On the Sensitivity of Channel Capacity for the Gaussian Bandlimited Channel

By I. W. SANDBERG

(Manuscript received June 24, 1966)

It is a classic result of Shannon that binary digits can be communicated with arbitrarily small error probability at any rate less than

$$W \log_2 \left(1 + \frac{P}{N}\right) \quad (bits/sec)$$

over a channel with bandwidth W and additive Gaussian noise of average power N, using signals of average power at most P. However, in Shannon's proof it is assumed that the input to the receiver is the sum of a linear combination of the bandlimited functions

$$\varphi_0(t - k/2W) \triangleq \frac{\sin 2\pi W(t - k/2W)}{2\pi W(t - k/2W)}, \qquad -\infty < t < \infty, \quad k = 1,2, \cdots$$

(which are of course of doubly infinite duration) and a sample function from an exactly bandlimited Gaussian random process. The fact that $\varphi_0(k/2W) = 0$ for all integers $k \neq 0$ plays a key role in that it implies the total absence of intersymbol interference.

As a result of these assumptions, there have been some objections to the Shannon model in connection with the notion of rate, the fact that the received signals are entire functions (which are predictable for all time from a knowledge of their values on any interval of nonzero length) and the fact that it is not clear whether the performance of the model is critically dependent on the assumptions that lead to the absence of intersymbol interference.

Since Shannon's model and his associated ingenious arguments are widely known and are of great interest, from the point of view of the system theorist, it is important to be able to prove an "insensitivity theorem" to the effect that if the model is modified to the extent that: (i) $\varphi_0(t)$ is replaced by an approximating function $\varphi(t)$ with the property that the signals are of average power at most $\tilde{P}$ where $\tilde{P}$ is approximately P, and $\varphi(t) = 0$ for $t < t_\varphi$ for some negative number t_φ, and (ii) the noise is approximately

bandlimited with bandwidth W, then, subject to some reasonable qualifications, it is possible to transmit information, with arbitrarily high reliability, at any rate less than

$$W\ log_2\left(1 + \frac{P}{N}\right).$$

We prove such a theorem in this paper. In fact, we show that if the noise has integrable power spectral density $S(\omega)$ for which

$$0 < \inf_{0 \leqq \omega < 2\pi W} \sum_{p=-\infty}^{\infty} S(\omega + 4\pi W p)$$

and

$$\tilde{N} \triangleq 2W \sup_{0 \leqq \omega < 2\pi W} \sum_{p=-\infty}^{\infty} S(\omega + 4\pi W p) < \infty$$

(these are very weak assumptions), then any rate

$$R < W\ log_2\left(1 + \frac{\gamma P}{\tilde{N}}\right)$$

is permissible if $\gamma \varepsilon (0,1)$ such that [with the understanding that $\varphi(0) = 1$]

$$\sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} \mid \varphi(k/2W) \mid < (1 - \gamma^{\frac{1}{2}}) \left(\frac{\tilde{N}\beta}{2WP\gamma}\right)^{\frac{1}{2}}$$

where β is an important positive number that depends on R, $(\tilde{N}/\gamma)$, P, and W.

Observe that if $S(\omega)$ is the ideal spectral density defined by

$$S(\omega) = \frac{N}{2W}, \qquad |\omega| \leqq 2\pi W$$
$$= 0, \qquad |\omega| > 2\pi W$$

then $\tilde{N} = N$.

I. INTRODUCTION

It is a classic result[1] of Shannon that binary digits can be communicated with arbitrarily small error probability at any rate less than

$$W \log_2\left(1 + \frac{P}{N}\right) \quad \text{(bits/sec)} \tag{1}$$

over a channel with bandwidth W and additive Gaussian noise of average power N, using signals of average power at most P. There are, however, some unrealistic assumptions in Shannon's argument. In particular, there have been some objections[2,3,4] to the Shannon model in connection with, for example, the notion of rate and the fact that the received signals are entire functions (which are predictable for all time from a knowledge of their values on any interval of nonzero length).

The purpose of this paper is to focus attention on Shannon's assumptions[1] and show that they can be modified so that the end result is a quite detailed and informative statement concerned with a much more realistic model* of a communication system.

II. REVIEW OF SHANNON'S ARGUMENT

2.1 *The Capacity of the Time-Discrete Gaussian Channel*

Shannon's result for the bandlimited time-continuous channel follows directly from a result concerned with the following type of time-discrete channel.

The channel receives one of M equally likely inputs (i.e., code words) every T seconds. Each input is a real n-vector $X \triangleq (x_1, x_2, \cdots, x_n)$ which satisfies

$$| X |^2 \leqq \rho T$$

where $| X |$ denotes the Euclidean norm of X and ρ is a positive constant independent of X. It is assumed that there exists a positive constant μ, independent of T, such that $n = 2\mu T$ (with the understanding that we consider only values of T for which $2\mu T$ is an integer).

The channel output (i.e., the receiver input) corresponding to the input X is the n-vector $X + Z$, in which the components of the "noise vector" Z are independent Gaussian random variables with mean zero and variance η. In its attempt to determine which of the M known code words was transmitted, the receiver may make an error, and we shall denote by p_{ei} the probability that an error is made given that code word i is transmitted.

It is assumed that the channel is used to transmit information in the following manner. Let a message source produce independent and equally likely binary digits at the rate R digits per second. Every T seconds,† one of 2^{RT} possible sequences is produced. We set $M = 2^{RT}$ and we represent each of the binary sequences by a particular code word.

* Some different results concerning the significance of the Shannon bound (1) are proved in Ref. 4. In particular, there, for certain models, converse propositions are established.

