigenfunctions, with eigenvalues smaller than σ^2, for an arbitrary example of a noise spectrum.

of $A_{i\ell}$. Substitution in Eq. 38 then gives the corresponding eigenfunction which may be normalized afterwards. The eigenvalues σ_i^2 are the result of the pure imaginary roots p_k which make the hyperbolic functions periodic. These pure imaginary roots may be found by drawing a horizontal line at height σ^2 in the noise power spectrum (Fig. 1). The intersection points of this line with the power spectrum give the values of $\omega_k = js_k$. When σ^2 is increased by such an amount that all values ω_k travel together, on the average, a distance of $4\pi/T$, both the even and the odd determinant will pass through zero; therefore, there will be, on the average, two eigenvalues in this interval. Only real roots for ω_k have to be considered, since all other roots move in pairs so that their influence on the determinant is approximately cancelled. Consequently, an estimate of the number of eigenvalues smaller than σ^2 may be found by multiplying the total length $\Delta\omega$ where $\sigma^2 > 2\pi N(\omega)$ with $T/2\pi$. The main frequencies occurring in an eigenfunction with an eigenvalue of approximately σ^2 are given by $\sigma^2 = 2\pi N(\omega)$.

III. CALCULATION OF PROBABILITIES

Suppose a signal $\vec{X}_k$ with components x_{ki} has been transmitted. The probability density of receiving a signal $\vec{Y}$ with components y_i will be

$$P\left(\vec{Y}\middle|\vec{X}_k\right) = \prod_i \frac{\exp\left(-\frac{(y_i - x_{ki})^2}{2\sigma_i^2}\right)}{(2\pi\sigma_i)^{1/2}} \tag{42}$$

or

$$P\left(\vec{Y}\middle|\vec{X}_k\right) = \frac{\exp\left(-\sum_i \frac{y_i^2}{2\sigma_i^2}\right)}{\prod_i (2\pi\sigma_i)^{1/2}} \cdot \exp(-V_{kk} + 2\phi_k) \tag{43}$$

with

$$V_{kk} = \frac{1}{2}\sum_i \frac{x_{ki}^2}{\sigma_i^2} \tag{44}$$

and

$$\phi_k = y_2 \sum_i \frac{y_i\, x_{ki}}{\sigma_i^2} \tag{45}$$

These expressions for V_{kk} and ϕ_k may be interpreted as the dot product of $\vec{X}_k$ and $\vec{Y}$ respectively, with a vector $\vec{Z}_k$ having components $z_{ki} = x_{ki}/\sigma_i^2$.

The corresponding time function

$$Z_k(t) = \sum_i \frac{x_{ki}}{\sigma_i^2} \psi_i(t) \tag{46}$$

is the solution of an integral equation (Eq. 19):

$$\begin{aligned} X_k(t) &= \sum_i x_{ki}\, \psi_i(t) = \sum_i \frac{x_{ki}}{\sigma_i^2} \int_{-T/2}^{+T/2} \Phi(t-\tau)\, \psi_j(\tau)\, d\tau \\ &= \int_{-T/2}^{+T/2} \Phi(t-\tau) \left(\sum_i \frac{x_{ki}}{\sigma_i^2} \psi_i(\tau) \right) d\tau \\ &= \int_{-T/2}^{+T/2} \Phi(t-\tau)\, Z_k(\tau)\, d\tau = X_k(t) \end{aligned} \tag{47}$$

When the solution of this integral equation (Eq. 47) is not unique, the difference of two solutions is an eigenfunction with the eigenvalue zero of Eq. 13. This case, where the definition, Eq. 46, obviously makes no sense, would indicate that there are signals that are not disturbed by noise, and has, therefore, no practical importance.

The computation of $Z_k(t)$ for a given $X_k(t)$ may be carried out in a simpler way than by determining the eigenfunctions and using Eqs. 6 and 46. The process, leading from Eq. 13 to Eq. 23, may be applied to Eq. 47. This leads to

$$U(t) = \Phi_\infty Z(t) \tag{48}$$

with

$$Z(t) = 0 \text{ for } t < -T/2 \text{ and } +T/2 < t \tag{49}$$

and

$$U(t) = X(t) \text{ for } -T/2 < t < +T/2 \tag{50}$$

where, for simplicity, the subscript k has been dropped.

