3.

DESIGN PROBLEMS IN PULSE TRANSMISSION

DONALD WINSTON TUFTS

TECHNICAL REPORT 368

JULY 28, 1960

Loan Copy Only

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS
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 U. S. Army (Signal Corps), the U.S. Navy (Office of Naval Research), and the U.S. Air Force (Office of Scientific Research, Air Research and Development Command), under Signal Corps Contract DA36-039-sc-78108, Department of the Army Task 3-99-20-001 and Project 3-99-00-000.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report 368 July 28, 1960

DESIGN PROBLEMS IN PULSE TRANSMISSION

Donald Winston Tufts

Submitted to the Department of Electrical Engineering, M. I. T., May 25, 1960, in partial fulfillment of the requirements for the degree of Doctor of Science.

Abstract

Criteria for determining optimum pulse shapes for use in a synchronous pulse transmission link are proposed and compared. Formulas for the optimum pulses are presented, and calculations are carried through in examples.

In a secondary problem, we present methods for calculating optimum interpolatory pulses for use in reconstructing a random waveform from uniformly spaced samples. The case of a finite number of samples and the case of an infinite number of samples are discussed. A mean-square criterion is used to judge the approximation. The results provide a generalization of the sampling principle.

TABLE OF CONTENTS

I. INTRODUCTION

This report is concerned with problems in the design of synchronous pulse transmission links like the link shown in Fig. 1. In this model the transmitter sends a voltage pulse into a transmission line every T seconds. The pulses differ only in their amplitudes and time delays. The peak amplitude of each pulse corresponds to a message. If the peak amplitude of a transmitted pulse can be determined at the receiver, then the message carried by that pulse can also be determined.

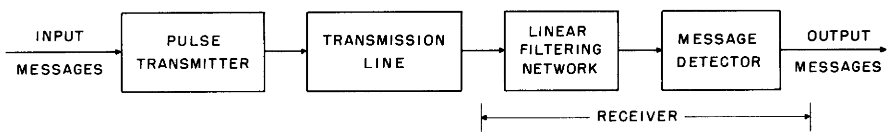

Fig. 1. Block diagram of a synchronous pulse transmission link.

The receiver consists of a linear network and a detector. The output of the transmission line is filtered by the linear network, and the detector operates on the resulting waveform every T seconds to extract one message. Ideally, the time delay between the transmission of a message and its detection is a constant. That is, the link is synchronous.

Our main problem is to determine transmitted pulse shapes and linear filtering networks that optimize the performance of such links. A secondary problem is to find optimum interpolatory pulses to use in reconstructing a random waveform from uniformly spaced samples. The same mathematical methods are required in each problem because the same type of nonstationary waveform occurs.

Our main problem is motivated by the following practical, theoretical, and mathematical considerations.

1. Synchronous pulse links are of great practical importance (1, 2) because they are naturally suited to the transmission of digital messages (3). We discover how signal power and bandwidth can best be used in our models to obtain good link performance. This is a step in attempting to improve present synchronous pulse links (1).

2. It has been shown that, under idealized conditions, pulse code modulation (PCM) exchanges increases in bandwidth for decreases in signal-to-noise ratio more efficiently than comparable systems, such as frequency-modulation systems (4). A PCM link is a special form of the synchronous pulse link. It is of interest to discover how well the theoretical advantages of pulse code modulation are preserved in link models that are different from the model that is considered by Oliver, Pierce, and Shannon (4).

3. The random waveforms that occur in synchronous pulse links are members of nonstationary ensembles. The mathematical properties of these waveforms have been only partially explored (5, 6, 7).

Our search for optimum interpolatory pulses is also motivated by mathematical interest in nonstationary ensembles. However, the main motivation is provided by two questions: How well can a random waveform be reconstructed from uniformly spaced samples of the waveform? Can present methods of interpolating samples be used for a wider class of random processes?

In answering the first question, we shall obtain estimates to a measure of the optimum reconstruction by considering a particular method of reconstruction. The present methods that are referred to in the second question are those discussed by Shannon and others (8, 9). Shannon considered the exact reconstruction of bandlimited waveforms from an infinite number of samples. We shall consider the approximate reconstruction of random waveforms from a finite number of samples.

II. THIN PULSE PERFORMANCE CRITERIA

2.1 The Basic Model

Let us use the block diagram of Fig. 2 as our basic synchronous pulse transmission link model — a model that follows Shannon's model for a general communications system (10). The message source produces a sequence of area modulated impulses, m(t), where

$$m(t) = \sum_{n=-\infty}^{\infty} a_n u_o(t-nT) \tag{1}$$

and $u_o(t)$ is the unit impulse or delta function. Each real number a_n, $n = 0, \pm 1, \pm 2, \ldots,$ corresponds to a transmitted message. For example, a particular message might be a page of English text or it might be a voltage amplitude. The waveform m(t) is the input for the linear pulse-forming network of Fig. 2, which has an impulse response s(t) and a transfer function S(f). Its output is

$$f_I(t) = \sum_{n=-\infty}^{\infty} a_n s(t-nT) \tag{2}$$

The waveform $f_I(t)$ represents the transmitted signal in a real synchronous pulse link.

The linear transmission network represents the transmission medium in a real link. It has impulse response $\ell(t)$, transfer function L(f), input $f_I(t)$, and output

$$f_s(t) = \sum_{n=-\infty}^{\infty} a_n r(t-nT) \tag{3}$$

where

$$r(t) = \int_{-\infty}^{\infty} s(u)\, \ell(t-u)\, du \tag{4}$$

The pulse r(t) is the response of the transmission network to an input pulse s(t).

The detector consists of a sampler and a clock. Ideally, the clock output, c(t), is a periodic train of unit impulses. And, ideally, each clock impulse coincides with a peak value of one of the pulses of $f_s(t)$. An instantaneous voltage sample of $f_s(t)$ is taken whenever the sampler receives a clock impulse. The ideal clock impulses and three typical pulse components of $f_s(t)$ are depicted in Fig. 2. It is assumed that r(t) attains

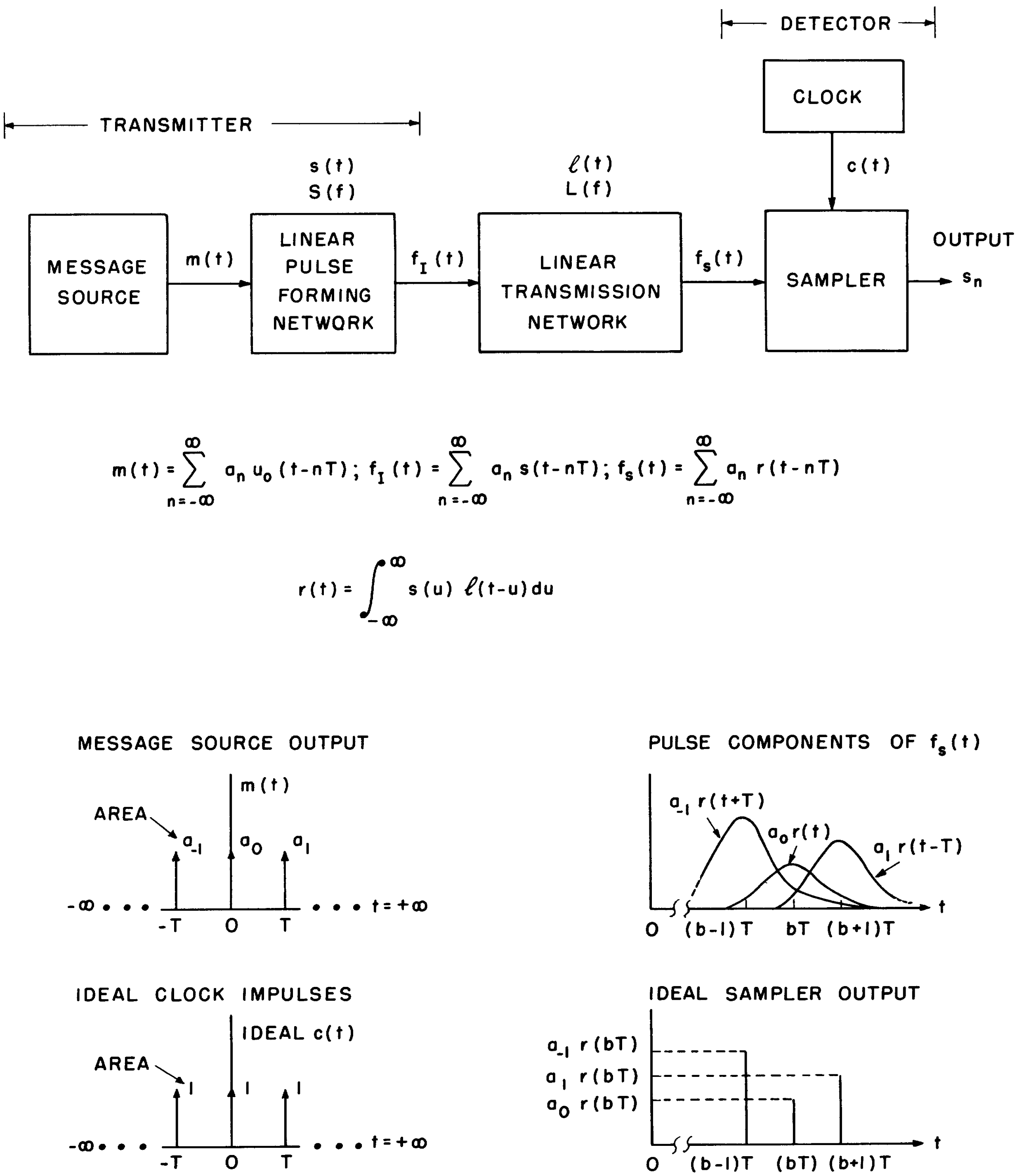

Fig. 2. Basic synchronous pulse transmission link model and associated waveforms.

a peak value when $t = bT$, where b is a positive integer. Thus, if the clock emits the ideal impulse train shown in Fig. 2 and if we say that the pulse $a_n\, s(t-nT)$ is transmitted at time nT, then the time delay between transmission and detection of a message is bT seconds.

The actual time delay is a random variable because we assume that there is jitter in the actual sampling instants. More specifically, the actual sampling instants are assumed to be in one-to-one correspondence with the ideal sampling instants. Each actual sampling instant does not differ from its corresponding ideal sampling instant by more than β seconds; and 2β is smaller than T, the time between ideal samples. Physically, this might come about because of thermal noise in the clock's oscillator. Any long-term drifting of the clock impulses in either time direction is assumed to be prevented, for example, by the use of special retiming pulses that are sent through the transmission network.

2.2 Definition of Interpulse Interference

If the clock impulses were ideal and $r(t)$ were zero at all ideal sampling instants, except $t = bT$, then the sampler output would be a sequence of real numbers, $\{a_n\, r(bT)\}$, that are proportional to the message sequence $\{a_n\}$. Three elements of an ideal output sequence are shown in Fig. 2. We denote the actual sequence of sampler outputs by $\{s_n\}$. We say that our link performs well if each element of the normalized sampler output sequence $\left\{\frac{s_n}{r(bT)}\right\}$ is close to the corresponding element of the message sequence $\{a_n\}$. The differences $\frac{s_n}{r(bT)} - a_n$, $n = 0, \pm 1, \pm 2, \ldots$, are called interpulse interference. Thus, the link performs well if interpulse interference is small.

Using this definition of interpulse interference, our assumptions about the sampler operation and the sampling time jitter, and the expression for the sampler input in Eq. 3, we can write the interpulse interference more explicitly. At any particular sampling instant, t_s, the interpulse interference is

$$I(t_s) = \frac{\sum_{j=-\infty}^{\infty} a_j\, r(t_s - jT) - a_n\, r(bT)}{r(bT)} \tag{5}$$

where

$$(b+n)\, T - \beta < t_s < (b+n)\, T + \beta \tag{6}$$

$$n = 0, \pm 1, \pm 2, \ldots \tag{7}$$

and 2β, the extent of the time uncertainty interval about each ideal sampling instant, is less than T. We recall that $1/T$ is the rate of transmitting messages.

2.3 Discussion of the Basic Model

In previous publications on synchronous pulse transmission, ideal clock impulses are almost always assumed (11, 12, 13). Under this assumption one can choose a band-limited received pulse r(t) in such a way that r(t) is zero at all ideal sampling instants except t = bT. Under these conditions the interpulse interference, $I(t_s)$, is zero at all sampling instants. This occurs, for example, when

$$r(t) = \frac{\sin[(\pi/T)(t-bT)]}{(\pi/T)(t-bT)} \tag{8}$$

In the absence of noise such a link would perform perfectly. That is, the output sequence of samples would be proportional to the input message sequence of amplitudes. However, the models of references 11, 12, and 13 do contain noise sources. In those papers major emphasis is placed on combating noise other than interpulse interference. Here, we place our emphasis on reducing interpulse interference. We assume that the effects of any additive noise source, such as that shown in Fig. 3, are negligible, compared with the interpulse interference. By the effects of the noise we mean the perturbations of the normalized sequence of samples that are due to the noise alone. If the noise source is removed, the linear filtering and transmission networks of Fig. 3 can be combined, and we have our basic model of Fig. 2. It can be seen from Eq. 5 that interpulse interference is independent of the amplitude of the standard input pulse s(t). This follows from the fact that the amplitude of the received pulse r(t) is proportional to the amplitude of s(t). But if the amplitudes of the received pulses are increased by transmitting larger pulses, the effects of the noise are reduced.

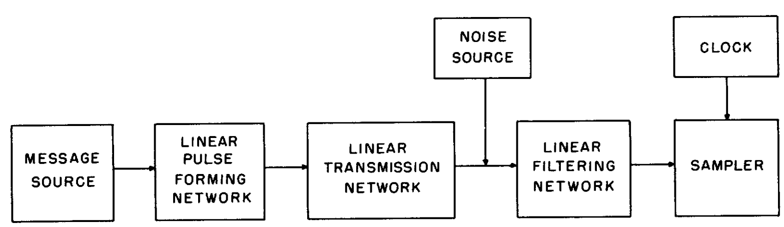

Fig. 3. Modified basic model.

Let us use our model of Fig. 2 and assume that the transmitted signal power is so large that, for a given rate of signaling, 1/T, the performance of the link is limited only by interpulse interference. For rapid rates of signaling this behavior can be achieved at a level of signal power that is small, compared with the level required at slow rates. This comes about because interpulse interference increases with the rate of signaling, due to greater received pulse overlapping, while the noise effects are independent of the signaling rate.

2.4 Output Energy Criterion

For a given rate of signaling we wish to make the interpulse interference, $I(t_s)$, small at all sampling instants so that the link will perform well. We also desire to signal as rapidly as possible so that our rate of flow of information is high. However, we do not attempt to signal so fast that the condition

$$2\beta < T \tag{9}$$

is violated. We recall that 2β is the extent of the sampling time uncertainty interval that surrounds each ideal sampling instant, and $1/T$ is the signaling rate. Inequality 9 preserves the synchronous property of our basic link model.

One way to make $I(t_s)$ small while signaling at a rapid rate is to make each pulse as thin as possible. This can keep pulse overlap small. A criterion for obtaining a thin output pulse is that

$$C_1 = \frac{1}{r(bT)} \int_{-\infty}^{\infty} [r(t)]^2 \, dt \tag{10}$$

shall be small. We assume that

$$r(bT) = \int_{-\infty}^{\infty} s(x) \, \ell(bT-x) \, dx = K_1 \tag{11}$$

where K_1 is a constant. As before, $r(t)$ is to attain a peak value at $t = bT$, and b is a positive integer. It is reasonable that, if the energy in a pulse is small for a prescribed peak value, the pulse must be thin. Strictly speaking, C_1 is proportional to the energy delivered by $r(t)$ only if the impedance seen by $r(t)$ is pure resistance. Further interpretation of C_1 is given in section 2.6.

Standard calculus of variations and Fourier methods can be used to show that

$$2S(f) \, |L(f)|^2 - \lambda L^*(f) \exp(-j2\pi bTf) = 0 \tag{12}$$

is a necessary condition for C_1 to be minimized by variation of $S(f)$, subject to the constraint of Eq. 11. (See ref. 14.) These methods are illustrated by a more detailed derivation in section 2.5. We recall that $S(f)$ is the Fourier transform of the input pulse $s(t)$; $L^*(f)$ is the complex conjugate of the transfer function of the transmission network; and λ is a Lagrange multiplier (14).

