INFORMATION AND CONTROL 6, 14–27 (1963)

Capacity and Error Bounds for a Time-Continuous Gaussian Channel*

R. B. Ash

Columbia University, New York, New York†

A model proposed by Fortet for a time-continuous Gaussian channel is analyzed. The model differs from that of Shannon in that a different constraint is imposed on the allowable input signals, and in addition the transmission of a code word in a given time interval is not assumed to interfere with the transmission of a word in any other interval. The capacity of the Fortet channel is established by proving a coding theorem and its weak converse. It is shown that the probability of error of an optimal code approaches zero exponentially with the time duration of the code words, provided the transmission rate is below channel capacity.

INTRODUCTION

Recently, Fortet (1961) introduced a new model for a time-continuous channel with additive Gaussian noise of arbitrary spectrum. The model[1] differs from that of Shannon (1948) in the following way. In each case the code words are truncated versions $s_T(t)$ of signals $s(t)$ limited to the same frequency band as the noise. In other words, $s_T(t) = s(t)$, $-T \leqq t \leqq T$; $s_T(t) = 0$ elsewhere, where the Fourier transform of $s(t)$ vanishes wherever the noise spectrum vanishes. A decision is made as to the identity of the code word transmitted during the interval $[-T, T]$ after observing the output during that interval. However, implicit in the Shannon theory is the assumption that the *entire* signal $s(t)$ is transmitted through the channel. Consequently the particular signal chosen for transmission during $[-T, T]$ could conceivably interfere with the reception of code words during all other intervals, past as well as future.

* This research was supported by the National Science Foundation under Grant No. G-15965.

† *Present address:* Department of Electrical Engineering, University of California, Berkeley, California.

[1] As interpreted by the author.

Shannon solves the interference problem by postulating a decoding procedure based on observation of the output at discrete sampling instants, and restricting the class of allowable signals to those with only a finite number of nonzero samples.

Fortet's model is static in the sense that no word-to-word dependence is assumed. Although the signals $s(t)$ from which the code words are derived have in general an infinite time duration, attention is restricted in the analysis to the finite interval $[-T, T]$. The effect of the portion of $s(t)$ outside $[-T, T]$ on past and future transmissions is ignored. In addition, Fortet imposes a constraint on the input signals which depends upon the spectrum of the noise. Fortet's constraint is not the same as the "average power" constraint imposed by Shannon, except in the case where the noise is band limited with a flat spectrum. In this case, the basic difference between the two approaches becomes clear. Both Shannon and Fortet derive their code words from band limited signals. However, in restricting his attention to the interval $[-T, T]$ Fortet is essentially time limiting his code words. Thus, it is not surprising that the capacity of the Fortet channel in this case agrees with Shannon's capacity formula for the limiting case of infinite bandwidth.

In this paper, the capacity of the Fortet channel is established by proving a coding theorem and its (weak) converse. In addition, it is shown that the probability of error of an optimal code approaches zero exponentially with the time duration of the code words.

It must be emphasized that there is no contradiction between the results of the present paper and those of Shannon; the models are quite different. The Shannon model is physically more realistic, but the Fortet model is more tractable mathematically. There are certain difficulties in the Shannon formulation of a model for a band limited channel which have not yet been resolved.

DEFINITIONS AND STATEMENTS OF RESULTS

Let $n(t)$ be a stationary Gaussian stochastic process with zero mean, continuous covarance function[2] $R(\tau)$, and spectral density $N(\omega)$, with

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} | N(\omega) |^2 \, d\omega < \infty$$

[2] To avoid degeneracy we assume that the eigenfunctions of the integral equation $\int_{-T}^{T} R(t - \tau) g(\tau) \, d\tau = \lambda g(t)$ span the entire L_2 space of square integrable functions over $[-T, T]$. This will be the case, according to Root and Pitcher (1955) if $n(t)$ is filtered white noise.