In order to solve Eq. 48 it is again necessary to assume that the power density of the noise is a rational function of the frequency; therefore

$$\Phi_{\infty} = \frac{\sum_{\ell=0}^{n} C_{\ell} s^{2\ell}}{\sum_{\ell=0}^{n} D_{\ell} s^{2\ell}} = \frac{C(s^2)}{D(s^2)} \qquad (24)$$

A function V(t) can be defined

$$V(t) = D(s^2)\ U(t) = C(s^2)\ Z(t) \qquad (51)$$

with conditions for V(t)

$$\left.\begin{aligned} V(t) &= C(s^2)\ Z(t) = 0 \text{ for } t < -T/2 \text{ and } +T/2 < t \\ V(t) &= D(s^2)\ U(t) = D(s^2)\ X(t) \text{ for } -T/2 < t < +T/2 \end{aligned}\right\} \qquad (52)$$

Addition of δ-functions and derivatives at $t = \pm T/2$ gives

$$V(t) = D(s^2)\ X(t) + A(s)\ \delta(t + T/2) + B(s)\ \delta(t - T/2) \qquad (53)$$

where A(s) and B(s) must again be determined from the conditions of Eqs. 49 and 50. Calculation of $Z(t) = 1/C(s^2)\ V(t)$ and $U(t) = 1/D(s^2)\ V(t)$ with Heaviside's expansion theorem gives, with the conditions of Eqs. 49 and 50,

$$\left.\begin{aligned} A(s_k)\exp(s_k T/2) + B(s_k)\exp(-s_k T/2) &= D\left(s_k^2\right)\int_{-T/2}^{+T/2} X(\tau)\exp(-s_k\tau)\,d\tau \\ A(s_h) = 0 \quad &(\mathrm{Re}(s_h) < 0) \\ B(s_h) = 0 \quad &(\mathrm{Re}(s_h) > 0) \end{aligned}\right\} \qquad (54)$$

where s_k are the zeros and s_h are the poles of the system function $N(\partial s)$. A and B are allowed to be of the grade $(2n - 1)$. Equation 54 allows calculation of the coefficients of

$$A(s) = \sum_{\ell=0}^{2n-1} A_{\ell} s^{\ell} \text{ and } B(s) = \sum_{\ell=0}^{2m-1} B_{\ell} s^{\ell}$$

When A(s) and B(s) are known, Heaviside's theorem gives

$$Z(t) = \sum_{1/2\,k}\left[\frac{1}{\frac{d}{ds}(C(s^2))}\left\{D(s^2)\int_{-T/2}^{+T/2}\exp(s\,|(t-\tau)|)\ X(\tau)\,d\tau + A(s)\exp(s(t+T/2)) + B(-s)\exp(-s(t-T/2))\right\}\right]_{s=s_k;\ \mathrm{Re}(s_k)<0} + \frac{D_n}{C_n}\cdot X(t) \qquad (55)$$

The computation of Z may be simplified by splitting X(t) into an even and an odd part, and treating these parts separately; setting $A_{even}(s) = B_{even}(-s)$ and $A_{odd}(s) = -B_{odd}(-s)$.

The dot product of two vectors is invariant under coordinate transformations

$$\sum k_i \ell_i = \sum k_i \int_{-T/2}^{+T/2} \psi_i(t)\ L(t)\ d(t)$$

$$= \int_{-T/2}^{+T/2} \sum k_i\ \psi_i(t)\ L(t)\ dt = \int_{-T/2}^{+T/2} K(t)\ L(t)\ dt$$

and therefore

$$V_{kk} = \frac{1}{2} \int_{-T/2}^{+T/2} X_k(t)\ Z_k(t)\ dt \tag{56}$$

and

$$\phi_k = \frac{1}{2} \int_{-T/2}^{+T/2} Y(t)\ Z_k(t)\ dt \tag{57}$$

From Eq. 43 it follows that the a posteriori probability $P(X_k|Y)$ of the signal X_k with an a priori probability $P(k)$ is (2)

$$P(X_k|Y) = \frac{P(X_k)\ \exp(-V_{kk} + 2\phi_k)}{\sum\limits_k P(X_k)\ \exp(-V_{kk} + 2\phi_k)} \tag{58}$$