Equation 12 implies that $|S(f)|$ can become arbitrarily large near a zero of $L(f)$. For example, if

$$|L(f)| = \exp(-f^2) \tag{13}$$

then

$$|S(f)| = \exp(f^2) \tag{14}$$

and $s(t)$ does not exist. That is, $S(f)$ does not have a Fourier transform.

Imposing a constraint on the input energy per pulse is a natural method of avoiding this difficulty. In the case of statistically independent messages the input energy per pulse is proportional to the average transmitted power. In other cases it is a convenient measure of average power. For economic reasons it is often important to use average power efficiently.

If the input admittance of the transmission network is Y(f), and s(t) is a voltage pulse, then the equation

$$\int_{-\infty}^{\infty} S(f)\, Y(f)\, S^{*}(f)\, df = K_2 \tag{15}$$

states that the input energy is a constant K_2. If Y(f) is a constant conductance, then Eq. 15 becomes

$$\int_{-\infty}^{\infty} S(f)\, S^{*}(f)\, df = K_2 \tag{16}$$

A necessary condition for C_1 of Eq. 10 to be minimized by varying S(f), subject to the constraints of Eqs. 11 and 16 is

$$S(f) = \frac{\lambda L^{*}(f) \exp(-j2\pi bTf)}{|L(f)|^{2} + \mu} \tag{17}$$

where λ and μ are Lagrange multipliers that must be chosen to satisfy our two constraints. Equation 17 has the advantages that S(f) is Fourier-transformable and its magnitude is bounded. One might think that values of μ and λ could be chosen to offset these advantages. However, such values lead either to infinite input pulse energy or to negative values of $|S(f)|$.

If we select the values K_1 and K_2 arbitrarily, we have no guarantee that we can solve the constraint equations 11 and 16 for μ and λ. But, given any pair of real numbers (μ, λ), we can compute K_1 and K_2. We can then vary μ and λ to bring us closer to the desired values of K_1 and K_2. This is a reasonable procedure because S(f), and hence K_1 and K_2, are continuous functions of λ and μ.

As an example, let us assume that

$$L(f) = \begin{cases} \exp(-j2\pi bTf) & |f| < B \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

Then, from Eq. 17, we have

$$S(f) = \begin{cases} \lambda/\mu+1 & |f| < B \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

Hence,

$$s(t) = \frac{2B\lambda}{\mu + 1} \frac{\sin 2\pi Bt}{2\pi Bt} \tag{20}$$

And r(t), the Fourier transform of $R(f) = S(f)\,L(f)$, is

$$r(t) = \frac{2B\lambda}{\mu + 1} \frac{\sin 2\pi B(t-bT)}{2\pi B(t-bT)} \tag{21}$$

From Eqs. 11 and 21, we obtain

$$r(bT) = \frac{2B\lambda}{\mu + 1} = K_1 \tag{22}$$

By use of Eqs. 16 and 19, we have

$$\int_{-\infty}^{\infty} S(f)\, S^*(f)\, df = 2B \left[\frac{\lambda}{\mu+1}\right]^2 = K_2 \tag{23}$$

Since K_1 and K_2 are both determined by the ratio $\lambda/\mu+1$, they cannot be chosen independently.

If we had used the constraint of Eq. 15, rather than the constraint of Eq. 16, then Eq. 17 is replaced by

$$S(f) = \frac{\lambda L^*(f) \exp(-j2\pi bTf)}{|L(f)|^2 + \mu \mathrm{Re}[Y(f)]} \tag{24}$$

Since Eqs. 17 and 24 appear as special cases of Eq. 46, we postpone additional interpretation to section 2.6.

2.5 Weighted Output Energy Criterion

We shall now consider a second measure of pulsewidth. It is

$$C_2 = \frac{1}{r(bT)} \int_{-\infty}^{\infty} (t-bT)^2 \,[r(t)]^2\, dt \tag{25}$$

This measure puts greater emphasis on portions of the pulse that are far away from the sampling instant $t = bT$. According to this measure, r(t) is concentrated at $t = bT$ if C_2 is small. The quantity C_2 is analogous to the variance of a probability density function (15).

We shall now minimize C_2 under the constraints of Eqs. 11 and 16. Expressed in the frequency domain, we want to minimize

$$C_2 = \frac{1}{r(bT)} \int_{-\infty}^{\infty} \frac{d}{df} \left|R(f) \exp(+j2\pi bTf)\right|^2 df \tag{26}$$

by varying S(f), subject to the constraints

$$r(bT) = \int_{-\infty}^{\infty} R(f) \exp(j2\pi bTf)\, df = K_1 \tag{27}$$

from Eq. 11 and

$$\int_{-\infty}^{\infty} |S(f)|^2\, df = K_2 \tag{28}$$

from Eq. 16.

The purpose of Eqs. 11 and 27 is to guarantee that our received pulses are large enough so that the effects of noise are negligible compared to the interpulse interference. The purpose of Eqs. 16 and 28 is to conserve the average transmitted power.

Equation 26 can be written more explicitly as

$$C_2 = \frac{1}{r(bT)} \int_{-\infty}^{\infty} \left\{ \left[\frac{d}{df} |R(f)| \right]^2 + \left[|R(f)| \frac{d}{df} \rho(f) \right]^2 \right\} df \tag{29}$$

where $\rho(f)$ is the phase function of $R(f) \exp(j2\pi bTf)$. Note that the second, positively contributing term of the integrand in Eq. 29 is zero if $\rho(f)$ is a constant. This constant must be zero or, equivalently, a multiple of 2π. Otherwise, $r(t)$ would be a complex function of t. We also note that the first term of the integrand in Eq. 29 is independent of $\rho(f)$. Thus, we should choose $\rho_S(f)$, the phase function of $S(f)$, in such a way that

$$\rho(f) = \rho_R(f) + 2\pi bTf = \rho_S(f) + \rho_L(f) + 2\pi bTf = 0 \tag{30}$$

where $\rho_R(f)$ and $\rho_L(f)$ are the phase functions of $R(f)$ and $L(f)$. We recall that $L(f)$ is the transmission network transfer function.

We now want to minimize

$$C_2 = \frac{1}{r(bT)} \int_{-\infty}^{\infty} \left[\frac{d}{df} |S(f)L(f)| \right]^2 df \tag{31}$$

by varying $|S(f)|$, subject to the constraint

$$r(bT) = \int_{-\infty}^{\infty} |S(f)R(f)|\, df = K_1 \tag{32}$$

and the constraint of Eq. 28. Using a well-known theorem of the calculus of variations, we can solve our problem by minimizing

$$\int_{-\infty}^{\infty} \left\{ \left[\frac{d}{df} |S(f)L(f)| \right]^2 - \lambda |S(f)L(f)| - \mu |S(f)|^2 \right\} df \tag{33}$$

subject to no constraint (16). The Lagrange multipliers, λ and μ, must be chosen in such a way that Eqs. 28 and 32 are satisfied.

Minimization of integral 33 by variation of $|S(f)|$ is a standard variational problem (17). A necessary condition for such a minimum is that $|S(f)|$ must satisfy the differential equation

$$|L(f)| \frac{d^2}{df^2} |S(f)L(f)| = \lambda |L(f)| + \mu |S(f)| \tag{34}$$

The optimum phase function $\rho_S(f)$ for S(f) must satisfy Eq. 30. That is,

$$\rho_S(f) + \rho_L(f) = -2\pi bTf \tag{35}$$

The writer has not obtained a general solution for Eq. 34, but special cases have been considered.

Let us remove our constraint on the input pulse energy by letting $\mu = 0$. We also assume that the transmission network is an ideal lowpass filter as in section 2.4 (cf. Eq. 18). Thus, Eq. 34 becomes

$$\frac{d^2}{df^2} |S(f)L(f)| = \frac{d^2}{df^2} |R(f)| = \lambda \qquad |f| < B \tag{36}$$

or

$$|S(f)L(f)| = |R(f)| = \lambda f^2 + K_3 \qquad |f| < B \tag{37}$$

where K_3 is a positive constant. There is no first-degree term in $|R(f)|$; from Fourier integral theory, $|R(f)|$ must be an even function. From Eqs. 35 and 37 a solution is

$$R(f) = \begin{cases} \left(\lambda f^2 + K_3\right) d \exp(-j2\pi bTf) & |f| < B \\ 0 & \text{otherwise} \end{cases} \tag{38}$$

and

$$r(t) = 2B(\lambda B^2 + K_3) \frac{\sin 2\pi B(t-bT)}{2\pi B(t-bT)} - 4\lambda B^3 \frac{\sin 2\pi B(t-bT)}{[2\pi B(t-bT)]^3} - \frac{\cos 2\pi B(t-bT)}{[2\pi B(t-bT)]^2} \tag{39}$$

If λ is chosen to be negative and K_3 chosen to be $-\lambda B^2$, then the first term of r(t) drops out, and r(t) falls off as $1/t^2$ for large t. In this case, L'Hospital's rule shows that

$$r(bT) = -\frac{4}{3} \lambda B^3 = K_1 \tag{40}$$

Thus, our constraint can be satisfied for any positive K_1.

Gabor's "signal shape which can be transmitted in the shortest effective time" is a particular solution of Eqs. 34 and 35 when there is no constraint on r(bT); that is, $\lambda = 0$, and L(f) is again an ideal lowpass transfer function. In this case

$$S(f) = \begin{cases} K_4 \cos(\pi f/2B) & |f| < B \\ 0 & \text{otherwise} \end{cases} \tag{41}$$

where K_4 is a constant. It follows that

$$s(t) = r(t+bT) = \frac{K_4 \cos 2\pi Bt}{\pi B[(1/4B^2)-4t^2]} \tag{42}$$

We note that the envelope of this pulse falls off as $K_4/4\pi Bt^2$ for large values of t. For comparison, a pulse whose spectrum has the same bandwidth and peak value is

$$s_1(t) = \frac{2BK_4 \sin^2 \pi Bt}{(\pi Bt)^2} \tag{43}$$

The envelope of $s_1(t)$ falls off as $2K_4/B\pi^2t^2$, or slightly slower than the pulse s(t) of Eq. 42. However, $s_1(t)$ has a larger peak value than that of s(t).

2.6 Least Squares Criterion

The thin pulse criteria were chosen to reduce the effects of pulse overlap. They both neglect an effect of sampling time jitter. For example, if the received pulses are very thin it is possible that a sample is taken when no pulse is present. In short, a good received pulse must not only have low amplitude when other pulses are sampled, but also have high amplitude whenever it can be sampled.

We continue to assume that each sampling instant occurs within β seconds of an ideal sampling instant and that 2β is less than T, the time between pulses. Thus, a desired received pulse-shape is

$$d(t) = \begin{cases} d & bT - \beta < t < bT + \beta \\ 0 & \text{otherwise} \end{cases} \tag{44}$$

where d is a positive real number that is chosen to satisfy our requirement for received pulse amplitude. We recall that this amplitude must be large so that we can neglect the effects of noise and use our model of Fig. 2. As before, we assume that the transmission delay is bT seconds.

Our criterion for a good received pulse is that

$$C_3 = \int_{-\infty}^{\infty} [d(t)-r(t)]^2 \, dt \tag{45}$$

shall be small. A necessary condition for C_3 to be minimum by varying S(f), the Fourier transform of the input pulse shape s(t), is

$$S(f) = \frac{D(f)\, L^*(f)}{|L(f)|^2 + \mu \mathrm{Re}[Y(f)]} \tag{46}$$

This condition is obtained by the standard Fourier and variational techniques that we employed in section 2.5. The total energy available from a transmitted pulse is constrained as in Eq. 15, and the Lagrange multiplier μ must be chosen to satisfy this constraint. The Fourier transform of the desired pulse shape d(t) is denoted by D(f); $L^*(f)$ is the complex conjugate of the transfer function L(f); and Re[Y(f)] is the real part of the input admittance of the transmission network.

Equations 17 and 24 can be considered as special cases of Eq. 46 because the mathematical steps leading from Eq. 45 to Eq. 46 are independent of our choice for d(t). If we choose d(t) to be an impulse of area λ, then Eq. 46 becomes Eq. 24. If, in addition, Re[Y(f)] is independent of f, then Eq. 17 is the result. We can now interpret our output energy criterion as a criterion for good approximation to an impulse.

Using Parceval's theorem and our optimum input pulse spectrum of Eq. 46, we can rewrite Eq. 45 as

$$C_3 = \int_{-\infty}^{\infty} \left| D(f) - \frac{D(f)\,|L(f)|^2}{|L(f)|^2 + \mathrm{Re}[Y(f)]} \right|^2 df \tag{47}$$

If the transmission network is bandlimited, the approximation error, C_3, can be considered as the sum of two components – one from bandlimiting and one that results from inaccuracy of approximation within the band. These components are called C_{3B} and C_{3I}, respectively.

If L(f) = 0 for frequencies that are such that $|f| > B$, and if d(t) is given by Eq. 44, with d = 1 and $\beta = 1$, then

$$C_{3B} = 2\int_{B}^{\infty} \frac{\sin^2 \pi f}{(\pi f)^2}\, df = \frac{2}{\pi}\left[\frac{\sin^2 \pi B}{\pi B} + \int_{2\pi B}^{\infty} \frac{\sin u}{u}\, du\right] \tag{48}$$

The quantity C_{3B} is plotted as a function of B in Fig. 4. For $B > 1$, we have $C_{3B} \approx \frac{0.1}{B}$.

The question then arises: How well can r(t) represent d(t) when r(t) must be

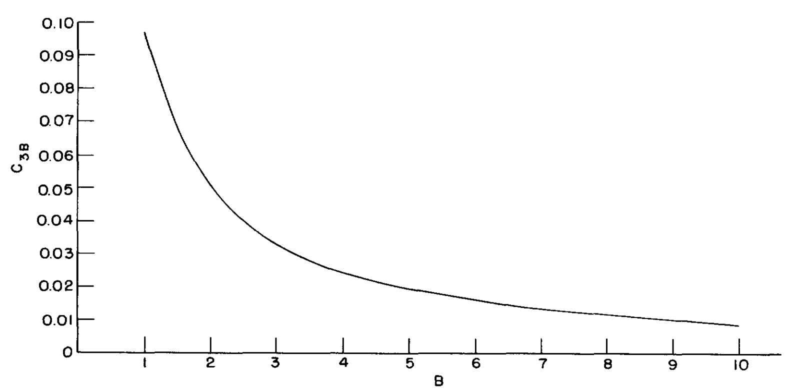

Fig. 4. C_{3B} vs. B.

bandlimited? Let us assume that R(f) matches D(f) exactly within the band (–B, B) and is zero elsewhere. Let r(t) under these conditions be called $r^*(t)$. We obtain $r^*(t)$ as follows:

$$r^*(t-bT) = \int_{-B}^{B} \frac{\sin \pi f}{\pi f} \cos 2\pi ft \, df$$

$$\frac{d}{dt} r^*(t-bT) = -2 \int_{-B}^{B} \sin \pi f \sin 2\pi ft \, df$$

$$= \frac{-2 \sin 2\pi B(t-(1/2))}{2\pi(t-(1/2))} + \frac{2 \sin 2\pi B(t+(1/2))}{2\pi(t+(1/2))}$$

$$r^*(t-bT) = \int_{-\infty}^{t} \frac{d}{dt} r^*(t) \, dt = \frac{1}{\pi} \int_{2\pi B(t-(1/2))}^{2\pi B(t+(1/2))} \frac{\sin x}{x} dx \tag{49}$$

In Fig. 5, $r^*(t)$ is shown with $d(t)$ for $t > bT$ and $B = 1$.