† We consider only values of T for which RT is an integer.

We say that a rate R is permissible if for each $\epsilon > 0$ there exists a T and a corresponding code such that

$$\max_i p_{ei} \leqq \epsilon.$$

It has been proven that the channel capacity C, the least upper bound of permissible rates, is given by

$$C = \mu \log_2 \left(1 + \frac{\rho}{2\mu\eta}\right) \quad \text{(bits/sec)}.$$

It has also been proven that for $R < C$ there exists a positive number $\beta = \beta(\eta,\rho,\mu,R)$ such that for each $T > 0$ there exists a code with the property that

$$\max_i p_{ei} = \exp\left[-\beta T + o(T)\right].$$

2.2 *The Time-Continuous Bandlimited Channel*

In order to use the ideas and results outlined above in his study of the time-continuous bandlimited channel, Shannon considers the model shown in Fig. 1, with the understanding that **H** represents an ideal low-pass filter with cut-off frequency W, and $z(\cdot)$ denotes a sample function of a bandlimited Gaussian random process with mean zero and power spectral density

$$\begin{aligned} S(\omega) &= \frac{N}{2W}, \qquad |\omega| \leqq 2\pi W \\ &= 0, \qquad |\omega| > 2\pi W, \end{aligned}$$

where N is a positive constant. Clearly the average power of $z(\cdot)$ is N.

As in the time-discrete case, the message source produces R binary digits per second, so that every T seconds one of $M = 2^{RT}$ possible sequences is produced. Consider the ith such sequence. The coder and signal generator associates with this sequence a particular n-vector

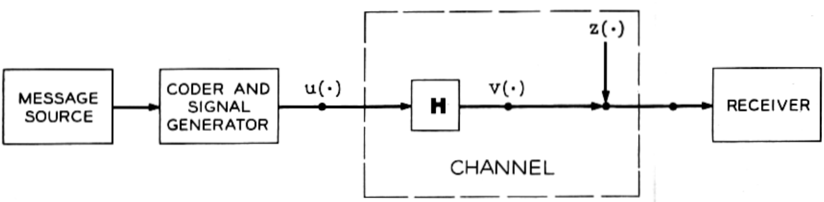

Fig. 1 — Model of a Communication System.

$X \triangleq (x_1, x_2, \cdots, x_n)$, where $n = 2WT$, and a corresponding signal

$$u(t) = \sum_{k=1}^{n} x_k \frac{\sin 2\pi W(t - k/2W)}{2\pi W(t - k/2W)}, \qquad t \, \varepsilon \, (-\infty, \infty)$$

which is transmitted. This process is repeated every T seconds. It is assumed that

$$| X |^2 \leqq 2WPT$$

for each code word, so that, for each signal, as can readily be verified,

$$\frac{1}{T} \int_{-\infty}^{\infty} u(t)^2 dt \leqq P. \tag{2}$$

Insofar as a physical interpretation of (2) is concerned, the object on the left is the total energy of $u(\cdot)$ divided by the length of the interval $[(4W)^{-1}, (4W)^{-1} + T]$ which, considering only the instants $t = k/2W$, contains all of the samples of $u(\cdot)$ that can be made nonzero. If (2) holds, then Shannon says that $u(\cdot)$ has average power at most P.

The received signal due to the noise and only the ith sequence is $u(\cdot) + z(\cdot)$, since the response of $\mathbf{H}$ to $u(\cdot)$ is $u(\cdot)$. The value of this signal at the instant $t = k/2W$ is

$$x_k + z(k/2W) \quad \text{for} \quad k = 1, 2, \cdots, n$$

in which the $z(k/2W)$ are independent* Gaussian random variables with mean zero and variance N. These sample values are the same as those that would have been obtained if we had not ignored the effect at the receiver of transmitted signals due to previous and subsequent sequences, since the values of such signals at $t = k/2W$ vanish for $k = 1, 2, \cdots, n$.

Thus, on the basis of the channel capacity result of the previous section, we see that our continuous channel can process information, with arbitrarily high reliability, at any rate less than the capacity of the time-discrete channel with parameters $\mu = W$, $\rho = 2WP$, and $\eta = N$, that is, at any rate R less than

$$W \log_2 \left(1 + \frac{P}{N}\right).$$

2.3 *Discussion*

The argument of the last section is based on the assumptions that the input to the receiver is the sum of a linear combination of the band-

* The autocorrelation function of the noise vanishes for $\tau = k/2W$, $k \neq 0$.

limited functions

$$\varphi_0(t - k/2W) \triangleq \frac{\sin 2\pi W(t - k/2W)}{2\pi W(t - k/2W)}, \qquad \begin{matrix} -\infty < t < \infty \\ k = 1,2,\ldots \end{matrix}$$

(which are of course of *doubly infinite* duration) and a sample function from an *exactly* bandlimited Gaussian random process. The fact that $\varphi_0(k/2W) = 0$ for all integers $k \neq 0$ plays a key role in that it implies the *total* absence of intersymbol interference.

As a result of these assumptions, there have been some objections to the Shannon model in connection with the notion of rate,* the fact that the received signals are entire functions (which are predictable for all time from a knowledge of their values on any interval on nonzero length), and the fact that it is not clear whether or not the performance of the model is critically dependent on the assumptions that lead to the absence of intersymbol interference.