As ϕ_k is obtained by a linear process from Z_k, and Z_k by a linear process from X_k, ϕ_k is a linear function of $X_k(t)$ (for a given $Y(t)$). Therefore, when there are linear relationships between the $X_k(t)$, linear relationships also exist between the ϕ_k. It will not be necessary in such cases to compute more correlation integrals (Eq. 21) than there are independent $X_k(t)$. The rest of the ϕ_k may be found as linear combinations. In practice, this crosscorrelation will always be carried out by constructing a filter (or other linear physical devices) with pulse response $Z_k(T/2 - t)$ and then feeding $Y(t)$ to it and sampling the output at the time $T/2$.

In a special case, which is often encountered in performing physical measurements, the possible signals all possess the same shape but different amplitude.

$$X_k(t) = k \cdot X_1(t) \tag{59}$$

where k is the quantity to be measured and $X_1(t)$ is a given function of time. Most often k is continuous. It may be assumed that the a priori probability distribution $P(k)$ is given.

Obviously

$$Z_k(t) = k \cdot Z_1(t) \tag{60}$$

and

$$V_{kk} = \frac{1}{2} k^2 \int_{-T/2}^{+T/2} X_1(t)\, Z_1(t)\, dt = \frac{1}{2} k^2\, V_{11} \tag{61}$$

$$\phi_k = \frac{1}{2} k \int_{-T/2}^{+T/2} Y(t)\, Z_1(t)\, dt = \frac{1}{2} k\, \phi_1 \tag{62}$$

Therefore, the a posteriori probability distribution of k will be

$$P(k\,|\,Y) = \frac{P(k) \exp - \frac{V_{11}}{2} \left(k - \frac{\phi_1}{V_{11}}\right)^2}{\int_{-\infty}^{+\infty} P(k) \exp - \frac{V_{11}}{2} \left(k - \frac{\phi_1}{V_{11}}\right)^2 dk} \tag{63}$$

When the a priori distribution is flat, the a posteriori distribution is gaussian with center value

$$k = \frac{\phi_1}{V_{11}} \tag{64}$$

and variance

$$\sigma_k^2 = \frac{1}{V_{11}} \tag{65}$$

Therefore, V_{11} may be interpreted as the signal-to-noise ratio for the signal $X_1(t)$ when the optimum filter is used to detect it.

When the optimum procedure is, in advance, assumed to be linear, the special case, described by Eq. 59, reduces to an optimum filtering problem. These optimum filters have been extensively dealt with in the literature (5, 6, 7, 8, 9). The work of Zadeh and Ragazzini (8) on finite memory filters gives the same results for the optimum filter as derived here. The mathematical methods, however, are slightly different.

IV. CHANNELS INTRODUCING LINEAR DISTORTION

The theory described so far may be used in connection with the problem of communication through a channel that introduces a linear distortion. The situation is shown schematically in Fig. 2. It is assumed that a sufficient approximation of the transfer function of the channel and of the power density of the noise is given as a rational function. At first, an additional restriction will be placed upon the channel transfer function: It will be assumed that it has no zeros on the imaginary axis or in the right half-plane. Under these circumstances the inverse network exists and may be used as a first step to the unravelling of the signal. This inverse filter multiplies the power density of the noise by $\left[F(j\omega)/E(j\omega)\right]^2$. The problem reduces therefore to the problem

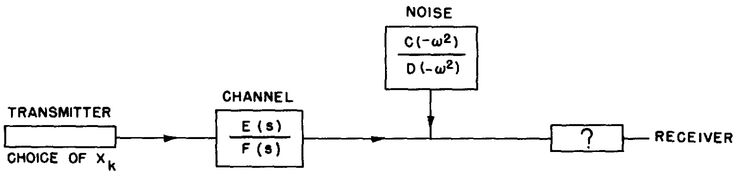

Fig. 2
A communication system introducing linear distortion and gaussian noise.

discussed in section I, with a noise power density $\left[F(j\omega)/E(j\omega)\right]^2 \cdot C(\omega^2)/D(\omega^2)$, the equivalent noise spectrum at the transmitter. The function Z(t), which must be crosscorrelated with the output of the inverse filter, may therefore be found. As a last step, the necessary practical outfit may be simplified by combining the filter F(s)/E(s) with the filter which performs the crosscorrelation, and making a physical device that realizes this combination to a sufficient degree of approximation.