Consider the class of real functions of integrable square over $(-\infty, \infty)$, whose Fourier transforms are zero whenever $N(\omega)$ is zero. If $S(\omega)$ is the Fourier transform of a function $s(t)$ in this class, let $F(\omega) = S(\omega)/\sqrt{N(\omega)}$ and let $f(t)$ be the inverse Fourier transform of $F(\omega)$. ($F(\omega)$ will have an inverse Fourier transform, at least in the sense of a limit in the mean, if $F(\omega)$ is of integrable square; this will be the case in the problem under consideration.) For any positive real number T, define $s_T(t) = s(t)$, $-T \leqq t \leqq T$; $s_T(t) = 0$ elsewhere; similarly define $n_T(t) = n(t)$, $-T \leqq t \leqq T$; $n_T(t) = 0$ elsewhere. A function $s(t)$ (and its corresponding $s_T(t)$) will be called *allowable* if

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|S(\omega)|^2}{N(\omega)} d\omega \leqq KT \tag{1}$$

where K is a positive constant which is fixed for the remainder of the discussion. The integrand is defined to be zero whenever $N(\omega) = 0$. We define a *time-continuous Gaussian channel* as follows. Let T be an arbitrary but fixed positive real number. The inputs to the channel will be allowable functions $s_T(t)$; the outputs will be functions $s_T(t) + n_T(t)$. An input is chosen from a specified set and transmitted through the channel, where it is corrupted by the additive "noise" $n_T(t)$. The received signal is observed over the interval $[-T, T]$ and then a decision is made as to the identity of the input signal. More formally, following Wolfowitz (1961) we define a *code* (T, M, β) as a set

$$\{(s_1(t), A_1), (s_2(t), A_2), \cdots, (s_M(t), A_M)\}$$

where each $s_i(t)$ is an allowable function $s_T(t)$ and the A_j are disjoint Borel sets in function space such that

$$P\{s_j(t) + n_T(t) \in A_j\} \geqq 1 - \beta \qquad (j = 1, 2, \cdots, M) \tag{2}$$

Thus the "probability of error" does not exceed β no matter which "code word" $s_i(t)$ is transmitted. A number R is called a *permissible rate of transmission* if for each T there is a code $(T, [e^{RT}], \beta(T))$ such that $\beta(T) \to 0$ as $T \to \infty$. The *channel capacity* C is the supremum of all permissible transmission rates. (Clearly zero is a permissible transmission rate so that the set of permissible rates is not empty.) The main results of this paper are:

THEOREM 1. $C = K/2$.

THEOREM 2. (*Exponential Bound*) *For each* $R = C - \epsilon$, $0 < \epsilon < C$, *there exist positive constants A and B (depending on K and* ϵ*) such that for each T there is a code* $(T, [e^{RT}], \beta(T))$ such that $\beta(T) \leqq Ae^{-BT}$.

Remarks: The result $C \geqq K/2$ is credited to Bethoux in Fortet (1961); the argument given here seems more direct than the one sketched by Fortet. The results $C \leqq K/2$ and Theorem 2 are new, although the method of attack leans heavily on the ideas of Feinstein (1958), Blackwell, Breiman, and Thomasian (1958, 1959) and Thomasian (1960). If $N(\omega) = N/2(-2\pi W \leqq \omega \leqq 2\pi W)$; $N(\omega) = 0$ elsewhere, and $K = 2P/N$, we have $(2\pi)^{-1}\int_{-\infty}^{\infty} | S(\omega) |^2 d\omega \leqq PT$, i.e., the "average signal power" is limited; this is the constraint imposed by Shannon (1948). Using the Sampling Theorem Shannon obtained $C = W \log (1 + P/NW)$, which is always less than $K/2 = P/N$ and approaches P/N only as the bandwidth W becomes infinite. Since in the present problem, the signals which are actually transmitted through the channel are time limited and hence not band limited, the Sampling Theorem is not applicable in this case. As indicated in the introduction, the time limiting of the code words accounts for the agreement between the channel capacity as given by Theorem 1 and the Shannon formula for the infinite bandwidth case. Note that if no constraint at all is put on the inputs, the capacity is clearly infinite; this has been pointed out by Good and Doog (1958) and Swerling (1960).