Since Shannon's model and his associated ingenious arguments are widely known and are of great interest, from the point of view of the system theorist, it is important to be able to prove an "insensitivity theorem" to the effect that if the model is modified to the extent that: (*i*) $\varphi_0(t)$ is replaced by an approximating function $\varphi(t)$ with the property that the signals are of average power at most $\tilde{P}$ where $\tilde{P}$ is approximately P, and $\varphi(t) = 0$ for $t < t_\varphi$ for some negative number t_φ, and (*ii*) the noise is approximately bandlimited with bandwidth W, then, subject to some reasonable qualifications, it is possible to transmit information, with arbitrarily high reliability, at any rate less than

$$W \log_2 \left(1 + \frac{P}{N}\right).$$

A quite explicit theorem of this type is stated in the next section.

III. THE MORE REALISTIC MODEL

We now consider the system of Fig. 1 to be an approximation to the Shannon model described in Section 2.2.

Here we assume that $z(\cdot)$ is a sample function from a Gaussian random process with zero mean and integrable power spectral density $S(\omega)$ with the property that

$$\sup_{0 \leqq \omega < 2\pi W} \sum_{p=-\infty}^{\infty} S(\omega + 4\pi W p)$$

* Shannon himself has indicated[5] that care must be taken in the physical interpretation of the result of Section 2.2. However, he does not discuss the effect of intersymbol interference or the effect of the departure of the noise spectrum from the ideal spectrum.

is finite. From the engineering viewpoint, this finiteness condition is a very weak assumption; it is certainly satisfied if there exists a constant $K > 0$ such that $S(\omega) \leqq K(1 + \omega^2)^{-1}$ for all real ω.

We again suppose that the message source produces one of $M = 2^{RT}$ equally likely binary sequences every T seconds. We assume that there is a first such sequence and that the coder assigns the code word $(x_1, x_2, \cdots, x_n)$ to it. After T seconds, the second sequence is assigned the code word $(x_{n+1}, x_{n+2}, \cdots, x_{2n})$, and so on. The integer n is equal to $2WT$.

The transmitted signal (i.e., the input to the channel) is assumed to be given by

$$u(t) = \sum_{k=1}^{n} x_k \psi(t - k/2W) + \sum_{k=n+1}^{2n} x_k \varphi(t - k/2W) + \ldots$$

in which $\psi(\cdot)$ is a real-valued function of t defined on $(-\infty, \infty)$ such that there exists a negative constant t_ψ with the property that $\psi(t) = 0$ for $t < t_\psi$. It is evident that each of the signal components (i.e., each sum) is associated with a particular code word, that is, with a particular input sequence to the coder. We note that the first signal component "begins" at $t = t_\psi + (2W)^{-1}$, the second at $t_\psi + (2W)^{-1} + T$, and so on.

The operator $\mathbf{H}$ in Fig. 1 is assumed here to be causal, linear, and time-invariant. Thus, the output of $\mathbf{H}$ is

$$v(t) = \sum_{k=1}^{n} x_k \varphi(t - k/2W) + \sum_{k=n+1}^{2n} x_k \varphi(t - k/2W) + \ldots$$

in which $\varphi(\cdot)$ is the response of $\mathbf{H}$ to $\psi(\cdot)$. Since $\mathbf{H}$ is causal, there exists a negative constant t_φ such that $\varphi(t) = 0$ for $t < t_\varphi$.

We assume that $\varphi(0) = 1$ and that $\varphi(\cdot)$ belongs to L_2 (i.e., is square integrable). We think of $\varphi(t)$ as being close to

$$\varphi_0(t) \triangleq \frac{\sin 2\pi W t}{2\pi W t}$$

in the sense that both $\| \varphi - \varphi_0 \|$ ($\| \cdot \|$ denotes the L_2 norm) and

$$\sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} | \varphi(k/2W) - \varphi_0(k/2W) | = \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} | \varphi(k/2W) |$$

are small. Of course this requires that $-t_\varphi$ be sufficiently large.*

* We may certainly take the view that $\psi(\cdot)$ and $\mathbf{H}$ are approximations to the ideal signal φ_0 and the ideal bandlimiting filter, respectively. However, the specific nature of these approximations is not pertinent to our development. Observe, in fact, that it makes sense for us to assume here that $\mathbf{H}$ is an approximation to the ideal bandlimiting filter, but that $\psi(\cdot)$ is an impulse-like function. The response $\varphi(\cdot)$ of $\mathbf{H}$ to $\psi(\cdot)$ is what we wish to focus attention on.

It is assumed also that

$$\sum_{k=1}^{n} (x_{k+jn})^2 \leqq 2WPT$$

for $j = 0, 1, 2, \cdots$, so that the "average power"

$$\frac{1}{T} \int_{-\infty}^{\infty} \left| \sum_{k=1}^{n} x_{k+jn} \varphi[t - (k + jn)/2W] \right|^2 dt$$

of the jth component of $v(\cdot)$ is bounded from above by $P + \zeta_j$, in which $\zeta_j \to 0$ as $\| \varphi - \varphi_0 \| \to 0$.