One property of $r^*(t)$, for $B = 1$, is that the envelope of $r^*(t)$ drops off faster than $(0.056/t^2)$ for $|t-bT| > 1$. Thus, if we receive pulses shaped like $r^*(t)$ once each second, if their separate peak values are ± 1.2, and if a sample is taken when each separate peak occurs, then the maximum possible interpulse interference is less than

$$2 \sum_{n=1}^{\infty} \frac{0.06}{n^2} = (0.12) \frac{\pi^2}{6} \approx 0.2 \tag{50}$$

In this example we are signaling at one-half of the Nyquist rate. If we signal at the Nyquist rate, the maximum interpulse interference is approximately equal to the peak value of the largest single received pulse.

The advantages of using digital message values become apparent from the preceding

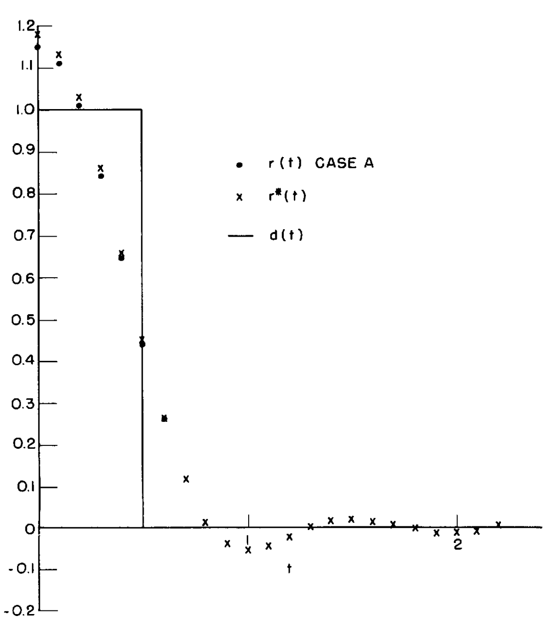

Fig. 5. Graphs of $r^*(t)$, $d(t)$, and $r(t)$ case A.

example. If we choose 4 input message values in such a way that an individual output pulse has peak value +1, $+\frac{1}{2}$, $-\frac{1}{2}$, or −1, and if we replace each sample of the received waveform by the nearest of these values, the resulting sequence will be ideal. That is, each element will be a fixed constant multiplied by an input message value. This is because the interpulse interference is again bounded by 0.2 and hence is never large enough to cause an error. Thus, by restricting our message values to a finite, discrete set we are able to transmit them without error through the idealized link of Fig. 2.

In the discussion of the preceding paragraphs we assumed ideal sampling instants. If the sampling instants drift more than one-fourth of the time T between messages, then, as shown by the plot of $r^*(t)$ of Fig. 5, the maximum interpulse interference is at least doubled. This means that a smaller number of messages is required to obtain perfect transmission. Of course, if the bandwidth B were increased, all other factors remaining the same, then $r^*(t)$ becomes flatter near the ideal sampling instant and thus is less susceptible to the effects of jitter.

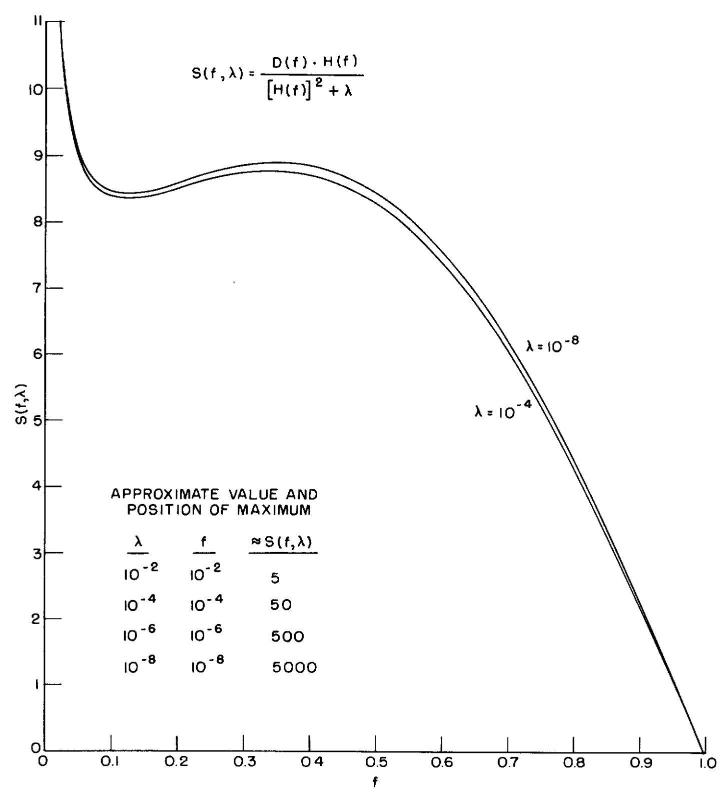

Fig. 6. Input pulse spectrum for two values of λ.

An interesting example of the use of Eq. 46 occurs when the transfer function L(f) has a zero within the band of frequencies to be used. We assume that Y(f) = 1 and that d(t) is given by Eq. 44. We compare two cases. For case A,

$$|L(f)| = |f|^{1/2} \exp(-10|f|^{1/2}) \qquad |f| < 1 \tag{51}$$

For case B

$$|L(f)| = \exp(-10|f|^{1/2}) \qquad |f| < 1 \tag{52}$$

In both cases L(f) is zero outside the band.

For case B the inband error can be made negligibly small. For example, if $\lambda = 10^{-4}$, then r(t) for case B cannot be distinguished from the optimum bandlimited pulse, $r^*(t)$, plotted in Fig. 5. For comparison, r(t), for case A and $\lambda = 10^{-4}$, is also plotted in Fig. 5. Note that the latter function is shifted slightly downward, because the average output of network A must be zero. The input pulse spectrum is plotted for case A and two values of λ in Fig. 6.

III. LOW ERROR PROBABILITY PERFORMANCE CRITERION

3.1 Introduction

Let us now discuss a criterion by which we can compare synchronous digital pulse links of the form depicted in Fig. 3. The adjective "digital" is intended to imply that only a finite, discrete set of amplitudes is possible for the transmitted pulses. This allows us to quantize the samples of the received waveform, and our messages can then be transmitted without errors if noise and interpulse interference are not too great. Using our criterion, we shall optimize the performance of our link model by proper choice of the input pulse shape and linear filtering network (cf. Fig. 3). Thus we shall emphasize problems in the design of digital links, although we shall find that our results apply in other cases.

Another feature that distinguishes Section III from Section II is that for much of Section III we assume knowledge of the power density spectrum of noise that is added in the transmission network (cf. Fig. 3). This will enable us to combat the effects of this noise by means other than increasing signal power.

3.2 Low Error Probability Criterion for a Noiseless Link

We shall now develop a criterion for judging digital links of the form shown in Fig. 2. Our particular assumptions about this model are:

1. The link is digital. That is, each element of the message sequence $\{a_n\}$ is assumed to be selected from the same set, A, of M real numbers, $(A_1, A_2, \dots, A_M)$. As discussed in section 2.1, each element of $\{a_n\}$ corresponds to one transmitted message.

2. We attempt to recover the transmitted sequence $\{a_n\}$ by quantizing the sequence of samples, $\{s_n\}$, rather than by dividing each element of $\{s_n\}$ by r(bT), as discussed in section 2.2. As before, the ideal sampler output is the sequence $\{a_n r(bT)\}$. Since our link is now digital, each element of this sequence is a member of the set $(A_1 r(bT), A_2 r(bT), \dots, A_M r(bT))$.

In the diagram of Fig. 7, we assume that, if a particular output sample s_n differs from $A_i r(bT)$ by less than $W_i/2$ volts, $i = 1,2,\dots,M$, then s_n is instantaneously replaced by $A_i r(bT)$. In Fig. 7 we assume $M = 3$. If s_n lies in none of these M voltage intervals, it is replaced by $A_{M+1}\, r(bT)$, where A_{M+1} is not a member of the set A. The widths W_i, $i = 1, 2, \dots, M$, are chosen in such a way that the voltage intervals about each ideal sampler output do not overlap. Thus, in Fig. 7, the quantizer replaces s_n by $A_1 r(bT)$, s_{n+1} by $A_{M+1}\, r(bT)$, and s_{n+2} by $A_{M+1}\, r(bT)$.

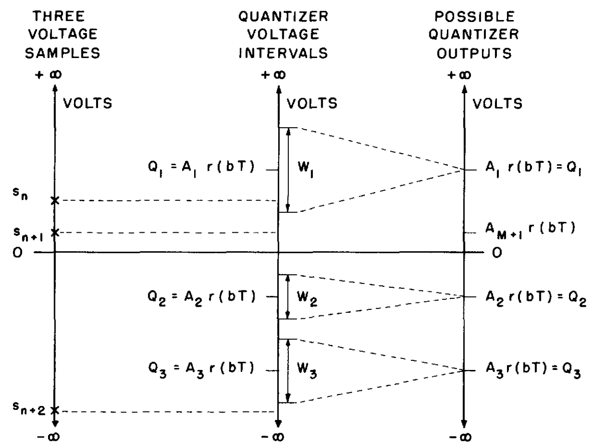

Fig. 7. Quantizer operation diagram.

3. The operation of the clock of Fig. 2 is characterized by the probability density function p(t), where

$$\int_{-\infty}^{t} p(u)\, du \tag{53}$$

is the probability that the actual sampling instant that corresponds to the ideal sampling instant $t = 0$ occurs before time t. The correspondence between the actual and ideal sampling instants was defined in section 2.1. The probability density p(t) is assumed to be nonzero only during the time interval $(-T/2, T/2)$. We also assume that the probability density function for the sampling instant that corresponds to the ideal sampling instant $t = jT$ is $p(t-jT)$, $j = 0, \pm 1, \pm 2, \dots$. Each sampling instant is statistically independent of all others. Physically, we assume that any drifting of the sampling instants in either time direction is prevented, for example, by transmitting special retiming pulses to resynchronize the clock. In this case, thermal noise in the clock could cause our statistically independent jitter.

We say that our link operates perfectly if the actual sequence of quantizer outputs $\{q_n\}$ is the same as our ideal sampler output sequence $\{a_n r(bT)\}$. Note that our definition of the quantizer operation (cf. assumption 2 and Fig. 7), and the fact that our link is digital (cf. assumption 1), imply that the sequence $\{a_n r(bT)\}$ would pass through the quantizer unaffected. The probability that q_n is not equal to $a_n r(bT)$ for a particular integer n, say n = 0, is called the error probability. If the ensemble of message sequences $[\{a_n\}]$ is ergodic, the error probability is independent of n.

If it is equally important that all of our M messages be transmitted correctly, low error probability is a natural criterion for good digital link performance. This is because each incorrect message receives equal weight. But there are two reasons for not using the error probability directly as a criterion. First, calculation of the error probability requires detailed a priori knowledge that would not normally be available to the system designer. For example, it is necessary to know the probability of occurrence of each message. Second, direct application of the error probability in our design problems is mathematically difficult.

Our criterion for low error probability is that an upper bound, U, to the error probability must be small. To derive U we first note that an error occurs whenever

$$|Q_i - f_s(t_s)| > \frac{W_i}{2} \tag{54}$$

where $f_s(t_s)$ is the voltage amplitude of the sampled waveform (cf. Fig. 2) at a sampling instant t_s. The quantities W_i and Q_i are the width and center of the voltage interval within which $f_s(t_s)$ must lie if the correct message is to be received (cf. Fig. 7). Since there are M possible messages, the integer i may be 1, 2, . . . , M−1, or M.

The error probability, P_E, can be written

$$P_E = \sum_{i=1}^{M} \Pr\left[|y_i(t_s)| > \frac{W_i}{2}\right] p'(i) \tag{55}$$

where $\Pr[|y_i(t_s)| > (W_i/2)]$ is the probability that $|y_i(t_s)|$ exceeds $(W_i/2)$; $p'(i)$ is the probability of occurrence of the transmitted message value A_i, $i = 1, 2, \ldots, M$; and $y_i(t_s) = f_d(t_s) - f_s(t_s)$ is the difference between a desired waveform, $f_d(t)$, and the actual sampled waveform, $f_s(t)$, at time $t = t_s$, given that the correct quantizer output at $t = t_s$ is Q_i.

The desired waveform at the sampler input, $f_d(t)$, is defined by

$$f_d(t) = \sum_{n=-\infty}^{\infty} a_n\, d(t-nT) \tag{56}$$

where $\{a_n\}$ is the input message sequence and

$$d(t) = \begin{cases} d & \text{for } p(t-bT) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{57}$$

The real number d is such that

$$A_i d = Q_i \qquad i = 1, 2, \ldots, M \tag{58}$$

As in Section II we assume that the approximate time delay of the link is bT seconds, where b is a positive integer. We recall that p(t−bT) is the probability density function for the sampling instant that occurs during the time interval $\left(bT-\frac{1}{2}T, bT+\frac{1}{2}T\right)$. Thus, at all possible sampling instants, the desired waveform lies in the center of the correct voltage quantization interval. The only possible values for $f_d(t)$ are zero and the center voltages Q_i, $i = 1, 2, \ldots, M$.

We now apply Tchebycheff's theorem to each term in the summation of Eq. 55 to obtain our upper bound, U. Tchebycheff's theorem can be stated as follows (18):

If z is a random variable with mean m and variance σ^2 and if P is the probability that $|z-m|$ exceeds $k\sigma$, then P does not exceed $\sigma^2/(k\sigma)^2$.

Thus, if m_i, the mean of $y_i(t_s)$, is zero, we can write

$$\Pr\left[|y_i(t_s)| > \frac{W_i}{2}\right] \leq \frac{4\sigma_i^2}{W_i^2} \tag{59}$$

by Tchebycheff's theorem, for $i = 1, 2, \ldots, M$. The quantity σ_i^2 is the variance of the random variable $y_i(t_s)$. We can now use Eqs. 55 and 59 to obtain

$$P_E \leq U = \sum_{i=1}^{M} \frac{4\sigma_i^2 p'(i)}{W_i^2} \tag{60}$$

If we assume that the quantization intervals are of equal width, that is, $W_1 = W_2 = \ldots = W_M = W$, then Eq. 60 can be written in the simpler form

$$U = \frac{4\sigma^2}{W^2} \tag{61}$$

where

$$\sigma^2 = \sum_{i=1}^{M} \sigma_i^2 p'(i) \tag{62}$$

3.3 A More Explicit Expression for the Low Error Probability Criterion

In obtaining a more explicit expression for the low error probability criterion we make the same assumptions that we made in section 3.2. In addition, we assume that the ergodic message sequence $\{a_n\}$ has correlation function $\phi_m(n)$ and that

$$E[a_n] = \phi_m(\infty) = 0 \tag{63}$$

where E[ ] denotes the ensemble average of [ ]. We also assume that

$$W_1 = W_2 = \ldots = W_M = W \tag{64}$$

This allows us to use Eq. 61 as our definition for U.

If we are given two possible input pulse shapes $s_1(t)$ and $s_2(t)$, we shall say that $s_1(t)$ is better than $s_2(t)$ if

$$U(s_1(t)) \leq U(s_2(t)) \tag{65}$$

where $U(s(t))$ is the value of U when the input pulse shape $s(t)$ is used. An optimum input pulse shape is one for which U is a minimum.

We recall that the validity of Eqs. 59, 60, and 61 depends on the conditional means, m_i, $i = 1, 2, \dots, M$, being zero. However, we shall now show that in the case of Eq. 61 we need only require that

$$m = E[f_d(t_s)-f_s(t_s)] = \sum_{i=1}^{M} m_i p'(i) = 0 \tag{66}$$

This follows from the fact that the error probability may be written

$$P_E = \Pr\left[|y(t_s)| > \frac{W}{2}\right] \tag{67}$$

if the quantizer voltage intervals have equal width W. The quantity $y(t_s)$ is the difference $f_d(t_s) - f_s(t_s)$. If $m = 0$, Tchebycheff's theorem can be applied directly to Eq. 67 to give Eq. 61. Physically, Eq. 66 requires that the expected value of each sample of the actual received waveform be equal to the value of the desired waveform at that instant.