PROOF OF THE DIRECT HALF OF THEOREM 1 ($C \geqq K/2$)

It is well known that the process $n(t)$ can be represented over $[-T, T]$ as

$$n_T(t) = \sum_{n=1}^{\infty} z_n g_n(t) \tag{3}$$

where the $g_n(t)$ are the orthonormalized eigenfunctions of the integral equation

$$\int_{-T}^{T} R(t - \tau) g(\tau)\, d\tau = \lambda g(t), \quad -T \leqq t \leqq T \tag{4}$$

The $z_n = \int_{-T}^{T} n_T(t) g_n(t)\, dt$ are independent, normally distributed random variables with zero mean and variance λ_n, where λ_n is the eigenvalue corresponding to $g_n(t)$. The series (3) converges in mean square for all t in $[-T, T]$.

Now let $h(t)$ be the inverse Fourier transform of $\sqrt{N(\omega)}$ ($h(t)$ exists since $(2\pi)^{-1} \int_{-\infty}^{\infty} N(\omega)\, d\omega = R(0) < \infty$; note that $h(-t) = h(t)$ since

$N(-\omega) = N(\omega)$.) Following Kelly, Reed, and Root (1960, Appendix 1), we define an auxiliary set of functions $\phi_n(t)$ by

$$\phi_n(t) = \frac{1}{\sqrt{\lambda_n}} \int_{-T}^{T} h(t-s) g_n(s)\, ds, \quad -\infty < t < \infty \tag{5}$$

To prove the direct half of Theorem 1 we shall consider signals generated in the following manner. Let $f(t)$ be any function of the form $\sum_{i=1}^{n} f_i \phi_i(t)$, $-\infty < t < \infty$, with $\sum_{i=1}^{n} f_i^2 \leqq KT$. Let $F(\omega)$ be the Fourier transform of $f(t)$. Let $s(t)$ be the signal whose Fourier transform is $S(\omega) = F(\omega)\sqrt{N(\omega)}$, i.e.,

$$s(t) = \int_{-\infty}^{\infty} h(t-s) f(s)\, ds \tag{6}$$

The "coordinates" of $s(t)$ with respect to the functions $g_n(t)$ are:

$$\begin{aligned} x_i &= \int_{-T}^{T} s(t) g_i(t)\, dt \\ &= \int_{-T}^{T} g_i(t) \left[\int_{-\infty}^{\infty} h(t-s) f(s)\, ds \right] dt \end{aligned} \tag{7}$$

The integral (7) is absolutely convergent, and thus the order of integration may be interchanged, whence

$$\begin{aligned} x_i &= \int_{-\infty}^{\infty} f(s) \left[\int_{-T}^{T} h(s-t) g_i(t)\, dt \right] ds \\ &= \sqrt{\lambda_i} \int_{-\infty}^{\infty} f(s) \phi_i(s)\, ds = f_i \sqrt{\lambda_i} \end{aligned} \tag{8}$$

The last equality of (8) follows since, as may be checked by a direct computation or by referring to Kelly, Reed, and Root (1960), the functions $\phi_n(t)$ are orthonormal on the real line.

Any signal obtained in the above manner is allowable, since

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|S(\omega)|^2}{N(\omega)}\, d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2\, d\omega = \sum_{i=1}^{n} f_i^2 \leqq KT$$

In fact, $s(t)$ is completely characterized by the vector $(x_1, x_2, \cdots, x_n)$ and

$$\sum_{i=1}^{n} \frac{x_i^2}{\lambda_i} \leqq KT \tag{9}$$

If the signal $s_T(t)$ is transmitted, the coordinates of the received signal with respect to the functions $g_n(t)$ are

$$y_i = \int_{-T}^{T} [s(t) + n(t)]g_i(t)\, dt = x_i + z_i \tag{10}$$

Having observed $\mathbf{y} = (y_1, \cdots, y_n)$ we attempt to determine the coordinates $\mathbf{x} = (x_1, \cdots, x_n)$ of the input signal. If we assume that the input signals are derived by the procedure described above, and that a decision is made at the receiver based only on the first n coordinates of the output, then the original channel is equivalent to a time-discrete memoryless channel. The input to the memoryless channel is a vector $\mathbf{x} = (x_1, \cdots, x_n)$. The output is a vector $\mathbf{y} = (y_1, \cdots, y_n) = \mathbf{x} + \mathbf{z}$ where $\mathbf{z} = (z_1, \cdots, z_n)$. The conditional density of $\mathbf{y}$ given $\mathbf{x}$ is