The receiver, which is assumed to be in possession of the code, samples the signal $v(\cdot) + z(\cdot)$ at the instants $t = k/2W$, $k = 1, 2, \cdots$, to obtain in succession the "received n-vectors"

$$Y_1 \triangleq (v_1, v_2, \cdots, v_n) + (z_1, z_2, \cdots, z_n)$$

$$Y_2 \triangleq (v_{n+1}, v_{n+2}, \cdots, v_{2n}) + (z_{n+1}, z_{n+2}, \cdots, z_{2n})$$

$$\vdots$$

in which $v_k = v(k/2W)$ and $z = z(k/2W)$. These vectors are used as inputs to a minimum distance decoder. Thus, for example, if

$$| Y_1 - X_i | < \min_{j \neq i} | Y_1 - X_j |,$$

in which $\{X_j\}$ denotes the set of code words, then Y_1 is decoded as X_i. We denote by p_{eij} the maximum probability, over all possible sequences of input code words with the jth code word X_i, that Y_j is not decoded as X_i. We let

$$p_{ei} \triangleq \sup_j p_{eij} .$$

Our result (which is proved in the next section) is

Theorem: Concerning the system described above, let

$$0 < \inf_{0 \leqq \omega < 2\pi W} \sum_{p=-\infty}^{\infty} S(\omega + 4\pi Wp)$$

and

$$\tilde{N} \triangleq 2W \sup_{0 \leqq \omega < 2\pi W} \sum_{p=-\infty}^{\infty} S(\omega + 4\pi Wp).$$

Then any rate

$$R < W \log_2\left(1 + \frac{\gamma P}{\tilde{N}}\right) \quad (bits/sec)$$

is permissible (in the sense of Section 2.1 with p_{ei} as defined above) provided that $\gamma \varepsilon$ (0,1) such that

$$\sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} |\varphi(k/2W)| < (1 - \gamma^{\frac{1}{2}}) \left(\frac{\tilde{N}\beta}{2WP\gamma}\right)^{\frac{1}{2}}$$

where $\beta = \beta[(\tilde{N}/\gamma), 2WP, W, R]$ is the number introduced in Section 2.1.

Remarks: Observe that if $S(\omega)$ is the ideal power spectral density defined by

$$S(\omega) = \frac{N}{2W}, \qquad |\omega| \leqq 2\pi W$$
$$= 0, \qquad |\omega| > 2\pi W$$

then $\tilde{N} = N$. The condition that

$$0 < \inf_{0 \leqq \omega < 2\pi W} \sum_{p=-\infty}^{\infty} S(\omega + 4\pi Wp)$$

is certainly satisfied if $S(\omega)$ is a reasonable approximation to the ideal spectrum.

If $S(\omega)$ is nonincreasing for $\omega \geqq 0$, then for $p = 1, 2, \cdots,$

$$\sup_{0 \leqq \omega < 2\pi W} S(\omega + 4\pi Wp) \leqq \frac{1}{2\pi W} \int_{4\pi Wp - 2\pi W}^{4\pi Wp} S(\omega) d\omega$$

and

$$\sup_{0 \leqq \omega < 2\pi W} S(\omega - 4\pi Wp) \leqq \frac{1}{2\pi W} \int_{-4\pi Wp + 2\pi W}^{-4\pi W(p-1)} S(\omega) d\omega.$$

Thus, for $S(\omega)$ nonincreasing for $\omega \geqq 0$, we have the bound

$$\tilde{N} \leqq 2W \sup_{0 \leqq \omega < 2\pi W} \sum_{p=-1}^{1} S(\omega + 4\pi Wp) + \frac{1}{\pi} \int_{4\pi W}^{\infty} S(\omega) d\omega.$$

The exponent β has been estimated by Shannon.[5]

The basic idea of the proof of the theorem is, roughly speaking, to (*i*) treat as an additional "noise source" the *departure* of the samples of $v(\cdot)$ from the corresponding samples in the case of zero intersymbol-interference (Sublemma 1 of Section IV provides an estimate of this departure), and (*ii*) to obtain a lower bound on the channel capacity of the more-realistic model by comparing its error probability performance with that of a model possessing zero intersymbol-interference and independent Gaussian noise samples (this is done in the proof of Sublemma 2 of Section IV).

IV. PROOF OF THE THEOREM

4.1 *The Discrete Channel*

Consider first a discrete channel with memory that receives one of M equally likely inputs (i.e., code words) every T seconds. As in Section 2.1, each input is a real n-vector X which satisfies $|X|^2 \leqq \rho T$, n is equal to $2\mu T$, and each input represents a particular sequence of RT binary digits. Let $(x_1, x_2, \cdots, x_n)$ denote the first code word, $(x_{n+1}, x_{n+2}, \cdots, x_{2n})$ the second code word, and so on.

At time $t = (j-1)T$, the receiver receives the n-vector

$$Y_j \triangleq \{y[1 + (j-1)n], y[2 + (j-1)n], \cdots, y[jn]\}$$

in which

$$y(p) = \sum_{k=1}^{\infty} x_k \varphi(p-k) + z(p), \qquad p = 1,2, \cdots$$

where here $\varphi(\cdot)$ is a function defined on the integers so that $\varphi(0) = 1$ and

$$\sum_{k=-\infty}^{\infty} |\varphi(k)| < \infty,$$

and each $z(p)$ is a Gaussian random variable with zero mean. For each j, let

$$Z_j \triangleq \{z[1 + (j-1)n], z[2 + (j-1)n], \cdots, z[jn]\}$$

and

$$V_j \triangleq \{v_j[1 + (j-1)n], v[2 + (j-1)n], \cdots, v[jn]\}.$$

where

$$v(p) = \sum_{k=1}^{\infty} x_k \varphi(p-k).$$

Then $Y_j = V_j + Z_j$.

We assume that the receiver attempts to determine the jth code word V_j by minimum distance decoding as in Section III. Let p_{ei} denote the error probability associated with the transmission of code word i, as defined in Section III. In Section 4.3 we prove the following result, which we shall exploit here, concerning this channel.