We shall now write m explicitly in order to obtain sufficient conditions for $m = 0$. Using the definitions of $f_d(t)$ and $f_s(t)$ (cf. Fig. 2 and Eq. 56), we obtain

$$E[f_d(t)-f_s(t)] = \sum_{n=-\infty}^{\infty} E[a_n](d(t-nT)-r(t-nT)) \tag{68}$$

Thus, if, as we have assumed, $E[a_n] = 0$, then

$$m = E[f_d(t_s)-f_s(t_s)] = 0 \tag{69}$$

Before we can minimize U by varying s(t) it is necessary to rewrite U in such a way that its dependence on s(t) becomes more apparent. We first calculate σ^2.

$$y(t) = f_d(t) - f_s(t) = \sum_{n=-\infty}^{\infty} a_n(d(t-nT)-r(t-nT)) \tag{70}$$

$$y^2(t) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_i a_j (d(t-iT)-r(t-iT))(d(t-jT)-r(t-jT)) \tag{71}$$

$$E[y^2(t)] = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \phi_m(i-j)(d(t-iT)-r(t-iT))(d(t-jT)-r(t-jT))$$

$$= \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \phi_m(k)(d(t-iT)-r(t-iT))(d(t+kT-iT)-r(t+kT-iT)) \tag{72}$$

We note that the ensemble average $E[y^2(t)]$ is not, in general, independent of t. That is, the ensemble $[y(t)]$ is nonstationary. This type of nonstationary random process has been discussed by W. R. Bennett (19).

To obtain $E[y^2(t_s)]$ we must also average over the possible instants at which any one sample could be taken. We assume that $(n-(1/2))\, T < t_s < (n+(1/2))\, T$, but we shall find that $E[y^2(t_s)]$ is independent of n, $n = 0, \pm 1, \pm 2, \ldots$. Since $m = 0$,

$$\sigma^2 = E\left[y^2(t_s)\right] = \int_{(n-(1/2))T}^{(n+(1/2))T} p(t-nT)\, E[y^2(t)]\, dt \tag{73}$$

Substitution of Eq. 72 and a change of variable $u = t - iT$ gives

$$\sigma^2 = \sum_{i=-\infty}^{\infty} \int_{(n-(1/2))T-iT}^{(n+(1/2))T-iT} \left(p(u+iT-nT) \sum_{k=-\infty}^{\infty} \phi_m(k)[d(u+kT)-r(u+kT)][d(u)-r(u)] \right) du \tag{74}$$

Performing the summation on i, we obtain

$$\sigma^2 = \int_{-\infty}^{\infty} \left(\sum_{i=-\infty}^{\infty} p(u+iT) \sum_{k=-\infty}^{\infty} \phi_m(k)[d(u+kT)-r(u+kT)][d(u)-r(u)] \right) du \tag{75}$$

For the case of uncorrelated messages

$$\phi_m(k) = 0 \qquad k = \pm 1, \pm 2, \ldots \tag{76}$$

and Eq. 75 becomes

$$\sigma^2 = \phi_m(0) \int_{-\infty}^{\infty} \left(\sum_{i=-\infty}^{\infty} p(u+iT)[d(u)-r(u)]^2 \right) dt \tag{77}$$

For this case we can obtain U by substituting Eq. 77 in Eq. 61.

$$U = \frac{4\phi_m(0)}{W^2} \int_{-\infty}^{\infty} \left(\sum_{i=-\infty}^{\infty} p(u+iT)[d(u)-r(u)]^2 \right) dt \tag{78}$$

We recall that r(t) is the response of the transmission network of Fig. 2 to an input pulse s(t); d(t) is an ideal received pulse shape that is defined in Eq. 57; p(t+iT) is the probability density function for the sampling instant that occurs in the time interval $\left(\left(-i-\frac{1}{2}\right) T, \left(-i+\frac{1}{2}\right) T\right)$; $\phi_m(k)$, $k = 0, \pm 1, \pm 2, \ldots$, is the discrete correlation function of the ensemble of message sequences $[\{a_n\}]$; and W is the voltage width of each of the M voltage quantization intervals (cf. Fig. 7).

We have shown that U is a suitable criterion with which to judge our link, because U is an upper bound to the error probability. Aside from this, U also satisfies more intuitive requirements for a good criterion.

Since there is no noise in our link, errors are caused only by interpulse interference, and U of Eq. 78 is very similar to the low interpulse interference criterion, C_3, of Eq. 45. Both U and C_3 are least squares criteria, but the integrand of U in Eq. 78 is weighted by the sum of the probability density functions p(t+iT), $i = 0, \pm 1, \pm 2, \ldots$. This weighting factor restricts the integration to possible sampling instants and gives more importance to probable sampling instants. This makes U a more satisfying criterion than C_3, because C_3 equally requires r(t) to be small, irrespective of the likelihood of a sample, when d(t) is zero. The choice of r(t) according to either criterion tends to keep r(t) large at the correct sampling instant, that is, the one closest to t = bT. However, the use of U again stresses the importance of the more probable sampling instants.

Let us now write the expression for U of Eq. 61 in the time and frequency domains for the case of correlated messages. Substituting Eq. 75 into Eq. 61 gives

$$U = \frac{4}{W^2} \int_{-\infty}^{\infty} \left(\sum_{i=-\infty}^{\infty} p(u+iT)[d(u)-r(u)] \sum_{k=-\infty}^{\infty} \phi_m(k)[d(u+kT)-r(u+kT)] \right) du \tag{79}$$

We now note that the Fourier transform of

$$f_1(u) = [d(u)-r(u)] \sum_{i=-\infty}^{\infty} p(u+iT) \tag{80}$$

is

$$F_1(f) = \frac{1}{T} \sum_{n=-\infty}^{\infty} P\left(\frac{n}{T}\right)\left[D\left(f-\frac{n}{T}\right)-S\left(f-\frac{n}{T}\right)L\left(f-\frac{n}{T}\right)\right] \tag{81}$$

where P(f), D(f), S(f), and L(f) are the Fourier transforms of p(t), d(t), s(t), and $\ell(t)$, respectively. For example,

$$S(f) = \int_{-\infty}^{\infty} s(t) \exp(-j2\pi ft)\, dt \tag{82}$$

In writing Eq. 81 we use the fact that R(f) = S(f) L(f) (cf. Fig. 2).

The Fourier transform of

$$f_2(u) = \sum_{k=-\infty}^{\infty} \phi_m(k)[d(u+kT)-r(u+kT)] \tag{83}$$

is

$$F_2(f) = \Phi_m(f)[D(f)-S(f)L(f)] \tag{84}$$

where

$$\Phi_m(f) = \sum_{k=-\infty}^{\infty} \phi_m(k) \exp(j2\pi kTf) = \sum_{k=-\infty}^{\infty} \phi_m(k) \cos 2\pi kTf \tag{85}$$

In writing Eq. 85 we use the fact that $\phi_m(k)$ is an even function. Thus, by use of Parceval's theorem, we obtain

$$U = \frac{4}{W^2}\int_{-\infty}^{\infty} F_1(f)\, F_2^*(f)\, df$$

$$= \frac{4}{W^2}\int_{-\infty}^{\infty}\left(\frac{1}{T}\sum_{n=-\infty}^{\infty} P\left(\frac{n}{T}\right)\left[D\left(f-\frac{n}{T}\right)-S\left(f-\frac{n}{T}\right)L\left(f-\frac{n}{T}\right)\right]\Phi_m^*(f)[D^*(f)-S^*(f)L^*(f)]\right) df \tag{86}$$

where the asterisk denotes a complex conjugate.

3.4 A Necessary Condition for an Optimum Transmitted Pulse Shape

We shall now obtain an equation that must be satisfied if U is to be minimized by proper choice of S(f). Using standard techniques of the calculus of variations we assume that there exists a signal pulse spectrum, S(f), that minimizes U (ref. 16). Such spectra exist. For example, if there are no constraints on S(f), we can choose S(f) in such a way that S(f) L(f) = D(f). In this case, U = 0, and this is as small as U can be.

We assume that S(f) is constrained to be nonzero only for some set of frequencies, F. It is convenient, but not necessary, to think of F as a base band of frequencies (-B, B). Physically, this constraint might result from (a) an artificial frequency restriction, such as the allotment of a band of frequencies to one channel of a frequency division system, or (b) a natural restriction, such as excessive transmission attenuation outside a certain band of frequencies or practical difficulties in generating pulses with appreciable energy at arbitrarily high frequencies.

We assume that S(f) is a spectrum that minimizes U and we replace S(f) of Eq. 86 by $S(f) + \alpha\beta(f)$, where α is a real parameter, and $\beta(f)$ is the Fourier transform of a real function. The function $\beta(f)$ is zero if f is not in the set F. Any possible signal pulse spectrum can be represented in the form $S(f) + \alpha\beta(f)$. Since S(f) is an optimum spectrum, U is a minimum if $\alpha = 0$. Thus, we require that

$$\left.\frac{\partial u}{\partial \alpha}\right|_{\alpha=0} = \int_{-\infty}^{\infty}\frac{1}{T}\sum_{n=-\infty}^{\infty} P\left(\frac{n}{T}\right)\left[D\left(f-\frac{n}{T}\right)-S\left(f-\frac{n}{T}\right)L\left(f-\frac{n}{T}\right)\right]\left[\Phi_m^*(f)-\beta^*(f)L^*(f)\right] df$$

$$+\int_{-\infty}^{\infty}\left(\frac{1}{T}\sum_{n=-\infty}^{\infty} P\left(\frac{n}{T}\right)\left[-\beta\left(f-\frac{n}{T}\right)L\left(f-\frac{n}{T}\right)\Phi_m^*(f)\right]\left[D^*(f)-S^*(f)L^*(f)\right]\right) df = 0 \tag{87}$$

This derivative is calculated from Eq. 86. We may rewrite Eq. 87 by making the change of variable $u = (n/T) - f$ in the second integral. We also use the fact that $\beta(-f) = \beta^*(f)$. This follows from our definition of $\beta(f)$. We then have

$$\left.\frac{\partial u}{\partial \alpha}\right|_{\alpha=0} = -\frac{1}{T}\int_{-\infty}^{\infty}\left(\beta^*(f)\,\Phi_m^*(f)\,L^*(f)\sum_{n=-\infty}^{\infty} P\left(\frac{n}{T}\right)\left[D\left(f-\frac{n}{T}\right)-S\left(f-\frac{n}{T}\right)L\left(f-\frac{n}{T}\right)\right]\right) df$$

$$-\frac{1}{T}\int_{-\infty}^{\infty}\left(\beta^*(u)\,L^*(u)\sum_{n=-\infty}^{\infty} P\left(\frac{n}{T}\right)\left[D\left(u-\frac{n}{T}\right)-S\left(u-\frac{n}{T}\right)L\left(u-\frac{n}{T}\right)\right]\Phi_m\left(u-\frac{n}{T}\right)\right) du = 0 \tag{88}$$

In writing the second integral in Eq. 88 we have used the fact from Fourier theory that $S(-f) = S^*(f)$, $L(-f) = L^*(f)$, and $D(-f) = D^*(f)$. If we can show that

$$\Phi_m^*(f) \sum_{n=-\infty}^{\infty} P\left(\frac{n}{T}\right) H\left(f-\frac{n}{T}\right) = \sum_{n=-\infty}^{\infty} P\left(\frac{n}{T}\right) H\left(f-\frac{n}{T}\right) \Phi_m\left(f-\frac{n}{T}\right) \tag{89}$$

where

$$H(f) = D(f) - S(f)\, L(f) \tag{90}$$

then, by substitution of Eq. 89 into Eq. 88, we obtain

$$2\int_{-\infty}^{\infty} -\beta^*(f)\, \Phi_m^*(f)\, L^*(f) \sum_{n=-\infty}^{\infty} \frac{1}{T} P\left(\frac{n}{T}\right) H\left(f-\frac{n}{T}\right) df = 0 \tag{91}$$

The validity of Eq. 89 is shown by transforming its right-hand side. By using the definition of $\Phi_m(f)$ in Eq. 85, the right-hand side of Eq. 89 is

$$\sum_{n=-\infty}^{\infty} P\left(\frac{n}{T}\right) H\left(f-\frac{n}{T}\right) \sum_{k=-\infty}^{\infty} \phi_m(k) \cos 2\pi kT\left(f-\frac{n}{T}\right) \tag{92}$$

Using the fact that $\cos 2\pi kT\left(f-\frac{n}{T}\right) = \cos 2\pi kTf$ and interchanging the order of summation gives

$$\Phi_m(f) \sum_{n=-\infty}^{\infty} P\left(\frac{n}{T}\right) H\left(f-\frac{n}{T}\right) \tag{93}$$

Eq. 85 shows that $\Phi_m^*(f) = \Phi_m(f)$. Thus, expression 93 is equivalent to the left-hand side of Eq. 89. Hence, Eqs. 89 and 91 are valid.

Since $\beta^*(f)$ is zero for all frequencies that are not in the set F and $\beta^*(f)$ can be nonzero at any or all of the frequencies in F, a necessary condition for Eq. 91 to hold is that

$$\Phi_m^*(f)\, L^*(f) \sum_{n=-\infty}^{\infty} \frac{1}{T} P\left(\frac{n}{T}\right) H\left(f-\frac{n}{T}\right) = 0 \qquad \text{for all } f \text{ in } F \tag{94}$$

Using Eq. 90 and the fact that $\Phi_m^*(f) = \Phi_m(f)$, we may rewrite Eq. 94 as

$$\Phi_m(f)\, L^*(f) \sum_{n=-\infty}^{\infty} \frac{1}{T} P\left(\frac{n}{T}\right) S\left(f-\frac{n}{T}\right) L\left(f-\frac{n}{T}\right) = \Phi_m(f)\, L^*(f) \sum_{n=-\infty}^{\infty} \frac{1}{T} P\left(\frac{n}{T}\right) D\left(f-\frac{n}{T}\right) \qquad \text{for all } f \text{ in } F \tag{95}$$

The term

$$\sum_{n=-\infty}^{\infty} \frac{1}{T} P\left(\frac{n}{T}\right) D\left(f-\frac{n}{T}\right)$$

can be recognized as the Fourier transform of $d(t) \sum_{j=-\infty}^{\infty} p(t+jT)$. From the definition of p(t) (Eq. 53 et seq.) and d(t) (Eq. 57) it follows that

$$d(t) \sum_{j=-\infty}^{\infty} p(t-jT) = dp(t-bT) \tag{96}$$

Thus, Eq. 95 may be written as

$$\Phi_m(f)\, L^*(f) \sum_{n=-\infty}^{\infty} \frac{1}{T} P\left(\frac{n}{T}\right) S\left(f-\frac{n}{T}\right) L\left(f-\frac{n}{T}\right)$$

$$= d\Phi_m(f)\, L^*(f)\, P(f) \exp(-j\pi bTf) \qquad \text{for all } f \text{ in } F \tag{97}$$

Therefore, if s(t) is an optimum transmitted pulse shape, then its Fourier transform S(f), must satisfy Eq. 97. We now wish to give some interpretation of Eq. 97.

For all frequencies that are such that

$$\Phi_m(f)\, L^*(f) \neq 0 \tag{98}$$

Eq. 97 is equivalent to

$$\sum_{n=-\infty}^{\infty} \frac{1}{T} P\left(\frac{n}{T}\right) S\left(f-\frac{n}{T}\right) L\left(f-\frac{n}{T}\right) = dP(f) \exp(-j2\pi bTf) \qquad \text{for all } f \text{ in } F' \tag{99}$$

where F' is the set of frequencies that are in the set F and that satisfy condition 98. If F' includes all frequencies, $(-\infty, \infty)$, then the result of taking the Fourier transform of both sides of Eq. 99 is

$$r(t) \sum_{n=-\infty}^{\infty} p(t-nT) = dp(t-bT) \tag{100}$$

Thus, the received pulse shape r(t) should be equal to d(t) when $\sum_{n=-\infty}^{\infty} p(t-nT)$ is nonzero, and r(t) is undefined for other values of t. This agrees with our discussion at the beginning of this section.

In this section we have shown that, if F, and hence F', do not include all frequencies, then Eq. 97, and Eq. 99, must be satisfied. That is, at every frequency at which we are allowed to transmit power, the Fourier transform of the left-hand side of Eq. 100 should equal the Fourier transform of the right-hand side. The left-hand side of Eq. 100 can be interpreted as a periodic function, $\sum_{n=-\infty}^{\infty} p(t-nT)$, that has been amplitude-modulated by a pulse, r(t). Equation 97 was derived previously by the writer in a less direct manner (20).