The restriction on the location of the zeros of E(s)/F(s) will not be met in practical cases. In all but trivial communication channels, there will be a zero at $\omega = \infty$, and often there will be one at $\omega = 0$. Strictly speaking, it is then no longer possible to carry out the inverse operation F(s)/E(s). The penalty for attempting this operation is that the noise power density found afterwards is no longer bounded. Therefore, the autocorrelation function $\Phi(t)$ does not exist and the calculation loses its sense.

We may, however, try to find the solution as a limit by solving the problem for the transfer function $E(s)/F(s) + a$, where a is a small positive real number tending towards zero. Although the autocorrelation function does not converge to a limit, the eigenvalues and eigenfunctions do, with the exception of the eigenfunctions built up with the "forbidden" frequencies. The eigenvalues of these eigenfunctions increase proportionally with $1/a$. Consequently, although the convergence of the set of eigenfunctions and eigenvalues is not uniform, the function Z_k derived from any bounded X_k converges to a limit. This limit, $Z_{k, a=0}$, pays no attention to the forbidden frequencies X_k contains. This is, of course, a physically sound procedure. It may, therefore, be assumed that this limit is the correct answer to the problem.

With a few minor alterations, the given procedure of computing the set of eigenfunctions or the function Z_k may be made to yield directly the limiting values. A zero at infinity means that the degree 2n of $D|E|^2$ is smaller than the degree 2m of $C|F|^2$. The number of constants A then becomes m + n. A finite zero requires that Eqs. 35 and 54 hold also for $\mathrm{Re}(s_h) = 0$. With these additions the correct number of equations is obtained under all circumstances. The set of eigenfunctions found with this procedure is complete only for the received signals, not for the transmitted signals. The forbidden frequencies are lacking. This is, of course, not of practical importance, as one would not be interested in transmitting signals that cannot be received.

Although F(s)/E(s) is not realizable when E has zeros on the imaginary axis (as it

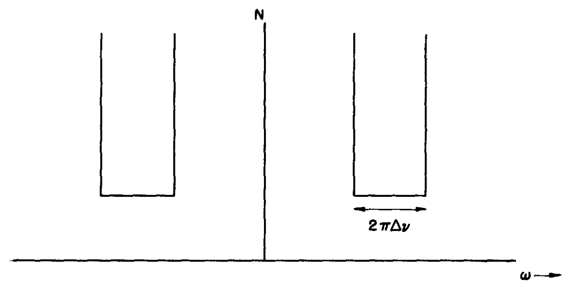

Fig. 3

Equivalent input noise for a sharply limited bandpass communication channel disturbed by uncorrelated noise.

would require an infinite gain for the corresponding frequencies), the combination of F(s)/E(s) and Z(t) is realizable, as the filter with impulse response Z(-t) has zeros on the imaginary axis where E(s) has them.

The second restriction imposed on E/F, the absence of zeros in the right half-plane, is necessary to insure the physical realizability of a network with the transfer function F/E. In practical cases, sometimes there will be zeros in the right half-plane; then the filter F/E is not realizable, nor are the combination filter of F/E and the cross-correlation filter. If one again interprets the function of this combination as a cross-correlation (or weighted averaging), the difficulty lies in the fact that, in this process, the whole future is involved. This future part of the weighting function is necessary in order to cancel the influence of transmitted signals outside the time interval. The weight attached to this future goes exponentially to zero with increasing time. For practical purposes a sufficient solution is found by introducing a reasonable time delay and cutting off the rest of the future. Of course this process does not give correct probabilities or statistically independent results, but the deviation may be made as small as required. In most practical cases there will be no objection to the delay, which is necessary to obtain a close approximation of the ideal situation. When, however, a shorter delay is required, it will be necessary to reconsider the problem in order to be able to prescribe the changes to be made in the past section of the weighting function as a result of the cutoff of the future. This problem, however, remains outside the scope of this report.