Lemma: Let Z_j, as defined above, possess the property that [with $\mathcal{E}$ the expectation operator and $(\cdot,\cdot)$ denoting the usual inner product of n-vectors] there exist constants ϵ and η such that for every real n-vector U of unit length:

$$0 < \epsilon \leqq \mathcal{E}\,|(U,Z_j)|^2 \leqq \eta$$

uniformly in j and n. Let $\gamma \,\varepsilon\, (0,1)$. Then any rate

$$R < \mu \log_2 \left(1 + \frac{\gamma\rho}{2\mu\eta}\right)$$

is permissible (in the sense of Section 2.1) provided that

$$\sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} |\varphi(k)| < (1 - \gamma^{\frac{1}{2}}) \left(\frac{\eta\beta}{\rho\gamma}\right)^{\frac{1}{2}}$$

where $\beta = \beta[(\eta/\gamma), \rho, \mu, R]$ is the number introduced in Section 2.1.

4.2 *Completion of the Proof of the Theorem*

$$\mathcal{E}\,|(U,Z_j)|^2 = \mathcal{E} \sum_{k,l} u_k u_l z_{[k+(j-1)n]} z_{[l+(j-1)n]}$$
$$= \sum_{k,l} u_k u_l R[(l-k)/2W]$$

for any real n-vector U, in which

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) e^{i\omega\tau} d\tau.$$

Thus,

$$\mathcal{E}\,|(U,Z_j)|^2 = \frac{1}{2\pi} \sum_{k,l} u_k u_l \int_{-\infty}^{\infty} S(\omega) e^{i\omega(l-k)/2W} d\omega$$
$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \sum_{k=1}^{n} u_k e^{-i\omega k/2W} \right|^2 S(\omega)\, d\omega$$
$$= \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \int_{-2\pi W+4\pi Wp}^{2\pi W+4\pi Wp} \left| \sum_{k=1}^{n} u_k e^{-i\omega k/2W} \right|^2 S(\omega)\, d\omega$$
$$= \frac{1}{2\pi} \int_{-2\pi W}^{2\pi W} \left| \sum_{k=1}^{n} u_k e^{-i\omega k/2W} \right|^2 \sum_{p=-\infty}^{\infty} S(\omega + 4\pi Wp)\, d\omega.$$

It follows at once that

$$\mathcal{E} \mid (U,Z_j) \mid^2 \leqq \sup_{0 \leqq \omega < 2\pi W} \sum_{p=-\infty}^{\infty} S(\omega + 4\pi W p) \frac{1}{2\pi} \int_{-2\pi W}^{2\pi W} \left| \sum_{k=1}^{n} u_k e^{-i\omega k/2W} \right|^2 d\omega,$$

and that

$$\mathcal{E} \mid (U,Z_j) \mid^2 \geqq \inf_{0 \leqq \omega < 2\pi W} \sum_{p=-\infty}^{\infty} S(\omega + 4\pi W p) \frac{1}{2\pi} \int_{-2\pi W}^{2\pi W} \left| \sum_{k=1}^{n} u_k e^{-i\omega k/2W} \right|^2 d\omega.$$

Since

$$\frac{1}{4\pi W} \int_{-2\pi W}^{2\pi W} \left| \sum_{k=1}^{n} u_k e^{-i\omega k/2W} \right|^2 d\omega = \mid U \mid^2,$$

we have

$$\mathcal{E} \mid (U,Z_j) \mid^2 \leqq 2W \sup_{0 \leqq \omega < 2\pi W} \sum_{p=-\infty}^{\infty} S(\omega + 4\pi W p)$$

$$\mathcal{E} \mid (U,Z_j) \mid^2 \geqq 2W \inf_{0 \leqq \omega < 2\pi W} \sum_{p=-\infty}^{\infty} S(\omega + 4\pi W p)$$

for $\mid U \mid = 1$, independent of j and n. Thus, we may view the time continuous system of Section III as a discrete-time communication system of the type described at the outset of this section with $\mu = W$, $\rho = 2WP$,

$$\epsilon = 2W \inf_{0 \leqq \omega < 2\pi W} \sum_{p=-\infty}^{\infty} S(\omega + 4\pi W p),$$

and

$$\eta = 2W \sup_{0 \leqq \omega < 2\pi W} \sum_{p=-\infty}^{\infty} S(\omega + 4\pi W p).$$

This proves the theorem.

4.3 *Proof of the Lemma*

With x_k as defined in Section 4.1, let

$$\tilde{V}_j \triangleq \{x_{[1+(j-1)n]}, x_{[2+(j-1)n]}, \cdots, x_{[jn]}\}.$$

Sublemma 1:

$$\mid V_j - \tilde{V}_j \mid^2 \leqq 2\rho T \left(\sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} \mid \varphi(k) \mid \right)^2$$

Proof:

$$| V_j - \tilde{V}_j |^2 = \sum_{p=1+(j-1)n}^{jn} \left| \sum_{k=1}^{\infty} x_k \varphi(p-k) - x_p \right|^2$$

$$= \sum_{p=1+(j-1)n}^{jn} \left| \sum_{k=-\infty}^{\infty} x_k \tilde{\varphi}(p-k) \right|^2$$

in which $x_k = 0$ for $k < 1$, $\tilde{\varphi}(0) = 0$, and $\tilde{\varphi}(k) = \varphi(k)$ for $k \neq 0$. Therefore,

$$| V_j - \tilde{V}_j |^2 = \sum_p \left| \sum_k x_{(p-k)} \tilde{\varphi}(k) \right|^2,$$

and, by the Schwarz inequality,

$$| V_j - \tilde{V}_j |^2 \leqq \sum_p \sum_k | x_{(p-k)} |^2 \cdot | \tilde{\varphi}(k) | \sum_k | \tilde{\varphi}(k) |$$

$$\leqq \sum_k | \tilde{\varphi}(k) | \sum_p | x_{(p-k)} |^2 \sum_k | \tilde{\varphi}(k) |.$$

Since

$$\sum_p | x_{(p-k)} |^2 \leqq 2\rho T,$$

we have

$$| V_j - \tilde{V}_j |^2 \leqq 2\rho T \left(\sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} | \varphi(k) | \right)^2$$

which is the assertion of Sublemma 1.