3.5 Optimum Pulse Shapes for a Noiseless Link

We shall now solve Eq. 99 for $R(f) = S(f)\,L(f)$. This is equivalent to solving for $S(f)$, because

(a) If f is not in the set F, defined at the beginning of section 3.4, then $R(f) = S(f) = 0$. This follows from the definition of F.

(b) If f is in the set F', defined in connection with Eq. 99, then $S(f) = \frac{R(f)}{L(f)}$, because $L(f)$ is not zero. A special case occurs if f is in F, but is not in F'. It can then happen that $L(f)$ is zero, but $R(f)$ is not zero, and thus $S(f)$ will be infinite. This special case, that is, Eq. 99 not equivalent to Eq. 97, will be taken up in section 3.7.

We wish to solve the equation

$$\sum_{n=-\infty}^{\infty} \frac{1}{T}\, P\left(\frac{n}{T}\right) R\left(f-\frac{n}{T}\right) = dP(f)\exp(-j2\pi bTf) \qquad \text{for all } f \text{ in } F' \tag{101}$$

This equation is to be solved for $R(f)$. We first note that, for the case in which F' is contained in an open interval $(-B, B)$ and $1/T \geq 2B$, Eq. 101 reduces to

$$R(f) = dTP(f)\exp(-j2\pi bTf) \qquad \text{for all } f \text{ in } F' \tag{102}$$

because here $F = F'$; $R(f)$ is zero outside F; and $(1/T)$ is chosen large enough so that none of the terms $R\left(f-\frac{n}{T}\right)$ is nonzero in F'. This is best visualized by considering the special case $F' = (-B, B)$.

For example, let $F = F'$ be the open interval ($-B$ cps, B cps) and let $p(t)$ be a unit impulse that occurs at $t = 0$. That is, we assume deterministic, periodic sampling. If $1/T \geq 2B$, Eq. 102 and the definition of F imply that

$$R(f) = \begin{cases} dT\exp(-j2\pi bTf) & \text{for } |f| < B \\ 0 & \text{otherwise} \end{cases} \tag{103}$$

Thus

$$r(t) = \frac{2BdT\sin 2\pi B(t-bT)}{2\pi B(t-bT)} \tag{104}$$

We note that, if $1/T = 2B$, then $r(bT) = d$. We recall that d is the desired received pulse amplitude at its sampling instant. As T decreases, $r(bT)$ decreases. This might be expected from the fact that, if we receive pulses like that of Eq. 104 faster than $2B$ per second, the pulses will overlap. The results of this example, that is, Eqs. 103 and 104, in the special case $1/T = 2B$ agree with those obtained by Nyquist (21).

As another example, let us again consider the assumptions of the previous example, except that we now assume $1/T = B$. In this case Eq. 101 becomes

$$R(f+B) + R(f) + R(f-B) = dT\exp(-j2\pi bTf) \qquad \text{for } |f| < B \tag{105}$$

Unlike the previous example, where the solution was unique, we now have many possible solutions. One solution is

$$R(f) = \begin{cases} \left(1 - \frac{|f|}{B}\right) dT \exp(-j2\pi bTf) & \text{for } |f| < B \\ 0 & \text{otherwise} \end{cases} \tag{106}$$

Another solution is

$$R(f) = \begin{cases} \frac{1}{2} dT \exp(-j2\pi bTf) & \text{for } |f| < B \\ 0 & \text{otherwise} \end{cases} \tag{107}$$

The received pulse shape that corresponds to Eq. 106 is

$$r(t) = \frac{d[\sin 2\pi B(t-bT)]^2}{[2\pi B(t-bT)]^2} \tag{108}$$

This example seems to indicate that if we are willing to signal slower than the Nyquist rate $1/T = 2B$, then we gain more freedom in the choice of our received pulseshape. However, we shall see that this conclusion applies only to deterministic, periodic sampling.

We shall now solve Eq. 101 by rewriting it as a set of simultaneous, linear equations. We assume that the probability density function $p(t)$ is positive over some time interval of positive extent. That is, we shall no longer consider $p(t)$ to be an impulse. The set F', over which $R(f)$ is nonzero, is assumed to lie within some interval $(-B, B)$.

Given any particular frequency f in F' we let $\{x_j\}$ be the set of ordered $(1/T)$ translates of f that lie in $(-B, B)$. That is,

(1) x_i is in $(-B, B)$ for $i = 1, 2, \ldots, N'$

(2) $x_{i-1} = x_i + \frac{1}{T}$ for $i = 2, 3, \ldots, N'$

(3) $x_j = f$ for some j, $j = 1, 2, \ldots, N'$

and N' is the number of elements of the set $\left\{f + \frac{n}{T}\right\}$, $n = 0, \pm 1, \pm 2, \ldots,$ that lie in $(-B, B)$. The integer N' must be finite, because B is assumed finite.

For $f = x_j$ we can write Eq. 101 as

$$\sum_{n=1}^{N'} P\left(\frac{n-j}{T}\right) R(x_n) = dTP(x_j) \exp(-j2\pi bTx_j) \tag{109}$$

If we now let $j = 1, 2, \ldots, N'$, we obtain a set of N' simultaneous linear equations. If N is the number of elements in the set $\{x_j\}$ that belong to the set F', then only N of these N' equations need hold, because Eq. 101 must only be satisfied for f in F'. Now let $\{f_j\}$, $j = 1, 2, \ldots, N$, be the elements of $\{x_j\}$ that are in F', and let $f_1 > f_2 > \ldots > f_N$. Of the N' unknowns in the remaining N equations, only $R(f_1)$, $R(f_2), \ldots, R(f_N)$ are not known to be zero. This follows from our assumption in this section that $F' = F$ and from our definition that $R(f) = 0$ for values of f that are not in F.

If F' is the interval $(-B, B)$, then $N = N'$, and the determinant of our set of linear equations is

$$D_N = \begin{vmatrix} P(0) & P\left(\frac{1}{T}\right) & P\left(\frac{2}{T}\right) & \dots & P\left(\frac{N-1}{T}\right) \\ P\left(\frac{-1}{T}\right) & P(0) & P\left(\frac{1}{T}\right) & \dots & P\left(\frac{N-2}{T}\right) \\ \vdots & \vdots & \vdots & & \vdots \\ P\left(\frac{-N+1}{T}\right) & P\left(\frac{-N+2}{T}\right) & P\left(\frac{-N+3}{T}\right) & \dots & P(0) \end{vmatrix} \tag{110}$$

For other choices of F' some of the columns and their corresponding rows do not appear in D_N.

If D_N is not zero, we can solve our set of equations by Cramer's rule (22). The unique solution for $f = f_j$ is then

$$R(f_j) = \begin{cases} \dfrac{P(f_1)\,D_N^{(1,j)} + P(f_2)\,D_N^{(2,j)} + \dots + P(f_N)\,D_N^{(N,j)}\,dT\ \exp(-j2\pi bTx_j)}{D_N} & \text{for } f_j \text{ in } F' \\ 0 \quad \text{otherwise} \end{cases} \tag{111}$$

where $D_N^{(i,j)}$ is the cofactor of the (i, j) element of D_N.

We shall now prove the validity of Eq. 111 by demonstrating that the determinant D_N is always positive. We note from Fourier theory that, because $p(t)$ is real, its Fourier transform, $P(f)$, has the property that $P^*(f) = P(-f)$. This implies that the matrix corresponding to D_N is Hermitian. It then follows that D_N is positive if its associated quadratic form

$$\sum_{i=1}^{N'} \sum_{k=1}^{N'} A_i^* A_k P\left(\frac{k-i}{T}\right) \tag{112}$$

is positive definite (23). If our set F' does not cover the whole band $(-B, B)$, then some of the coefficients A_l are zero. We wish to show that expression 112 is positive, unless $A_1 = A_2 = \dots = A_{N'} = 0$.

Schetzen has shown that this expression is non-negative definite (24). We shall extend his argument to obtain our result. We first rewrite expression 112.

$$\begin{aligned} \sum_{i=1}^{N'} \sum_{k=1}^{N'} A_i^* A_k P\left(\frac{k-i}{T}\right) &= \int_{-T/2}^{T/2} \sum_{i=1}^{N'} \sum_{k=1}^{N'} A_i^* A_k p(t) \exp\left(-j2\pi\left(\frac{k-i}{T}\right)t\right) dt \\ &= \int_{-T/2}^{T/2} p(t) \left| \sum_{k=1}^{N'} A_k \exp\left(-j2\pi \frac{k}{T} t\right) \right|^2 dt \end{aligned} \tag{113}$$

We now note that, unless $A_1 = A_2 = \dots A_{N'} = 0$, $\sum_{k=1}^{N'} A_k \exp\left(-j2\pi\frac{k}{T}t\right)$ can be zero for at most N' values of t. This follows from the fact that the polynomial $\sum_{k=1}^{N'} A_k z^k$ can have only N' roots, unless $A_1 = A_2 = \dots = A_{N'} = 0$. Since we assumed that p(t) is positive over some subinterval of $(-(T/2),(T/2))$, it follows that the last integral in Eq. 113 will be positive, unless $A_1 = A_2 = \dots = A_{N'} = 0$. This proves that D_N is positive and hence that, under our assumptions, a unique solution to Eq. 101 always exists and is given by Eq. 111.

We note that if p(t) is an impulse our proof does not apply because a zero of $\sum_{k=1}^{N'} A_k \exp\left(-j2\pi\frac{k}{T}t\right)$ can coincide with and cancel the impulse. This agrees with our second example, given earlier in this section, in which the solution of Eq. 101 was not unique and p(t) was assumed to be an impulse.

We shall now work out an example to demonstrate what computations must be made when we apply Eq. 111. The set F' is assumed to be the open interval $(-B,B)$, $1/T = B$, and

$$p(t) = \begin{cases} \frac{2}{T} & \text{for } |t| < \frac{T}{4} \\ 0 & \text{otherwise} \end{cases} \tag{114}$$

Hence,

$$P(f) = \frac{\sin \pi(T/2)\, f}{\pi(T/2)\, f} \tag{115}$$

If we choose $f = x_j$ in $(-B, B)$ but not equal to zero, then $N = N' = 2$. That is, in the set $\left\{f+\frac{n}{T}\right\}$, only f and $f \pm (1/T)$ can lie in $(-B, B)$ because $1/T = B$. If $f = 0$, then $N = N' = 1$, because $\pm B$ do not lie within $(-B, B)$

$$D_2 = \begin{vmatrix} P(0) & P\left(\frac{1}{T}\right) \\ P\left(\frac{-1}{T}\right) & P(0) \end{vmatrix} = P^2(0) - P\left(\frac{1}{T}\right) P\left(\frac{-1}{T}\right) = 1 - \frac{4}{\pi^2} \tag{116}$$

$$D_1 = P(0) = 1 \tag{117}$$

Application of formula 111 now yields

$$R(f) = \begin{cases} \dfrac{\dfrac{\sin[\pi(T/2)f]}{\pi(T/2)\,f} - \dfrac{2}{\pi}\dfrac{\sin[\pi(T/2)(f+1/T)]}{\pi(T/2)(f+1/T)}}{1 - 4/\pi^2}\, dT \exp(-j2\pi bTf) & -B < f < 0 \\ dT \quad f = 0 & \\ \dfrac{\dfrac{\sin[\pi(T/2)f]}{\pi(T/2)\,f} - \dfrac{2}{\pi}\dfrac{\sin[\pi(T/2)(f-1/T)]}{\pi(T/2)(f-1/T)}}{1 - 4/\pi^2}\, dT \exp(-j2\pi bTf) & 0 < f < B \end{cases} \tag{118}$$

It is important to note that the form of Eq. 111 — in particular, the order of the determinant D_N — depends strongly on the rate of signaling, $(1/T)$, because this rate controls the number of elements of $\left\{f+\frac{n}{T}\right\}$ that lie in F'. If F' is a base band of frequencies, (-B, B), and if we signal faster than the Nyquist rate 2B, then Eq. 111 can be written in the simpler form of Eq. 102. As we reduce our signaling rate, R(f), for any particular f, becomes a linear combination of many values of P(f).

3.6 Low Error Probability Criterion for a Noisy Link

In this section we assume that noise added during the transmission of our signal can no longer be neglected. This might come about because an excessive amount of signal power is required to overcome the noise, or because we know the power density spectrum of the noise and wish to use this knowledge to reduce the effects of the noise. We shall consider the synchronous link model of Fig. 8.

With the exception of the gaussian noise generator, the dc voltage source, and the decoder, all of the blocks of this model have been discussed in connection with Figs. 1,2, and 3. We recall (from sec. 2.1) that the message source produces a periodic train of area modulated impulses, m(t), where

$$m(t) = \sum_{n=-\infty}^{\infty} a_n \, u_o(t-nT) \tag{119}$$

and $u_o(t)$ is a unit impulse that occurs at $t = 0$. Each element a_n of the message sequence $\{a_n\}$ is chosen from a finite set A. That is, the link is assumed to be digital. The real numbers, $(A_1, A_2, \ldots, A_M)$, that constitute A, are assumed to be in one-to-one correspondence with M possible transmitted messages. The message probabilities are such that A_i occurs at any particular place in a message sequence with probability $p'(i)$, $i = 1, 2, \ldots, M$, and the correlation function of the message sequence ensemble $[\{a_n\}]$ is $\phi_m(n)$.

The linear, pulse-forming network has output response s(t) to a unit impulse input. The Fourier transform of s(t) is denoted by S(f). The input to the transmission network is then

$$f_I(t) = \sum_{n=-\infty}^{\infty} a_n \, s(t-nT) \tag{120}$$

In a real link the individual input pulses might be transmitted by triggering some nonlinear device every T seconds. However, an input pulse train like $f_I(t)$ can always be represented as the response of a linear network to an impulse train, and this latter representation is convenient conceptually and mathematically.

The linear transmission network has output response $\ell(t)$ to a unit impulse input. The Fourier transform of $\ell(t)$ — that is, the network transfer function — is denoted by L(f). The noise waveform n(t) is added to the output of the transmission network to represent the effects of noise in the transmission network. The random waveform n(t)

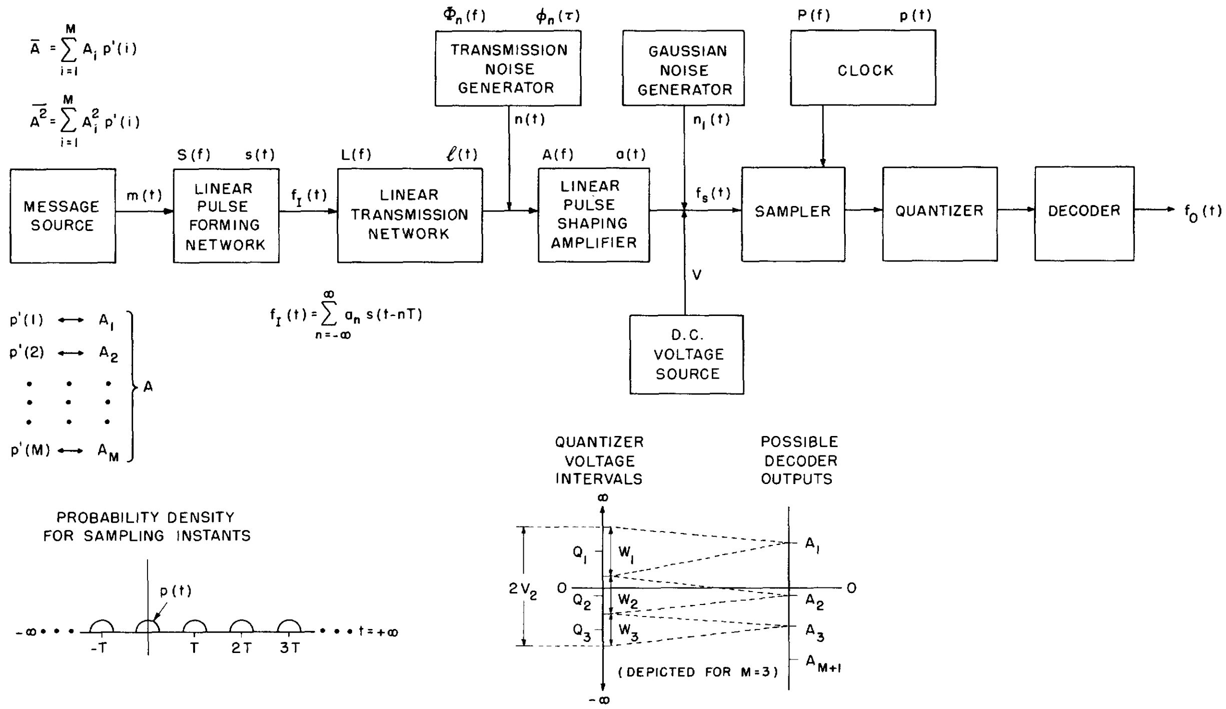

Fig. 8. Synchronous digital link model.

is assumed to be a member of a noise ensemble $[n(t)]$ that has correlation function $\phi_n(\tau)$ and power density spectrum $\Phi_n(f)$.