$$p(\mathbf{y}/\mathbf{x}) = \left[\prod_{i=1}^{n} (2\pi\lambda_i)^{-1/2}\right] \exp\left(-\sum_{j=1}^{n} \frac{(y_j - x_j)^2}{2\lambda_j}\right) \tag{11}$$

Let us now construct a probability distribution on the set of input vectors $\mathbf{x}$ by assuming that the coordinates of $\mathbf{x}$ are independent, normally distributed random variables with zero means and variances σ_i^2, $i = 1, 2, \cdots, n$. In other words, let

$$p(x_1, \cdots, x_n) = \left[\prod_{i=1}^{n} (2\pi\sigma_i^2)^{-1/2}\right] \exp\left(-\sum_{j=1}^{n} x_j^2/2\sigma_j^2\right) \tag{12}$$

The densities $p(\mathbf{x})$ and $p(\mathbf{y}/\mathbf{x})$ induce a density $p(\mathbf{y})$ on the set of output vectors.

For any positive number a, let A be the set of pairs $(\mathbf{x}, \mathbf{y})$ such that $\log [p(\mathbf{y}/\mathbf{x})/p(\mathbf{y})] > a$. Let E be *any* set of input vectors. The following result was proved by Thomasian (1960, Theorem 2) as an extension of a result of Feinstein:

FEINSTEIN-THOMASIAN LEMMA: *Given any integer* $M \geqq 1$, *there exist* M *distinct inputs* $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M$, *all belonging to* E, *and* M *disjoint output sets* $B_1, B_2, \cdots, B_M$ *such that*

$$P(\bar{B}_i/\mathbf{x}_i) \leqq Me^{-a} + P(\bar{E}) + P(\bar{A}), \qquad i = 1, 2, \cdots, M \tag{13}$$

The bar over a set denotes complementation; $P(\bar{B}_i/\mathbf{x}_i)$ *is the probability that the output does not belong to the set* B_i, *given that the input* $\mathbf{x}_i$ *is transmitted.*

Given an arbitrary but fixed ϵ, with $0 < \epsilon < K/2$, assume

$$\sum_{i=1}^{n} \frac{\sigma_i^2}{\lambda_i} \leqq \left(K - \frac{\epsilon}{2}\right) T \tag{14}$$

Let us take E to be the set of $\mathbf{x}$ satisfying (9), and let $a = (K - \epsilon)T/2$. The Feinstein-Thomasian Lemma yields a code consisting of M input vectors satisfying (9), with a probability of error bounded above by

$$Me^{-(K-\epsilon)T/2} + P\left\{\sum_{j=1}^{n} \frac{x_j^2}{\lambda_j} > KT\right\} + P\left\{\sum_{j=1}^{n} \log\left[\frac{p(y_j/x_j)}{p(y_j)}\right] \leqq \frac{(K-\epsilon)T}{2}\right\} \tag{15}$$

Let us first consider the term $P(\bar{A})$. From (15) we obtain

$$P(\bar{A}) = P\left\{\sum_{j=1}^{n}\left[\frac{1}{2}\log\left(1 + \frac{\sigma_j^2}{\lambda_j}\right) + \frac{y_j^2}{2(\lambda_j + \sigma_j^2)} - \frac{(y_j - x_j)^2}{2\lambda_j}\right] \leqq \frac{(K-\epsilon)T}{2}\right\} \tag{16}$$

It is easy to verify that the mean and variance of

$$\frac{y_j^2}{2(\lambda_j + \sigma_j^2)} - \frac{(y_j - x_j)^2}{2\lambda_j}$$

are respectively 0 and $\sigma_j^2/(\lambda_j + \sigma_j^2)$. We are free to choose the σ_i^2 subject to (14). Let us take

$$\sigma_i^2 = \frac{\lambda_i(K - \epsilon/2)T}{n}, \qquad i = 1, \cdots, n \tag{17}$$