Therefore, with Y_j and Z_j as defined in Section 4.1, we have

$$Y_j = \tilde{V}_j + E_j + Z_j$$

in which

$$| E_j |^2 \leqq 2\rho T \left(\sum_{k\neq 0} \varphi(k) \right)^2.$$

This fact when combined with the following result* proves the lemma.

Sublemma 2: Consider a time-discrete channel of the type described in Section 2.1. Replace Z by the n-vector $(E + Q)$ in which E is a fixed vector and the components of Q are Gaussian random variables with zero mean with the property that there exist constants ϵ and η such that for every real n-vector U of unit length:

$$0 < \epsilon \leqq \mathcal{E} | (U,Q) |^2 \leqq \eta$$

* See Ref. 3, Appendix D, for a result related to Sublemma 2.

uniformly in n. Let $\gamma \varepsilon$ (0,1). Then any rate

$$R < \mu\, log_2 \left(1 + \frac{\gamma \rho}{2\mu\eta}\right)$$

is permissible (in the sense of Section 2.1) provided that

$$| E |^2 \leqq \vartheta T$$

for all $T > 0$, in which

$$\vartheta < 2(1 - \gamma^{\frac{1}{2}})^2 \frac{\eta\beta}{\gamma}$$

where $\beta = \beta(\eta/\gamma, \rho, \mu, R)$ is the number introduced in Section 2.1.

Proof: Let $T_0 \varepsilon (0, \infty)$. Consider the time-discrete channel of Section 2.1 with noise vector Z, but with η replaced with $(1/\gamma)\eta$. Here for

$$R < \mu \log_2 \left(1 + \frac{\gamma \rho}{2\mu\eta}\right)$$

and $T \geqq T_0$, there exists a code $\{X_j\}$ such that $X_i \neq X_j$ for $i \neq j$, and the error probability (using minimum distance decoding) given that the ith code word was transmitted

$$\hat{p}_{ei} \triangleq Pr \bigcup_{j \neq i} \{| X_i + Z - X_j | \leqq | Z |\}$$

is at most $\exp [-\beta T + \theta(T)]$ independent of i, where

$$\beta = \beta[(\eta/\gamma), \rho, \mu, R]$$

and $\theta(T)/T \to 0$ as $T \to \infty$. For this code, the error probability (using minimum distance decoding) for the channel described in Sublemma 2 is

$$p_{ei} \triangleq Pr \bigcup_{j \neq i} \{| X_i + E + Q - X_j | \leqq | E + Q |\}.$$

Let $c_{ij} \triangleq | X_i - X_j |$, and let U_{ij} denote the unit-length vector $(X_i - X_j)/c_{ij}$. Then it can easily be shown that

$$| X_i + E + Q - X_j | \leqq | E + Q |$$

if and only if

$$(U_{ij}, Q) \leqq -\tfrac{1}{2}c_{ij} - (U_{ij}, E),$$

in which $(\cdot,\cdot)$ denotes the usual inner product of n-vectors. Thus,

$$p_{ei} = Pr \bigcup_{j \neq i} \{(U_{ij}, Q) \leqq -\tfrac{1}{2}c_{ij} - (U_{ij}, E)\}$$

and similarly,

$$\hat{p}_{ei} = Pr \bigcup_{j \neq i} \{(U_{ij}, Z) \leqq -\tfrac{1}{2}c_{ij}\}. \tag{3}$$

Consider (3). Let the n-vector $P \triangleq (p_1, p_2, \cdots, p_n)$ represent a general point in Euclidean n-space $\mathcal{E}_n$, and let $\mathcal{R}_{ij}$ denote the closed half-space of $\mathcal{E}_n$ throughout which $(U_{ij}, P) \leqq -\tfrac{1}{2}c_{ij}$. Let $\mathcal{R}_i = \bigcup_{j \neq i} \mathcal{R}_{ij}$. Then

$$\hat{p}_{ei} = (2\pi)^{-n/2} \left(\frac{\eta}{\gamma}\right)^{-n/2} \int_{\mathcal{R}_i} \exp\left[-\frac{1}{2}\frac{\gamma}{\eta}\sum_{i=1}^{n} z_k^2\right] dz_1 \cdots dz_n .$$

Similarly, let $\mathcal{S}_{ij}$ denote the closed half-space throughout which

$$(U_{ij}, P) \leqq -[\tfrac{1}{2}c_{ij} + (U_{ij}, E)]\gamma^{-\frac{1}{2}},$$

and let

$$\mathcal{S}_i \triangleq \bigcup_{j \neq i} \mathcal{S}_{ij} .$$

Then, since

$$p_{ei} = Pr \bigcup_{j \neq i} \{(U_{ij}, \gamma^{-\frac{1}{2}}Q) \leqq -[\tfrac{1}{2}c_{ij} + (U_{ij}, E)]\gamma^{-\frac{1}{2}}\},$$

we have, with Λ the covariance matrix of the random variables $\{q_i\gamma^{-\frac{1}{2}}\}$,

$$p_{ei} = (2\pi)^{-n/2} (\det \Lambda)^{-\frac{1}{2}} \int_{\mathcal{S}_i} \exp[-\tfrac{1}{2}Q^t\Lambda^{-1}Q]dq_1 \cdots dq_n .$$