The resulting waveform is the input for a linear, pulse-shaping amplifier that has output response a(t) to a unit impulse input. The Fourier transform of a(t) is A(f). White, gaussian noise, $n_1(t)$, with mean $\bar{n}_1$ and variance $\sigma^2_{n_1}$ and dc voltage V are added to the output of the amplifier. Noise in the amplifier is represented by $n_1(t)$. The purpose of the dc voltage will be discussed later.

The waveform at the input to the sampler is called $f_s(t)$, and

$$f_s(t) = \sum_{n=-\infty}^{\infty} a_n \, r(t-nT) + n_1(t) + V + \int_{-\infty}^{\infty} n(x) \, a(t-x) \, dx \tag{121}$$

where

$$r(t) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} s(x)(u-x) \, dx \right] a(t-u) \, du \tag{122}$$

The linearity of the transmission network and the amplifier enables us to interpret each term of $f_s(t)$ as the result of a separate cause. The first term is the result of passing the input waveform, $f_I(t)$, through the transmission network and amplifier; the next two terms are the direct result of their addition to the amplifier output; and the fourth term is the result of passing noise waveform n(t) through the amplifier.

The sampler instantaneously samples its input whenever it receives an impulse from the clock. We assume, as before, that there exists a small time interval about each time instant jT, $j = 0, \pm 1, \pm 2, \ldots$. These time intervals have the property that one and only one sampling instant occurs within each interval. The probability density function for the sampling instant that occurs in the j^{th} interval is $p(t-jT)$, $j = 0, \pm 1, \pm 2, \ldots$. See Fig. 8.

We assume that the quantizer operates as discussed in connection with Fig. 7. That is, the sequence of sampler outputs, $\{s_n\}$, is sorted sequentially into M voltage intervals, each of which corresponds to a distinct member of the message set A. The decoder instantaneously replaces each sample by an element of A according to this correspondence. If a voltage sample happens to lie in none of the M intervals, the decoder replaces it by the number A_{M+1}, where A_{M+1} is not an element of the set A. The operation of the quantizer and decoder is depicted in Fig. 8, where it is assumed that M = 3, and the M voltage intervals are contiguous and lie within a voltage interval of width $2v_2$ volts.

We say that the link performs perfectly if the sequence of decoder outputs is the same as the message sequence, $\{a_n\}$, except for a finite time delay. The probability that these sequences will differ in any one place is called the error probability. The discussion in section 3.2 on the use of an upper bound to the error probability as a criterion for link performance and our reasons for not using the error probability directly apply to our present case as well.

We also note that the necessary and sufficient condition for an error given in Eq. 54 is valid for our noisy model. Also, the steps that lead from Eq. 54 to Eqs. 60 and 61 hold because they did not depend on the fact that our link was noiseless. In using Eqs. 54-62 in connection with our model of Fig. 8 we need only use the fact that $f_s(t)$ is now defined by Eq. 121.

Thus, from Eqs. 60 and 61, our criterion for good link performance in our model of Fig. 8 is that U be small, where

$$U = \sum_{i=1}^{M} \frac{4\sigma_i^2 p'(i)}{W_i^2} \tag{123}$$

in the general case, and

$$U = \frac{4\sigma^2}{W^2} \tag{124}$$

in the case of equally wide quantization intervals.

3.7 Necessary Conditions for Optimum Pulse Shapes in a Noisy Link

We shall now use the methods developed in section 3.4 to minimize U by varying the input pulse spectrum, S(f), and the amplifier transfer function, A(f).

How does this problem arise? We recall from the discussion of Fig. 3 that the linear amplifier of Fig. 8 is unnecessary in the sense that the link behavior is determined by the over-all transfer function S(f) L(f) A(f), if the link is noiseless. That is, if we vary A(f), keeping S(f) L(f) A(f) constant, the performance of the link is unaffected.

If transmission noise with known power density spectrum is present, it is possible to choose A(f) in such a way that much of the noise can be filtered out. However, such a choice can make it impossible to achieve received pulses that are optimum in the sense of Section II or section 3.5. For example, good noise filtering might require a zero in the transfer function A(f) at a frequency at which an optimum received pulse spectrum is not zero. The following questions then arise: At any given frequency should A(f) be chosen primarily to cut out noise or primarily to enhance the received pulse shape? If the transmitted power is to be conserved, should S(f) be chosen primarily to give a high signal-to-noise ratio at the amplifier input or primarily to obtain low interpulse interference?

These questions can be answered within the framework of our model by choosing S(f) and A(f) to minimize U and to conserve transmitted power. As in section 3.4, we assume that all M voltage quantization intervals are of equal width. This allows us to use the expression for U in Eq. 124. We also assume that the noise random variables are statistically independent of each other and of the message random variable.

As shown in section 3.3 (cf. Eq. 66), if Eq. 61, and hence Eq. 124, are to be valid, it must be true that

$$m = E[y(t_s)] = E[f_d(t_s)-f_s(t_s)] = 0 \tag{125}$$

To calculate m we first use our definitions of $f_s(t)$ in Eq. 121 and $f_d(t)$ in Eq. 56 to write

$$y(t) = f_d(t) - f_s(t) = \sum_{n=-\infty}^{\infty} a_n[d(t-nT)-r(t-nT)] - n_1(t) - V - n'(t) \tag{126}$$

where

$$n'(t) = \int_{-\infty}^{\infty} n(x)\, a(t-x)\, dx \tag{127}$$

Thus

$$E[y(t)] = \overline{A} \sum_{n=-\infty}^{\infty} [d(t-nT)-r(t-nT)] - \overline{n}_1 - V - \overline{n'} \tag{128}$$

where $\overline{A}$, $\overline{n}_1$, and $\overline{n'}$ are the ensemble averages of a_n, $n_1(t)$, and $n'(t)$, respectively. Since $E[y(t)] = 0$ implies that $E[y(t_s)] = 0$, sufficient conditions for Eq. 125 to be satisfied are

$$\overline{A} = \sum_{i=1}^{M} A_i p'(i) = 0 \tag{129}$$

and

$$V = -\overline{n}_1 - \overline{n'} \tag{130}$$

We assume that the set A has been chosen in such a way that Eq. 129 is satisfied and that the dc voltage V satisfies Eq. 130.

Thus, Eq. 125 is satisfied, and U of Eq. 124 is a valid upper bound to the error probability. Physically, Eq. 125 requires that the expected value of each actual sample be equal to the value of the desired waveform at the sampling instant.

We calculate U by first calculating σ^2. Using Eqs. 129 and 130 and our assumption that the random variables a_n, $n_1(t)$, and n(t) are statistically independent, we have

$$E[y^2(t)] = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \phi_m(i-j)[d(t-jT)-r(t-jT)][d(t-iT)-r(t-iT)]$$
$$+ E\left[(n_1(t)+n'(t)+V)^2\right] \tag{131}$$

where

$$E\left[(n_1(t)+n'(t)=V)^2\right] = \int_{-\infty}^{\infty} A(f)\, A^*(f)\, \phi_n(f)\, df + \sigma_{n_1}^2 - (\overline{n'})^2 \tag{132}$$

In general, $E[y^2(t)]$ is not independent of t. That is, $[y(t)]$ is a nonstationary ensemble. To calculate σ^2 we average over the possible times that any one sample can be taken. We assume that $(n-(1/2))\,T < t_s < (n+(1/2))\,T$, but, as before, we shall find that σ^2 is independent of n.

$$\sigma^2 = E\left[y^2(t_s)\right] = \int_{(n-(1/2))T}^{(n+(1/2))T} p(t-nT)\, E[y^2(t)]\, dt \tag{133}$$

If we make a change of summation index k = i - j in Eq. 131, as we did in Eq. 72, and substitute the result in Eq. 133, making a change of variable u = t - iT, then we have

$$\sigma^2 = \sum_{i=-\infty}^{\infty} \int_{(n-(1/2))T-iT}^{(n+(1/2))T-iT} \left(p(u+iT-nT) \sum_{k=-\infty}^{\infty} \phi_m(k)[d(u+kT)-r(u+kT)][d(u)-r(u)] \right) du$$
$$+ \int_{-\infty}^{\infty} A(f)\, A^*(f)\, \Phi_n(f)\, df + \sigma_{n_1}^2 - (\overline{n'})^2 \tag{134}$$

Noting that the first term of Eq. 134 is the same as the right-hand side of Eq. 74, we can use Eqs. 74, 79, and 86, and Eq. 124 to write

$$U = \frac{4}{W^2} \int_{-\infty}^{\infty} \left(\frac{1}{T} \sum_{n=-\infty}^{\infty} P\left(\frac{n}{T}\right) \left[D\left(f-\frac{n}{T}\right) - S\left(f-\frac{n}{T}\right) L\left(f-\frac{n}{T}\right) A\left(f-\frac{n}{T}\right) \right] \Phi_m(f)[D^*(f)-S^*(f)L^*(f)A^*(f)] \right) df$$
$$+ \frac{4}{W^2} \left[\int_{-\infty}^{\infty} A(f)\, A^*(f)\, \Phi_n(f)\, df + \sigma_{n_1}^2 - (\overline{n'})^2 \right] \tag{135}$$

In writing Eq. 135, we use the fact that $\Phi_m(f) = \Phi_m^*(f)$, from Eq. 85, and the fact that in this section $R(f) = S(f)\ L(f)\ A(f)$, instead of $R(f) = S(f)\ L(f)$, as in section 3. 3. The function $R(f)$ is the Fourier transform of the pulse shape $r(t)$.

To obtain our optimum input pulse spectrum, $S(f)$, and optimum amplifier transfer function, $A(f)$, we must minimize U by varying $S(f)$ and $A(f)$, subject to the constraint that our measure of average transmitter power,

$$\int_{-\infty}^{\infty} S(f)\ S^*(f)\ df \tag{136}$$

is held constant. As in Section II, we incorporate our constraint by the method of Lagrange multipliers (17).

We shall obtain an equation that specifies the optimum $S(f)$ for any given $A(f)$ and another equation giving the optimum $A(f)$ for any given $S(f)$. These equations will be solved simultaneously to give the optimum pair $(S(f), A(f))$. We choose this method because the methods and results of section 3. 4 can be used in a natural manner. Would the same optimum pair $(S(f), A(f))$ result if we considered varying them simultaneously? The answer is yes, and this can be proved by comparing the results of both procedures. We will not go through this proof. However, it will be made plausible by demonstrating the uniqueness of our optimum pair.

We first note that varying $S(f)$ affects only the first term of Eq. 135, and this term is equivalent to Eq. 86 because, if $A(f)$ is given, we can combine $A(f)$ and $L(f)$ into one known function. For convenience, we denote the right-hand side of Eq. 86 by U_{86}. Our problem is then to minimize $U_{86} + \lambda \int_{-\infty}^{\infty} S^*(f)\ S(f)\ df$ by varying $S(f)$ where λ is a Lagrange multiplier (17). Using the discussion and definitions of the paragraph that precedes Eq. 87, a necessary condition for such a minimum is that

$$\left.\frac{\partial U_{86}}{\partial \alpha}\right|_{\alpha=0} + \left.\frac{\partial}{\partial \alpha}\left[\lambda \int_{-\infty}^{\infty} [S^*(f)+\alpha\beta^*(f)][S(f)+\alpha\beta(f)]\ df\right]\right|_{\alpha=0} = 0 \tag{137}$$

By evaluating the derivative in the second term of Eq. 137 and using Eqs. 88-91 to transform the first term, Eq. 137 becomes

$$-2\int_{-\infty}^{\infty} \beta^*(f)\left[\Phi_m^*(f)\ L^*(f) \sum_{n=-\infty}^{\infty} \frac{1}{T}\ P\left(\frac{n}{T}\right) H\left(f-\frac{n}{T}\right) - \lambda S(f)\right] df = 0 \tag{138}$$

where $H(f)$ is given by Eq. 90, with $L(f)$ replaced by $L(f)\ A(f)$. Since $\beta^*(f)$ can be nonzero at any or every frequency, it follows that

$$\phi_m^*(f)\ L^*(f)\ A^*(f) \sum_{n=-\infty}^{\infty} \frac{1}{T}\ P\left(\frac{n}{T}\right)\left[S\left(f-\frac{n}{T}\right)L\left(f-\frac{n}{T}\right)A\left(f-\frac{n}{T}\right)-D\left(f-\frac{n}{T}\right)\right] + \lambda S(f) = 0 \tag{139}$$

Eq. 139 is our necessary condition for U of Eq. 135 to be minimized by varying $S(f)$, subject to the constraint that expression 136 remains constant.

Similarly, if (a) we consider S(f) given and combine it with L(f) in Eq. 135 and (b) identify A(f) of Eq. 135 with S(f) of Eq. 86, then minimizing U of Eq. 135 by varying A(f) is equivalent to minimizing $U_{86} + \int_{-\infty}^{\infty} S(f)\, S^*(f)\, \Phi_n(f)\, df$ by varying S(f). The arguments used in obtaining Eqs. 137, 138, and 139 may then be repeated to show that a necessary condition for such a minimum is

$$\Phi_m^*(f)\, L^*(f) \sum_{n=-\infty}^{\infty} \frac{1}{T} P\left(\frac{n}{T}\right)\left[S\left(f-\frac{n}{T}\right)L\left(f-\frac{n}{T}\right)-D\left(f-\frac{n}{T}\right)\right] + \Phi_n(f)\, S(f) = 0 \tag{140}$$

To obtain from Eq. 140 a necessary condition for U of Eq. 135 to be a minimum by varying A(f) we must reverse substitutions (a) and (b). Thus, we replace S(f) by A(f) and L(f) by S(f) L(f) in Eq. 140. The result is

$$\Phi_m^*(f)\, S^*(f)\, L^*(f) \sum_{n=-\infty}^{\infty} \frac{1}{T} P\left(\frac{n}{T}\right)\left[S\left(f-\frac{n}{T}\right)L\left(f-\frac{n}{T}\right)A\left(f-\frac{n}{T}\right)-D\left(f-\frac{n}{T}\right)\right] + \Phi_n(f)\, A(f) = 0 \tag{141}$$

In obtaining Eq. 97 from Eq. 95 it is shown that

$$\sum_{n=-\infty}^{\infty} \frac{1}{T} P\left(\frac{n}{T}\right) D\left(f-\frac{n}{T}\right) = dP(f) \exp(-j2\pi bTf) \tag{142}$$

The time delay bT originates in the definition of d(t) in Eq. 57. This delay can presumably be chosen by the system designer. If we substitute Eq. 142 into Eqs. 139 and 141 and use the fact that $\Phi^*(f) = \Phi(f)$ (cf. Eq. 85), we have

$$\Phi_m(f)\, L^*(f)\, A^*(f) \sum_{n=-\infty}^{\infty} \frac{1}{T} P\left(\frac{n}{T}\right) S\left(f-\frac{n}{T}\right) L\left(f-\frac{n}{T}\right) A\left(f-\frac{n}{T}\right) + \lambda S(f)$$

$$= d\Phi_m(f)\, L^*(f)\, A^*(f)\, P(f) \exp(-j2\pi bTf) \tag{143}$$

and

$$\Phi_m(f)\, L^*(f)\, S^*(f) \sum_{n=-\infty}^{\infty} \frac{1}{T} P\left(\frac{n}{T}\right) A\left(f-\frac{n}{T}\right) L\left(f-\frac{n}{T}\right) S\left(f-\frac{n}{T}\right) + \Phi_n(f)\, A(f)$$

$$= d\Phi_m(f)\, L^*(f)\, S^*(f)\, P(f) \exp(-j2\pi bTf) \tag{144}$$

Equations 143 and 144 are necessary conditions for U of Eq. 135 to be minimized by varying S(f) and A(f), subject to the constraint that expression 136 be held constant. These equations are written in a form that emphasizes their symmetry in A(f) and S(f).