Then

$$P(\bar{A}) = P\left\{\frac{n}{2}\log\left(1 + \frac{(K-\epsilon/2)T}{n}\right) + W_n \leqq \frac{(K-\epsilon)T}{2}\right\} \tag{18}$$

where W_n is a random variable with zero mean and variance

$$\frac{(K-\epsilon/2)Tn}{n + (K-\epsilon/2)T} < (K - \epsilon/2)T$$

Now for sufficiently large n, in particular

$$n \geqq \max\left[\frac{2(K-\epsilon/2)^2T}{\epsilon}, 2(K-\epsilon/2)T\right] \tag{19}$$

we have

$$n\log\left(1 + \frac{(K-\epsilon/2)T}{n}\right) > (K-\epsilon/2)T - \frac{(K-\epsilon/2)^2T^2}{2n} \tag{20}$$

It follows from (18) and (20) that for any n satisfying (19),

$$P(\bar{A}) \leqq P\{W_n \leqq (K - \epsilon)T/2 - (K - \epsilon/2)T/2 + (K - \epsilon/2)^2T^2/4n\}$$
$$= P\{W_n \leqq -\epsilon T/4 + (K - \epsilon/2)^2T^2/4n\}$$
$$\leqq P\{W_n \leqq -\epsilon T/8\} \leqq P\{|W_n| \geqq \epsilon T/8\} \quad (21)$$

Chebyshev's inequality yields

$$P(\bar{A}) \leqq \frac{64(K - \epsilon/2)}{\epsilon^2 T} \quad (22)$$

We now consider the term $P(\bar{E})$. It follows by a direct computation that $\sum_{j=1}^{n} x_j^2/\lambda_j$ has mean $(K - \epsilon/2)T$ and variance $2(K - \epsilon/2)^2T^2/n$. Thus

$$P\left\{\sum_{j=1}^{n} \frac{x_j^2}{\lambda_j} > KT\right\} = P\left\{\sum_{j=1}^{n} \frac{x_j^2}{\lambda_j} - \left(K - \frac{\epsilon}{2}\right)T > \frac{\epsilon T}{2}\right\}$$
$$\leqq \frac{8(K - \epsilon/2)^2}{n\epsilon^2} \quad (23)$$

Finally, for a given T choose any n satisfying (19); then $n \to \infty$ as $T \to \infty$. Let M be any integer $\leqq \exp\left(\left(\frac{1}{2}K - \epsilon\right)T\right)$. The foregoing procedure yields a code $(T, M, \beta(T))$ where

$$\beta(T) \leqq e^{-\epsilon T/2} + \frac{64(K - \epsilon/2)}{\epsilon^2 T} + \frac{8(K - \epsilon/2)^2}{n\epsilon^2} \to 0 \quad \text{as} \quad T \to \infty$$

This completes the proof.

PROOF OF THEOREM 2

The derivation of the exponential bound is based on a sharper estimate of the terms of the expression (15). First of all from (21) we have

$$P(\bar{A}) \leqq P\{W_n \leqq -\epsilon T/8\}$$

where (24)

$$W_n = \sum_{j=1}^{n} \left[\frac{y_j^2}{2(\lambda_j + \sigma_j^2)} - \frac{(y_j - x_j)^2}{2\lambda_j}\right]$$

The following result is due to Thomasian (1960):

If $x_1^*, \cdots, x_n^*$ (resp. $z_1^*, \cdots, z_n^*$) are independent, identically distributed Gaussian random variables with mean zero and variance

Q (resp. N), the random vectors $(x_1^*, \cdots, x_n^*)$ *and* $(z_1^*, \cdots, z_n^*)$ *are independent, and* $y_j^* = x_j^* + z_j^*(j = 1, 2, \cdots, n)$ *then*

$$P\left\{\sum_{j=1}^{n}\left[\frac{(y_j^*)^2}{2(Q+N)} - \frac{(z_j^*)^2}{2N}\right] \leqq -n\delta\right\} \leqq \exp\left(\frac{-n}{4}\left[\sqrt{1+4\delta^2\left(1+\frac{N}{Q}\right)} - 1\right]\right) \tag{25}$$