Let us assume that

$$[\tfrac{1}{2}c_{ij} + (U_{ij}, E)]\gamma^{-\frac{1}{2}} \geqq \tfrac{1}{2}c_{ij} \tag{4}$$

for all $j \neq i$. Then $\mathcal{S}_{ij} \subseteq \mathcal{R}_{ij}$, $\mathcal{S}_i \subseteq \mathcal{R}_i$, and hence

$$p_{ei} \leqq (2\pi)^{-n/2} (\det \Lambda)^{-\frac{1}{2}} \int_{\mathcal{R}_i} \exp[-\tfrac{1}{2}Q^t\Lambda^{-1}Q]dq_1 \cdots dq_n .$$

Let $Q = \Xi Y$, where Ξ is the orthogonal matrix such that $\Xi^{-1}\Lambda^{-1}\Xi = \text{diag}(\lambda_1^{-1}, \lambda_2^{-1}, \cdots, \lambda_n^{-1})$, with the understanding that λ_1 and λ_n denote the smallest and largest eigenvalues of Λ, respectively. Then

$$p_{ei} \leqq (2\pi)^{-n/2}(\lambda_1\lambda_2 \cdots \lambda_n)^{-\frac{1}{2}} \int_{\mathcal{R}_i'} \exp\left[-\frac{1}{2}\sum_{k=1}^{n} \lambda_k^{-1}y_k^2\right] dy_1 \cdots dy_n$$

in which $\mathcal{R}_i'$ denotes the inverse image of $\mathcal{R}_i$ under the transformation represented by Ξ. Similarly,

$$\hat{p}_{ei} = (2\pi)^{-n/2}\left(\frac{\eta}{\gamma}\right)^{-n/2}\int_{\mathcal{R}_i'}\exp\left[-\frac{1}{2}\frac{\gamma}{\eta}\sum_{k=1}^{n} y_k^2\right]dy_1\cdots dy_n .$$

Since, by assumption,

$$\epsilon \leqq \mathcal{E}\,|(U,Q)|^2 \leqq \eta$$

for every real n-vector U of unit length and every positive integer n, it follows that $\lambda_1 \geqq \epsilon\gamma^{-1}$ and $\lambda_n \leqq \eta\gamma^{-1}$. We note that for $0 < \lambda_j < \eta\gamma^{-1}$:

$$\lambda_j^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\lambda_j^{-1}y_j^2\right] \leqq \left(\frac{\eta}{\gamma}\right)^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\frac{\gamma}{\eta}y_j^2\right]$$

provided that $y_j^2 \geqq \eta/\gamma$. Thus,

$$\begin{aligned}(2\pi)^{-n/2}(\lambda_1\lambda_2\cdots\lambda_n)^{-\frac{1}{2}}&\int_{(\mathcal{R}_i'-\mathcal{C})}\exp\left[-\frac{1}{2}\sum_{k=1}^{n}\lambda_k^{-1}y_k^2\right]dy_1\cdots dy_n\\ &\leqq (2\pi)^{-n/2}\left(\frac{\eta}{\gamma}\right)^{-n/2}\int_{(\mathcal{R}_i'-\mathcal{C})}\exp\left[-\frac{1}{2}\frac{\gamma}{\eta}\sum_{k=1}^{n}y_k^2\right]dy_1\cdots dy_n\\ &\leqq \hat{p}_{ei},\end{aligned}$$

in which $\mathcal{C}$ denotes the hypercube in $\mathcal{E}_n$ defined by the inequalities: $p_j^2 \leqq \eta/\gamma$ for $j = 1, 2, \cdots, n$.

Therefore,

$$\begin{aligned}p_{ei} &\leqq \int_{\mathcal{R}_i'} \leqq \int_{(\mathcal{R}_i'-\mathcal{C})} + \int_{\mathcal{C}}\\ &\leqq \hat{p}_{ei} + (2\pi)^{-n/2}(\lambda_1\lambda_2\cdots\lambda_n)^{-\frac{1}{2}}\int_{\mathcal{C}}\exp\left[-\frac{1}{2}\sum_{k=1}^{n}\lambda_k^{-1}y_k^2\right]dy_1\cdots dy_n .\end{aligned}$$

However,

$$\begin{aligned}(2\pi)^{-n/2}(\lambda_1\lambda_2\cdots\lambda_n)^{-\frac{1}{2}}&\int_{\mathcal{C}}\exp\left[-\frac{1}{2}\sum_{k=1}^{n}\lambda_k^{-1}y_k^2\right]dy_1\cdots dy_n\\ &= \prod_{k=1}^{n}(2\pi)^{-\frac{1}{2}}\lambda_k^{-\frac{1}{2}}\int_{-\eta/\gamma}^{\eta/\gamma}e^{-\frac{1}{2}\lambda_k^{-1}y^2}dy\\ &\leqq r^n,\end{aligned}$$

in which

$$r = (2\pi)^{-\frac{1}{2}}\left(\frac{\gamma}{\epsilon}\right)^{\frac{1}{2}}\int_{-\eta/\gamma}^{\eta/\gamma}\exp\left(-\frac{1}{2}\frac{\gamma}{\epsilon}y^2\right)dy.$$

Thus,

$$p_{ei} \leqq \hat{p}_{ei} + r^n \leqq \exp[-\beta T + \theta(T)] + r^{2\mu T}. \tag{5}$$

Since $r < 1$, the right-side of (5) approaches zero as $T \to \infty$. Therefore, to complete the proof of Sublemma 2, it suffices to show that there exist values of T_0 such that (4) is satisfied (for all $j \neq i$) for all $T \geqq T_0$.