3.8 Optimum Pulse Shapes for a Noisy Link

We now wish to solve Eqs. 143 and 144 for S(f) and A(f). Let us first consider the case in which L(f) is bandlimited to frequencies less than B cps and the signaling rate, (1/T), is not less than the Nyquist rate, 2B. In this case Eqs. 143 and 144 can be written as

$$\frac{1}{T}\Phi_m(f)\, L^*(f)\, A^*(f)\, S(f)\, L(f)\, A(f) + \lambda S(f) = d\Phi_m(f)\, L^*(f)\, A^*(f)\, P(f) \exp(-j2\pi bTf) \tag{145}$$

and

$$\frac{1}{T}\Phi_m(f)\, L^*(f)\, S^*(f)\, A(f)\, L(f)\, S(f) + \Phi_n(f)\, A(f) = d\Phi_m(f)\, L^*(f)\, S^*(f)\, P(f) \exp(-j2\pi bTf) \tag{146}$$

We have used the fact that $P(0) = 1$, since $p(t)$ is a probability density function.

For example, let us consider the special case in which $1/T \geq 2B$, $\Phi_n(f) = 0$, and $\lambda = 0$. That is, we remove our constraint on transmitted power. Then both Eqs. 145 and 146 reduce to

$$R(f) = S(f)\, L(f)\, A(f) = dTP(f) \exp(-j2\pi bTf) \tag{147}$$

Thus, when we remove the noise, our equations imply the results of our noiseless case, because Eq. 147 is the same as Eq. 102.

We now rewrite Eqs. 145 and 146 in the following form:

$$S(f) = \frac{d\Phi_m(f)\, L^*(f)\, A^*(f)\, P(f) \exp(-j2\pi bTf)}{(1/T)\, \Phi_m(f) |L(f)A(f)|^2 + \lambda} \tag{148}$$

and

$$A(f) = \frac{d\Phi_m(f)\, L^*(f)\, S^*(f)\, P(f) \exp(-j2\pi bTf)}{(1/T)\, \Phi_m(f) |L(f)S(f)|^2 + \Phi_n(f)} \tag{149}$$

If we denote the phase functions of $L(f)$, $A(f)$, $S(f)$, and $P(f)$ by $\rho_L(f)$, $\rho_A(f)$, $\rho_S(f)$, and $\rho_P(f)$, then matching phases on both sides of Eqs. 148 and 149 shows that $\rho_S(f)$ and $\rho_A(f)$ must be chosen in such a way that

$$\rho_S(f) + \rho_L(f) + \rho_A(f) = \rho_P(f) - 2\pi bTf \tag{150}$$

We note that the phases of the power density spectra $\Phi_m(f)$ and $\Phi_n(f)$ are zero.

Equating the magnitudes of both sides of Eqs. 148 and 149 gives

$$|S(f)| = \frac{d\Phi_m(f) |L(f)|\, |A(f)|\, |P(f)|}{(1/T)\, \Phi_m(f) |L(f)|^2 |A(f)|^2 + \lambda} \tag{151}$$

and

$$|A(f)| = \frac{d\Phi_m(f) |L(f)|\, |S(f)|\, |P(f)|}{(1/T)\, \Phi_m(f) |L(f)|^2 |S(f)|^2 + \Phi_n(f)} \tag{152}$$

This pair may be solved simultaneously to give

$$|S(f)|^2 = \frac{dT|P(f)|}{|L(f)|}\left(\frac{\Phi_n(f)}{\lambda}\right)^{1/2} - \frac{T\Phi_n(f)}{\Phi_m(f)|L(f)|^2} \tag{153}$$

and

$$|A(f)|^2 = \frac{dT|P(f)|}{|L(f)|}\left(\frac{\lambda}{\Phi_n(f)}\right)^{1/2} - \frac{\lambda T}{\Phi_m(f)|L(f)|^2} \tag{154}$$

The symmetry between S(f) and A(f) is now quite apparent because Eqs. 153 and 154 are identical except for an interchange in the positions of λ and $\Phi_n(f)$. Equations 153 and 154 are valid only when

$$d|P(f)| > \frac{(\lambda\Phi_n(f))^{1/2}}{\Phi_m(f)|L(f)|} \tag{155}$$

This condition results from the fact that $|A(f)|$ and $|S(f)|$ cannot be negative. At frequencies for which inequality 155 is not satisfied we must use the other admissible solution of Eqs. 151 and 152. That is

$$S(f) = A(f) = 0 \tag{156}$$

We can now use Eqs. 153 and 154 to calculate the magnitude of the Fourier transform of the received pulse shape r(t). We have

$$\begin{aligned} |S(f)|^2|L(f)|^2|A(f)|^2 &= \left(dT|P(f)|\left(\frac{\Phi_n(f)}{\lambda}\right)^{1/2} - \frac{T\Phi_n(f)}{\Phi_m(f)|L(f)|}\right) \\ &\quad \times \left(dT|P(f)|\left(\frac{\lambda}{\Phi_n(f)}\right)^{1/2} - \frac{T\lambda}{\Phi_m(f)|L(f)|}\right) \\ &= \left(dT|P(f)| - \frac{T(\lambda\Phi_n(f))^{1/2}}{\Phi_m(f)|L(f)|}\right)^2 \end{aligned} \tag{157}$$

Therefore,

$$|R(f)| = |S(f)L(f)A(f)| = dT|P(f)| - \frac{T(\lambda\Phi_n(f))^{1/2}}{\Phi_m(f)|L(f)|} \tag{158}$$

provided inequality 155 holds. Otherwise

$$|R(f)| = 0 \tag{159}$$

The phase function of R(f) is, from Eq. 150,

$$\rho_R(f) = \rho_P(f) - 2\pi bTf \tag{160}$$

With the results of Eqs. 150, 153, 154, and 158 and inequality 155, it is now possible to answer the questions posed at the beginning of section 3.7. We first consider λ,

$\Phi_m(f)$, and L(f) to be fixed. Also, we denote $\dfrac{(\lambda\Phi_n(f))^{1/2}}{\Phi_m(f)|L(f)|}$ by X(f).

(a) If X(f) is small compared to $d|P(f)|$, that is, $\Phi_n(f)$ is small, then from Eqs. 153 and 154

$$|A(f)|^2 \approx \frac{dT|P(f)|}{|L(f)|}\left(\frac{\lambda}{\Phi_n(f)}\right)^{1/2} \tag{161}$$

$$|S(f)|^2 \approx \frac{dT|P(f)|}{|L(f)|}\left(\frac{\Phi_n(f)}{\lambda}\right)^{1/2} \tag{162}$$

and from Eq. 158

$$|R(f)| \approx dT|P(f)| \tag{163}$$

We can think of $|S(f)|^2$ in terms of its three factors, $dT|P(f)|$, $\dfrac{1}{|L(f)|}$, and $\left(\dfrac{\Phi_n(f)}{\lambda}\right)^{1/2}$. The first factor is the spectrum of the optimum noiseless pulse discussed in section 3.5. Thus, we should try to transmit pulses that result in low interpulse interference at the receiver. The second factor compliments the loss of the transmission medium to help preserve our desired frequency content. The third factor modifies our expression in such a way that more transmitted power is allotted to frequencies at which the noise is highest. Thus, we should also try to obtain a favorable signal-to-noise ratio at the amplifier input. We note that $|A(f)|^2$ of Eq. 161 contains the inverse factor $\left(\dfrac{\lambda}{\Phi_n(f)}\right)^{1/2}$. Thus, we reduce our gain where the most noise power lies.

(b) If $\Phi_n(f)$ becomes larger, so that $X(f) \approx d|P(f)|$, then $|A(f)|^2$, $|S(f)|^2$, and $|R(f)|$ all become very small, and, from Eqs. 156 and 159, if $X(f) \geq d|P(f)|$ then we should not transmit or amplify at such frequencies; that is, $A(f) = S(f) = R(f) = 0$.

We have explored the critical nature of inequality 155 by increasing the noise power, $\Phi_n(f)$. The same behavior results from decreasing the input signal power, i.e., increasing λ, or decreasing the message power density, $\Phi_m(f)$, or increasing the loss of the transmission medium, i.e., reducing $|L(f)|$.

The optimum transmitted and received pulse spectra, S(f) and R(f), and the optimum amplifier transfer function, A(f), are given by Eqs. 150-158. Some limitations of these results are: (a) We did not constrain our pulses s(t) and a(t) to be zero for $t < 0$. (b) We assumed the quantizer intervals are equally wide. (c) We assumed that the signaling rate is not less than the Nyquist rate.

The constraint mentioned in limitation (a) is not used because it is mathematically inconvenient; the results are difficult to interpret physically; a solution with fewer constraints exhibits the ideal behavior that we are seeking; and, if enough transmission time delay is allowed, our optimum pulse shapes can be closely approximated by realizable pulse shapes.

Limitation (b) cannot be easily disposed of. The difficulty is that we must use Eq. 60, rather than Eq. 61, as our definition for the upper bound U. The variational and computational problems connected with this bound are much more difficult.

Limitation (c) can be removed by using the methods of section 3.5. That is, Eqs. 143 and 144 must each be expressed as a set of simultaneous linear equations. These sets are solved separately. We then have two equations similar to Eqs. 148 and 149. These must be solved simultaneously for S(f) and A(f).

We have also assumed that the sampler was able to instantaneously sample its received waveform. Physically, we are only able to take smeared samples of a waveform. We assume a smeared sample of x(t) at time t_s can be represented by

$$\int_{-\infty}^{\infty} w(t)\, x(t)\, dt \tag{164}$$

where $w(t) \geq 0$, and $w(t) \neq 0$ for $t_s - \beta < t < t_s + \beta$ and $2\beta < T$. If in addition, the sampling time, t_s, is a random variable that is characterized by the probability density p(t), then the expected value of our smeared sample is

$$\int_{-\infty}^{\infty} p(t)\left[\int_{-\infty}^{\infty} w(u-t)\, x(u)\, du\right] dt = \int_{-\infty}^{\infty} p'(u)\, x(u)\, du \tag{165}$$

where

$$p'(u) = \int_{-\infty}^{\infty} p(t)\, w(u-t)\, dt \tag{166}$$

and $p'(u) \geq 0$. Thus, we can replace p(t) by p'(t) to obtain optimum pulse shapes for the case of noninstantaneous sampling. The only change in our formulas will be the introduction of the normalizing constant

$$C = \int_{-\infty}^{\infty} p'(u)\, du \tag{167}$$

due to the fact that the area of p(t) is 1, while the area of p'(t) depends on the form of w(t). For example, if the receiver integrates its input over a finite time interval

$$w(t) = \begin{cases} 1/2\beta & t_s - \beta < t < t_s + \beta \\ 0 & \text{otherwise} \end{cases} \tag{168}$$

Thus, C = 1 and the formulas of Section III apply directly, with p(t) replaced by p'(t).

IV. RECONSTRUCTION OF A RANDOM WAVEFORM FROM UNIFORMLY SPACED SAMPLES

4.1 The Problem

Let $[x(t)]$ be an ensemble of random waveforms that has correlation function $\phi(\tau)$, and let $[f(t)]$ be an ensemble of waveforms that are approximations to the waveforms of $[x(t)]$. We assume that there exists a one-to-one correspondence between the waveforms of $[x(t)]$ and their approximations in $[f(t)]$.

If $x(t)$ is a particular member of $[x(t)]$, then $f(t)$, the corresponding member of $[f(t)]$, is given by

$$f(t) = \sum_{n=-N}^{N} x(nT)\, s(t-nT) \tag{169}$$

where the sequence $\{x(nT)\}$ represents samples of $x(t)$ that are taken every T seconds within the time interval $(-NT, NT)$, and $s(t)$ is an interpolatory function to be determined.

We wish to discover what interpolatory function, $s(t)$, will make $f(t)$ the best approximation to $x(t)$ for all members of the ensemble $[x(t)]$. If N is finite, our criterion for best approximation is that

$$E_N = \frac{1}{2NT}\int_{-NT}^{NT} E\left[(x(t)-f(t))^2\right] dt \tag{170}$$

be as small as possible. If N is infinite, we require that

$$E_\infty = \frac{1}{T}\int_0^T E\left[(x(t)-f(t))^2\right] dt \tag{171}$$

be as small as possible, where E[ ] denotes the ensemble average of [ ].

Because it is both practically and theoretically interesting, the method of reconstructing a random waveform that is discussed in connection with Eq. 169 has received considerable attention (e.g., refs. 8, 9, and 25). However, this attention has been centered on the reconstruction of bandlimited waveforms from uniformly spaced samples. This is because, if the samples are taken at a rate that is not slower than the Nyquist rate, the waveform can be reconstructed with zero mean-square error. This result has been extended to certain types of non-uniform sampling (9). Through our investigation we wish to discover how well waveforms of arbitrary bandwidth can be represented (a) over a finite time interval by a finite number of samples and (b) over an infinite interval by an infinite number of samples.

Our criteria of Eqs. 170 and 171 are essentially mean-square criteria. However, if we wish to grade the reconstructions that use different interpolatory functions by attaching a positive, real number to each, then it is not sufficient to use the ensemble average of $(x(t)-f(t))^2$. This is because $[x(t)-f(t)]$ is generally a nonstationary ensemble, (cf. ref. 19 and the discussion following Eq. 72). In the case of a finite number of samples, we perform an additional time average over the time interval (-NT, NT) to remove the time dependence of the ensemble average. In the case of an infinite number of samples the ensemble average is periodic in time, with period T (19). In this case we need only perform our time average over one period.

4.2 The Case of a Finite Number of Samples

In this section our problem is to minimize E_N of Eq. 170 by varying the interpolatory pulse s(t). We must express E_N as a functional of s(t) and begin by calculating $E[(x(t)-f(t))^2]$.

$$(x(t)-f(t))^2 = x^2(t) - 2 \sum_{m=-N}^{N} x(t)\, x(mT)\, s(t-mT) + \sum_{m=-N}^{N} \sum_{n=-N}^{N} x(mT)\, x(nT)\, s(t-mT)\, s(t-nT) \tag{172}$$

Since $\phi(\tau)$ is the autocorrelation of $[x(t)]$, we have

$$E[x(t)-f(t)^2] = E[x^2(t)] - 2 \sum_{m=-N}^{N} \phi(t-mT)\, s(t-mT) + \sum_{m=-N}^{N} \sum_{n=-N}^{N} \phi([m-n]T)\, s(t-mT)\, s(t-nT) \tag{173}$$

Since $[x(t)]$ is assumed to be stationary, $E[x^2(t)]$ is a constant, which we shall call C. The expression to be minimized becomes

$$E_N = \frac{1}{2NT} \int_{-NT}^{NT} \left(C - 2 \sum_{m=-N}^{N} \phi(t-mT)\, s(t-mT) + \sum_{m=-N}^{N} \sum_{n=+N}^{-N} \phi([m-n]T)\, s(t-mT)\, s(t-nT) \right) dt \tag{174}$$

Next we change summation index, letting $m - n = k$.

$$E_N = \frac{1}{2NT} \int_{-NT}^{NT} \left(C - 2 \sum_{m=-N}^{N} \phi(t-mT)\, s(t-mT) + \sum_{m=-N}^{N} \sum_{k=m-N}^{m+N} \phi(kT)\, s(t-mT)\, s(t-mT+kT) \right) dt \tag{175}$$

The order of summation can be interchanged by the formula

$$\sum_{m=-N}^{N} \sum_{k=m-N}^{m+N} = \sum_{k=-2N}^{0} \sum_{m=-N}^{k+N} + \sum_{k=1}^{2N} \sum_{m=k-N}^{N} \tag{176}$$

We also allow the limits of integration to become infinite by introducing a function, $g_N(t)$, in the integrand; $g_N(t)$ is 1 in the time interval $(-NT, NT)$ and is zero elsewhere.