To apply (25) take $x_j^* = x_j/\sqrt{\lambda_j}$, $z_j^* = z_j/\sqrt{\lambda_j}$, $y_j^* = y_j/\sqrt{\lambda_j}$, $N = 1$, $Q = \sigma_j^2/\lambda_j = (K - \epsilon/2)T/n$, $\delta = \epsilon T/8n$. From (24) and (25) we obtain

$$P(\bar{A}) \leq \exp\left(\frac{-n}{4}\left[\sqrt{1+4\left(\frac{\epsilon T}{8n}\right)^2\left(1+\frac{n}{(K-\epsilon/2)T}\right)} - 1\right]\right) \leqq \exp\left(\frac{-n}{4}\left[\sqrt{1+4\left(\frac{\epsilon T}{8n}\right)^2} - 1\right],\right. \tag{26}$$

We next bound $P(\bar{E})$ as follows:

$$P(\bar{E}) = P\left\{\sum_{j=1}^{n}\frac{x_j^2}{\lambda_j} > KT\right\} = P\left\{\sum_{j=1}^{n}\frac{x_j^2}{\sigma_j^2}\frac{\sigma_j^2}{\lambda_j} > KT\right\} = P\left\{\sum_{j=1}^{n}\frac{x_j^2}{\sigma_j^2} > \frac{Kn}{(K-\epsilon/2)}\right\} \tag{27}$$

The following result is also proved by Thomasian (1960):

If $x_1^*, \cdots, x_n^*$ *are independent random variables, each normally distributed with mean zero and variance one, and d is any real number greater than one, then*

$$P\left\{\sum_{j=1}^{n}(x_j^*)^2 > nd\right\} \leqq (de^{1-d})^{n/2} \tag{28}$$

Comparing (27) and (28)

$$P(\bar{E}) \leqq \left[\frac{K}{(K-\epsilon/2)}\exp\left(-\frac{\epsilon/2}{(K-\epsilon/2)}\right)\right]^{n/2} \tag{29}$$

or

$$P(\bar{E}) \leqq \exp\left[-\frac{n}{2}\log b(K,\epsilon)\right] \tag{30}$$

where

$$b(K, \epsilon) = \frac{K - \epsilon/2}{K} \exp\left(\frac{\epsilon/2}{K - \epsilon/2}\right) > 1.$$

Thus if M is any integer $\leqq \exp((\frac{1}{2}K - \epsilon)T)$, we can construct a code $(T, M, \beta(T))$ with

$$\beta(T) \leqq e^{-\epsilon T/2} + \exp\left(-\frac{n}{4}\left[\sqrt{1 + 4\left(\frac{\epsilon T}{8n}\right)^2} - 1\right]\right) + \exp\left[-\frac{n}{2}\log b(K, \epsilon)\right] \quad (31)$$

For each T we are free to choose n. If T is an integer and $n = T$, an examination of (31) shows that Theorem 2 is proved, with $A = 3$ and

$$B = \min\left[\tfrac{1}{2}\epsilon, \tfrac{1}{4}(\sqrt{1 + \epsilon^2/16} - 1), \tfrac{1}{2}\log b(K, \epsilon)\right] \quad (32)$$

A similar bound is easily obtained when T is not an integer.

PROOF OF THE CONVERSE HALF OF THEOREM 1 ($C \leqq K/2$)

We shall show that any code (T, M, β) with $\beta < \frac{1}{2}$ must satisfy

$$\log M < \frac{KT/2 + \log 2}{1 - 2\beta} \quad (33)$$

and therefore if $M \geqq \exp((\frac{1}{2}K + \epsilon)T)$, then

$$2\beta > \frac{\epsilon - (1/T)\log 2}{\epsilon + K/2} \quad (34)$$

Thus the probability of error cannot approach zero as $T \to \infty$.