V. EXAMPLES

1. COMMUNICATION

As a first example, it may be shown that the approach of this report leads to the points of view of Shannon (10) and Rice (11) when applied to the problem of communication in a limited band, disturbed by uncorrelated noise. The equivalent input noise shows the spectrum (given in Fig. 3). The rule about the number of eigenfunctions for

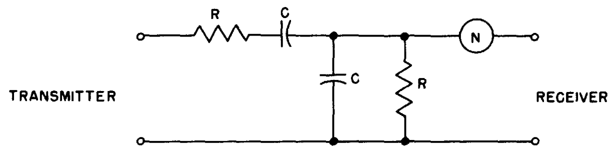

Fig. 4

A specific example of a communication channel introducing linear distortion and gaussian noise.

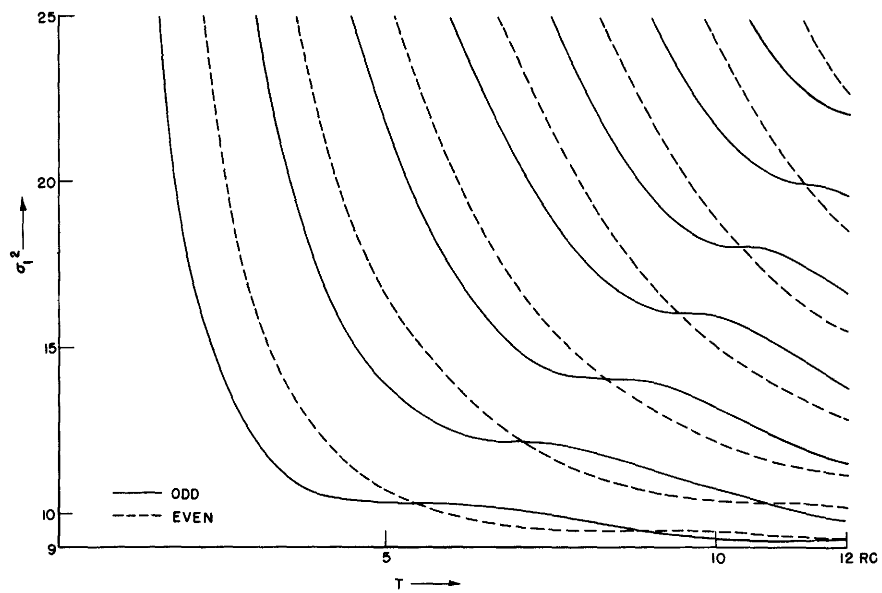

Fig. 5

The eigenvalues as a function of the basic time T for the channel of Fig. 4.

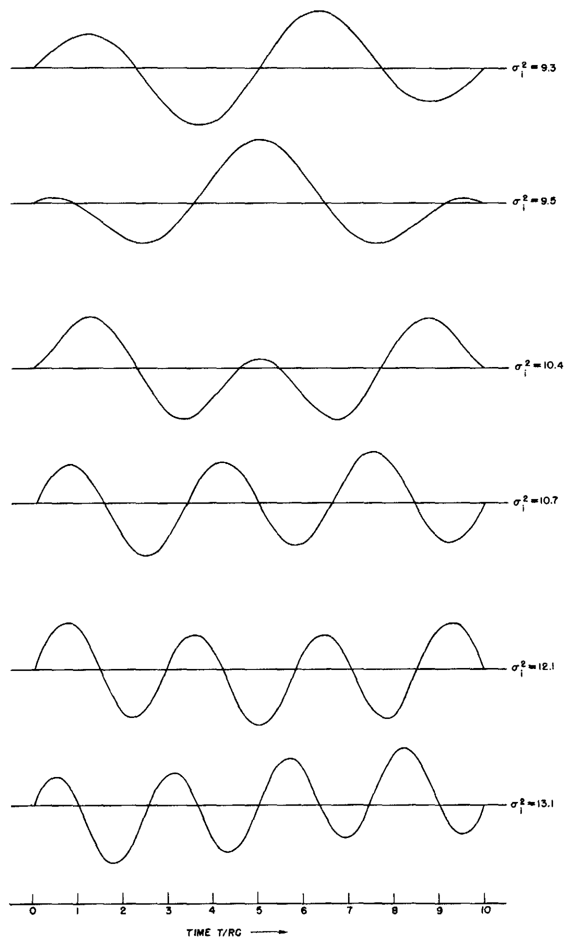

Fig. 6

Eigenfunctions for the channel of Fig. 4 and a basic time T = 10RC.

this case states that there are about $2T\Delta\nu$ eigenfunctions, all with the same eigenvalue. In such a degenerated case, all linear combinations of eigenfunctions are again eigenfunctions, and therefore any orthogonal set of about $2T\Delta\nu$ functions, substantially limited to the frequency band and time interval, may be used to describe the problem.