Interchanging orders of summation and integration, we have

$$E_N = \frac{1}{2NT} \sum_{m=-N}^{N} \int_{-\infty}^{\infty} g_N(t)[-2\phi(t-mT)s(t-mT)]\, dt$$

$$+ \sum_{k=-2N}^{0} \sum_{m=-N}^{k+N} \int_{-\infty}^{\infty} \phi(kT)\, s(t-mT)\, s(t-mT+kT)\, g_N(t)\, dt$$

$$+ \sum_{k=1}^{2N} \sum_{m=k-N}^{N} \int_{-\infty}^{\infty} \phi(kT)\, s(t-mT)\, s(t-mT+kT)\, g_N(t)\, dt + C \tag{177}$$

We now make a change of variable, $u = t - mT$, and interchange the orders of summation and integration.

$$E_N = \frac{1}{2NT} \int_{-\infty}^{\infty} \left(-2\phi(u)\, s(u) \left[\sum_{m=-N}^{N} g_N(u+mT) \right] \right.$$

$$+ \left[\sum_{k=-2N}^{0} \phi(kT)\, s(u+kT) \right] s(u) \left[\sum_{m=-N}^{k+N} g_N(u+mT) \right]$$

$$\left. + \left[\sum_{k=1}^{2N} \phi(kT)\, s(u+kT) \right] s(u) \left[\sum_{m=k-N}^{N} g_N(u+mT) \right] \right) du + C \tag{178}$$

The expression $f(u) = \sum_{m=-N}^{N} g_N(u+mT)$ can be described as follows.

$$f(u) \begin{cases} = 2N & \text{for } |u| < T \\ = (2N-1) & \text{for } T < |u| < 2T \\ \vdots & \\ = 1 & \text{for } (2N-1)\, T < |u| < 2NT \\ = 0 & \text{elsewhere} \end{cases} \tag{179}$$

Let

$$f_k(u) \begin{cases} = \sum_{m=-N}^{k+N} g_N(u+mT) & \text{for } -2N \leqslant k \leqslant 0 \\ = \sum_{m=k-N}^{N} g_N(u+mT) & \text{for } 1 \leqslant k \leqslant 2N \\ = 0 & \text{for all other } k \end{cases} \tag{180}$$

Note that $f_o(u) = f(u)$, and $f_k(u-kT) = f_{-k}(u)$.

We can rewrite Eq. 178 as

$$E_N = \frac{1}{T}\int_{-\infty}^{\infty} \left(-2\phi(u)\ s(u)\ \frac{1}{2N} f(u) + s(u) \sum_{k=-2N}^{2N} \phi(kT)\ s(u+kT)\ \frac{1}{2N} f_k(u)\right) du + C \tag{181}$$

If we now substitute $s(u) + \alpha\beta(u)$ for $s(u)$ in Eq. 181 and assume that the function $s(u)$ minimizes E_N, then a necessary condition for E_N to be minimum is that

$$\left.\frac{\partial E_N}{\partial \alpha}\right|_{\alpha=0} = 0$$

In our case,

$$\begin{aligned}\left.\frac{\partial E_N}{\partial \alpha}\right|_{\alpha=0} &= \frac{1}{T}\int_{-\infty}^{\infty} \beta(u)\left[-2\phi(u)\ \frac{f(u)}{2N}\right] du \\ &+ \frac{1}{T}\int_{-\infty}^{\infty} \beta(u)\left[\sum_{k=-2N}^{2N} \phi(kT)\ s(u+kT)\ \frac{f_k(u)}{2N}\right] du \\ &+ \frac{1}{T}\int_{-\infty}^{\infty} s(u)\left[\sum_{k=-2N}^{2N} \phi(kT)\ \beta(u+kT)\ \frac{f_k(u)}{2N}\right] du\end{aligned} \tag{182}$$

Making the change of variable, $w = u + kT$, in the last integral, and noting that $\phi_x(kT) = \phi_x(-kT)$, and $f_k(w-kT) = f_{-k}(w)$, we have

$$\left.\frac{\partial E_N}{\partial \alpha}\right|_{\alpha=0} = \frac{1}{T}\int_{-\infty}^{\infty} \beta(u)\left[-2\phi(u)\ \frac{f(u)}{2N} + 2 \sum_{k=-2N}^{2N} \phi(kT)\ s(u+kT)\ \frac{f_k(u)}{2N}\right] du \tag{183}$$

Since $\beta(u)$ is arbitrary, our necessary condition for E_N to be minimum becomes

$$\phi(u)\ \frac{f(u)}{2N} = \sum_{k=-2N}^{2N} \phi(kT)\ s(u+kT)\ \frac{f_k(u)}{2N} \tag{184}$$

Let us now consider some properties of the solution of Eq. 184. Since $f(u)$, $f_k(u)$, and $f_{-k}(u)$ are all zero for $|u| > 2NT$, $s(u)$ is arbitrary for $|u| > 2NT$. For simplicity, we shall let $s(u)$ be zero for $|u| > 2NT$.

For the special case in which $\phi(kT) = 0$ for $k = \pm 1, \pm 2, \ldots, \pm 2N$, Eq. 184 becomes

$$\phi(u)\left[\frac{f(u)}{2N}\right] = \phi(0)\ s(u)\left[\frac{f(u)}{2N}\right] \tag{185}$$

The solution is

$$\begin{aligned} s(u) &= \frac{\phi(u)}{\phi(0)} \qquad \text{for } |u| < 2NT \\ s(u) &= 0 \qquad\quad \text{for } |u| \geqslant 2NT \end{aligned} \tag{186}$$

Another example is given by the following case:

$$\phi(\tau) = 2 - \frac{|\tau|}{T} \qquad \text{for } 0 \leq |\tau| \leq 2T$$
$$\phi(\tau) = 0 \qquad \text{elsewhere} \tag{187}$$

In this case, Eq. 184 may be written as follows for N = 1:

$$\left(2-\frac{u}{T}\right) 1 = 1s(u-T)\ 1 + 2s(u)\ 1 + 1s(u+T)\frac{1}{2} \tag{188}$$

with $0 \leq u \leq T$, and may be written as

$$\left(2-\frac{w}{T}\right)\frac{1}{2} = 1s(w-T)\frac{1}{2} + 2s(w)\frac{1}{2} \tag{189}$$

for $T < w \leq 2T$.

The form of Eq. 184 indicates that we should look for a symmetric solution; that is, we assume that $s(u-T) = s(T-u)$. Making the change of variable, $w = u + T$, gives us the following system of equations which is valid for $0 < u < T$:

$$2 - \frac{u}{T} = s(T-u) + 2s(u) + \frac{1}{2}s(u+T) \tag{190}$$

$$\frac{1}{2}\left(1-\frac{u}{T}\right) = \frac{1}{2}s(u) + s(u+T) \tag{191}$$

A solution to these equations is

$$s(u)\begin{cases} = 1 - \dfrac{|u|}{T} & \text{for } 0 \leq |u| \leq T \\ = 0 & \text{elsewhere} \end{cases} \tag{192}$$

Let us now substitute Eq. 184 in Eq. 181 to obtain an expression for the minimum value of E_N. We shall call this minimum value E_m. It is given by the formula

$$E_m = C - \frac{1}{T}\int_{-\infty}^{\infty} \phi(u)\ s(u)\ \frac{f(u)}{2N}\, du \tag{193}$$

For the special case that has just been discussed, C is 2. From the definition of f(u) we recall that f(u)/2N is 1 in the interval (−T, T), and s(u) and $\phi(u)$ have been defined. Thus Eq. 193 becomes

$$E_m = 2 - \frac{1}{T}\int_{-T}^{T}\left(2-\frac{|\tau|}{T}\right)\left(1-\frac{|\tau|}{T}\right)d\tau$$
$$= 2 - \frac{2}{T}\int_{0}^{T}\left(2-\frac{|\tau|}{T}\right)\left(1-\frac{|\tau|}{T}\right)d\tau = 2 - \frac{2}{T}\left[2 - \frac{3\tau^2}{2T} + \frac{\tau^3}{3T^2}\right]_0^T$$
$$= \frac{1}{3} \tag{194}$$

We might interpret this in terms of a ratio of average signal power to average "noise" (or error) signal power into a unit resistor load. For our special case,

$$\frac{S}{N} = \frac{2}{1/3} = \frac{6}{1} \tag{195}$$

4.3 The Case of an Infinite Number of Samples

We have been discussing the reconstruction, over a finite interval, of a member of a stationary ensemble from a finite number of uniformly spaced samples. We have constrained the problem by requiring that each sample be treated the same; that is, the same interpolatory function is used for each sample value.

We now turn to the case of reconstruction from an infinite number of samples. As before, we do not restrict x(t) to be bandlimited. In this situation,

$$E_\infty = \frac{1}{T}\int_0^T E\left[(x(t)-f(t))^2\right] dt$$

$$= \frac{1}{T}\int_0^T \left(C - 2\sum_{n=-\infty}^{\infty} \phi(t-nT)\, s(t-nT) + \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} \phi(nT-mT)\, s(t-mT)\, s(t-nT)\right) dt \tag{196}$$

Letting $u = t - nT$, we have

$$E_\infty = C - \frac{1}{T}\sum_{n=-\infty}^{\infty}\int_{-nT}^{(-n+1)T} \left(2\phi(u)\, s(u) - s(u)\sum_{m=-\infty}^{\infty} \phi(nT-mT)\, s(u+nT-mT)\right) du \tag{197}$$

Making the change of summation index, $k = n - m$, and then carrying out the summation on n, yields

$$E_\infty = C - \frac{1}{T}\int_{-\infty}^{\infty} \left(2\phi(u)\, s(u) - s(u)\sum_{k=-\infty}^{\infty} \phi(kT)\, s(u+kT)\right) du \tag{198}$$

A necessary condition for E_∞ to be minimum (by varying s) is

$$\phi(u) = \sum_{k=-\infty}^{\infty} \phi(kT)\, s(u+kT) \tag{199}$$

If $\Phi(f)$ is the Fourier transform of $\phi(u)$, then Eq. 199 can be written in the frequency domain as

$$S(f) = \frac{\Phi(f)}{\sum_{k=-\infty}^{\infty} \phi(kT)\exp(+j2k\pi Tf)} = \frac{\Phi(f)}{\sum_{k=-\infty}^{\infty} \phi(kT)\cos 2\pi kTf} \tag{200}$$

where S(f) is the Fourier transform of s(t), provided that the series in the denominator has no zeros, where $\Phi(f)$ is nonzero.

Analysis given by Bennett (19) shows that we can rewrite Eq. 200 as

$$S(f) = \frac{\Phi(f)}{(1/T) \sum_{n=-\infty}^{\infty} \Phi(f-nf_r)} \tag{201}$$

where $f_r = 1/T$ (ref. 19).

We can obtain an expression for the minimum value of E_∞, which we shall call E'_m, by substituting Eq. 199 in Eq. 198. Thus

$$E'_m = C - \frac{1}{T}\int_{-\infty}^{\infty} \phi(u)\, s(u)\, du \tag{202}$$

Use of Eq. 201 and Parseval's theorem gives

$$E'_m = C - \int_{-\infty}^{\infty} \frac{\Phi^2(f)}{\sum_{n=-\infty}^{\infty} \Phi(f-(n/T))}\, df \tag{203}$$

We recall that $C = E[x^2(t)]$.

For the special case in which $\Phi(f)$ is nonzero only within the band $(-B, B)$ and $1/T \geqslant 2B$, we have, from Eq. 201,

$$S(f) = \begin{cases} T & \text{for } |f| < B \\ 0 & \text{for } \left|\frac{n}{T}-f\right| < B, \quad n = \pm 1, \pm 2, \dots \\ \text{otherwise undefined} \end{cases} \tag{204}$$

One possible solution is

$$S(f) = \begin{cases} T & \text{for } |f| < B \\ 0 & \text{otherwise} \end{cases} \tag{205}$$

Thus,

$$s(t) = \frac{2BT \sin 2\pi Bt}{2\pi Bt} \tag{206}$$

in agreement with the sampling theorem (8, 9). We note that, if $1/T = 2B$, then $S(f)$ is defined for all frequencies by Eq. 204.

As another example, let us assume that $\phi(\tau)$ is given by Eq. 187. Equation 199 then reduces to

$$\phi(u) = s(u-T) + 2s(u) + s(u+T) \tag{207}$$

A solution of Eq. 207 is

$$s(u) = \begin{cases} 1 - \dfrac{|u|}{T} & |u| < T \\ 0 & \text{otherwise} \end{cases} \tag{208}$$

in agreement with Eq. 192. Thus, in this case, the optimum interpolatory pulse is the same for an infinite number of samples as it is for three samples. This is not true if we are given an arbitrary correlation function, as shown by Eqs. 185 and 186. In the latter case the function s(u) changes with N, in general.

Acknowledgment

The present form of this report owes much to the interest and constructive criticism of Professor Y. W. Lee. His emphasis on the importance of good technical speech and writing has been especially helpful. Discussions with Professor C. E. Shannon led to improvements in the technical results. The writer was introduced to this research area by Dr. W. R. Bennett of Bell Telephone Laboratories.

References

1. W. R. Bennett and C. B. Feldman, Band width and transmission performance, Bell System Tech. J. 28, 490-595 (July 1949).
2. W. M. Goodall, Television by pulse code modulation, Bell System Tech. J. 28, 33-49 (Jan. 1951).
3. W. R. Bennett, Synthesis of active networks, Proc. Symposium on Modern Network Synthesis, Polytechnical Institute of Brooklyn, April 1955, p. 47.
4. B. M. Oliver, J. R. Pierce, and C. E. Shannon, The philosophy of PCM, Proc. IRE 36, 1324-1331 (Nov. 1948).
5. W. R. Bennett, Statistics of regenerative digital transmission, Bell System Tech. J. 37, 1504-1518 (Nov. 1958).
6. S. P. Lloyd and B. McMillan, Linear least squares filtering and prediction of sampled signals, Proc. Symposium on Modern Network Synthesis, Polytechnical Institute of Brooklyn, April 1955, p. 225.
7. Ibid., p. 227.
8. C. E. Shannon, Communication in the presence of noise, Proc. IRE 37, 11-12 (Jan. 1949).
9. A. V. Balakrishnan, A note on the sampling principle for continuous signals, Trans. IRE IT-3, 143-146 (June 1957).
10. C. E. Shannon, Communication in the presence of noise, Proc. IRE 37, 10-11 (Jan. 1949).
11. B. M. Oliver, J. R. Pierce, and C. E. Shannon, The philosophy of PCM, Proc. IRE 36, 1324 (Nov. 1948).
12. W. M. Goodall, Television by pulse code modulation, Bell System Tech. J. 28, 37 (Jan. 1951).
13. H. S. Black, Modulation Theory (D. Van Nostrand Co., Inc., New York, 1953), pp. 33, 303.
14. F. B. Hildebrand, Methods of Applied Mathematics (Prentice-Hall, Englewood Cliffs, N. J., 1952), pp. 120-143.
15. H. Cramér, Mathematical Methods of Statistics (Princeton University Press, Princeton, N. J., 1946), p. 180.
16. F. B. Hildebrand, op. cit., p. 143.
17. F. B. Hildebrand, op. cit., p. 127.
18. H. Cramér, op. cit., p. 182.
19. W. R. Bennett, Statistics of regenerative digital transmission, Bell System Tech. J. 37, 1507-1508 (Nov. 1958).
20. D. W. Tufts, Some optimization problems of digital pulse transmission, S.M. Thesis, Department of Electrical Engineering, M. I. T., February 1958, p. 18.
21. H. Nyquist, Certain topics in telegraph transmission theory, Trans. AIEE 47, 617-644, Appendix II-A (April 1928).
22. F. B. Hildebrand, op. cit., p. 12.
23. F. B. Hildebrand, op. cit., p. 51.
24. M. Schetzen, Fourier transforms of positive functions, Quarterly Progress Report No. 52, Research Laboratory of Electronics, M. I. T., January 15, 1959, p. 85.