We note first that (4) is satisfied if

$$-(U_{ij}, E) \leqq \tfrac{1}{2}(1 - \gamma^{\frac{1}{2}})c_{ij} \tag{6}$$

for all $j \neq i$. Since $-(U_{ij}, E) \leqq |E|$, (6) is satisfied if

$$|E| \leqq \tfrac{1}{2}(1 - \gamma^{\frac{1}{2}})c_{ij} \tag{7}$$

for all $j \neq i$.

We now estimate the numbers c_{ij}. We have,[4] with $a \triangleq \frac{1}{2}c_{ij}(\gamma/\eta)^{\frac{1}{2}}$,

$$\exp[-\beta T + \theta(T)] \geqq \hat{p}_{ei} \geqq Pr\{(U_{ij},Z) \leqq -\tfrac{1}{2}c_{ij}\} = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{-a} e^{-\frac{1}{2}x^2} dx,$$

for any i and any $j \neq i$, since the variance of (U_{ij}, Z) is η/γ. Therefore,

$$\exp[-\beta T + \theta(T)] \geqq (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{-a} e^{-\frac{1}{2}x^2} dx = (2\pi)^{-\frac{1}{2}} \int_{a^2/2}^{\infty} \cdot e^{-y}(2y)^{-\frac{1}{2}} dy. \tag{8}$$

Let $\delta > 0$ be a constant, and let $\alpha(\delta)$ denote the smallest nonnegative number such that

$$(2y)^{-\frac{1}{2}} \geqq e^{-\delta y} \quad \text{for all} \quad y \geqq \alpha(\delta).$$

Then

$$\begin{aligned}\exp[-\beta T + \theta(T)] &\geqq (2\pi)^{-\frac{1}{2}} \int_{a^2/2}^{\infty} \exp[-(1 + \delta)y] dy \\ &\geqq (2\pi)^{-\frac{1}{2}}(1 + \delta)^{-1} \exp[-\tfrac{1}{2}(1 + \delta)a^2]\end{aligned}$$

for $a^2 \geqq 2\alpha(\delta)$, from which it follows at once that

$$a^2 \geqq 2(1 + \delta)^{-1}\beta T - 2(1 + \delta)^{-1}\{\ln[(2\pi)^{\frac{1}{2}}(1 + \delta)] + \theta(T)\}$$

for $a^2 \geqq 2\alpha(\delta)$. Since $\exp[-\beta T + \theta(T)] \to 0$ as $T \to \infty$, we see from (8) that for each $\alpha(\delta) > 0$, there exists a constant $T_\delta > 0$ such that $a^2 > 2\alpha(\delta)$ for all $T \geqq T_\delta$. Thus, for each $\delta > 0$ there exists a $T_\delta \,\varepsilon\, (0, \infty)$

such that

$$c_{ij}^{2} \geqq 8(1+\delta)^{-1}\gamma^{-1}\eta\beta T - 8(1+\delta)^{-1}\gamma^{-1}\eta\{\ln[(2\pi)^{\frac{1}{2}}(1+\delta)] + \theta(T)\} \tag{9}$$

for all $T \geqq T_\delta$.

Inequality (7) is therefore satisfied for all $T \geqq T_0$ if $T_0 > T_\delta$ and

$$|E|^2 \leqq 2(1-\gamma^{\frac{1}{2}})^2(1+\delta)^{-1}\gamma^{-1}\eta\beta T - 2(1-\gamma^{\frac{1}{2}})^2(1+\delta)^{-1}\gamma^{-1}\eta\{\ln[(2\pi)^{\frac{1}{2}}(1+\delta)] + \theta(T)\} \tag{10}$$

for all $T \geqq T_0$. By assumption: $|E|^2 \leqq \vartheta T$ for all $T > 0$, in which

$$\vartheta < 2(1-\gamma^{\frac{1}{2}})^2\gamma^{-1}\eta\beta.$$

Choose $\delta > 0$ so that

$$\vartheta < 2(1-\gamma^{\frac{1}{2}})^2(1+\delta)^{-1}\gamma^{-1}\eta\beta,$$

and then let $T_0 \,\varepsilon\, [T_\delta, \infty)$ be so large that

$$\vartheta \leqq 2(1-\gamma^{\frac{1}{2}})^2(1+\delta)^{-1}\gamma^{-1}\eta\beta - 2(1-\gamma^{\frac{1}{2}})^2(1+\delta)^{-1}\gamma^{-1}\eta T^{-1}\{\ln[(2\pi)^{\frac{1}{2}}(1+\delta)] + \theta(T)\}$$

for all $T \geqq T_0$. Then (10) is satisfied for all $T \geqq T_0$. This completes the proof of Sublemma 2.

V. FINAL REMARKS

The writer is indebted to D. Hamming and L. A. Shepp for discussions concerning this work, and to J. Savage, D. Slepian, and A. Wyner for commenting on the draft.

REFERENCES

1. Shannon, C. E., Communication in the Presence of Noise, Proc. IRE, *37*, January, 1949, pp. 10–21.
2. Slepian, D., Bounds on Communication, B.S.T.J., *42*, 1963, p. 681, (see Appendix B).
3. Ash, R. B., *Information Theory*, John Wiley & Sons, Inc., 1965, p. 256.
4. Wyner, A. D., The Capacity of the Bandlimited Gaussian Channel, B.S.T.J., *45*, March, 1966, p. 364.
5. Shannon, C. E., Probability of Error for Optimal Codes in a Gaussian Channel, B.S.T.J., *38*, May, 1959, p. 611.