Figure 4 shows the simplest example of a low- and high-cutoff communication channel, disturbed by uncorrelated noise. Figure 5 shows the eigenvalues as a function of the basic time T. Figure 6 shows a number of the lowest eigenfunctions for T = 10RC.

The lowest eigenfunction gives, for a fixed energy, the largest number of distinguishable levels (with a certain probability of error). When the average energy is limited, it seems, therefore, advisable to choose the transmitted signals as combinations of a number of the lower eigenfunctions. When only a small alphabet is allowed, the number of possibilities is limited and the best may be selected. It seems probable that when a large alphabet is allowed, a random distribution of signal points in accordance with Shannon's law (10), requiring that signal energy plus noise energy be constant as far as possible, will approach asymptotically the theoretical limit of the rate of transmission. A proof similar to the one Rice gave for the case of noncorrelated noise (11) would be very difficult here, as there is no spherical symmetry in the signal space.

2. PHYSICAL MEASUREMENTS

A typical example of the theory (12) is the point-by-point measurement of a quantity as a function of some parameter; for instance, the collector current in a mass spectrometer as a function of the magnet current. When a time T is available to determine a point, the magnet current is kept constant during that time; therefore the signal, in this case the collector current, is constant during that interval. Nothing is known about the signal before and after the measuring time. The current measuring device usually is a voltage measuring instrument equipped with a current feedback. This feedback effects a transformation of the nearly uncorrelated noise of the voltage meter into a noise with a power density approximately proportional to ω^2 (at least for frequencies where the feedback is effective). The feedback resistor adds a constant power density. An additional constant power density may serve as a first approximation for contact-potential variations and flicker effect. The total noise spectrum becomes the form $N = a\omega^2 + b$. A typical example is $a = 10^{-33}\ A^2 sec^3$; $b = 10^{-33}\ A^2 sec$. Optimal pulse responses for T = 1 sec and T = 10 sec are given in Fig. 7.

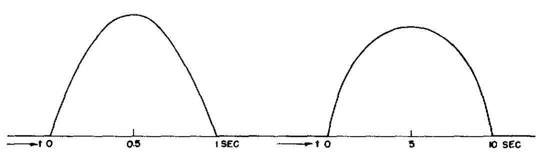

Fig. 7

Two optimal pulse responses for a specific example of a current-measuring instrument.

Acknowledgment

The author is indebted to Professor R. M. Fano for many helpful discussions and constant encouragement of the research presented in this report.

References

1. P. M. Woodward, I. L. Davies: Phil. Mag. 41, 1001, 1950
2. R. M. Fano: Communication in the Presence of Additive Gaussian Noise, Symposium on Applications of Communication Theory, London, 1952
3. J. von Neumann: Die Mathematische Grundlagen der Quantenmechanik, Springer, Berlin, 1932
4. U. Grenander: Arkiv Mat. (Stockholm) 1, 195, 1950
5. D. O. North: Report PTR-6C, RCA Laboratories, 1943
6. H. den Hartog, F. A. Muller: Physica 13, 571, 1947
7. B. M. Dwork: Proc. I.R.E. 38, 771, 1950
8. L. A. Zadeh, J. R. Ragazzini: Proc. I.R.E. 40, 1223, 1952
9. K. Halbach: Helv. Phys. Acta 26, 65, 1953
10. C. E. Shannon: Proc. I.R.E. 37, 10, 1949
11. S. O. Rice: Bell System Tech. J. 29, 60, 1950
12. F. A. Muller: Measurement of Currents and Voltages with a Vibrating-plate Electrometer, Doctoral Thesis, University of Amsterdam, 1951