The idea is to approximate a given code (T, M, β) by a code for a discrete memoryless channel. Since the eigenfunctions of the integral equation

$$\int_{-T}^{T} R(t - \tau)g(\tau)\, d\tau = \lambda g(t), \qquad -T \leqq t \leqq T,$$

span the Hilbert space $L_2[-T, T]$, there is a one to one correspondence between square integrable functions over $[-T, T]$ and square summable sequences. The sequence $(x_1, x_2, \cdots)$ corresponding to a function $x(t)$ consists of the coordinates of $x(t)$ with respect to the "basis functions" $g_n(t)$. Thus for each decoding set A_i in function space there is a corresponding Borel set A_i^* in sequence space with the following property. If the code word $s_i(t)$ has the representation

$$s_i(t) = \sum_{n=1}^{\infty} s_{in} g_n(t), \qquad -T \leqq t \leqq T;\ i = 1, 2, \cdots, M \quad (35)$$

where

$$s_{in} = \int_{-T}^{T} s_i(t) g_n(t)\, dt$$

and the series (35) converges in the mean, then we may write

$$P\{(s_{i1}, s_{i2}, \cdots) + (z_1, z_2, \cdots) \in A_i^*\} \geqq 1 - \beta (i = 1, 2, \cdots, M) \quad (36)$$

For each A_i^* there is a measurable cylinder $B_i \subset A_i^*$ such that

$$P\{(s_{i1}, s_{i2}, \cdots) + (z_1, z_2, \cdots) \in B_i\} \geqq 1 - 2\beta$$

Now membership in a measurable cylinder in sequence space is determined by a finite number of coordinates. Since there is only a finite number of code words, there is an integer n such that the base of each B_i is n-dimensional. Consequently,

$$P\{(s_{i1}, s_{i2}, \cdots, s_{in}) + (z_1, z_2, \cdots, z_n) \in B_{in}\} \geqq 1 - 2\beta \quad (i = 1, 2, \cdots, M) \quad (37)$$

where B_{in} is the base of B_i.

Equivalently,

$$P\left\{\left(\frac{s_{i1}}{\sqrt{\lambda_1}}, \cdots, \frac{s_{in}}{\sqrt{\lambda_n}}\right) + \left(\frac{z_1}{\sqrt{\lambda_1}}, \cdots, \frac{z_n}{\sqrt{\lambda_n}}\right) \in B'_{in}\right\} \geqq 1 - 2\beta \qquad (i = 1, 2, \cdots, M) \quad (38)$$

where B'_{in} is formed from B_{in} by dividing the jth component of each vector in B_{in} by $\sqrt{\lambda_j} (j = 1, 2, \cdots, n)$. The sets B'_{in} (as well as the B_{in}, B_i and A_i^*) are of course disjoint. Thus the vectors $(s_{i1}/\sqrt{\lambda_1}, \cdots, s_{in}/\sqrt{\lambda_n})$, $i = 1, 2, \cdots, M$, may be considered as code words of a code $(n, M, 2\beta)$ (i.e., a code consisting of M vectors of dimension n with a probability of error $\leqq 2\beta$), for a time-discrete memoryless channel with noise variance unity. Since each code word is allowable, i.e., satisfies

(1), it follows (e.g., Kelly, Reed and Root (1960, Appendix 1, equation (I-18)), or Fortet (1961)) that

$$\sum_{j=1}^{n} \frac{s_{ij}^2}{\lambda_j} \leqq KT \quad (i = 1, 2, \cdots, M) \tag{39}$$

Therefore the coordinates $(x_1, x_2, \cdots, x_n)$ of any code word satisfy

$$\frac{1}{n} \sum_{j=1}^{n} x_j^2 \leqq \frac{KT}{n} \tag{40}$$

By the weak converse for the zero memory Gaussian channel (proved in the appendix), any code $(n, M, 2\beta)$ whose code words meet the constraint (40) must satisfy

$$\log M < \frac{\frac{1}{2}n \log [1 + (KT/n)] + \log 2}{1 - 2\beta} \tag{41}$$

Thus

$$\log M < \frac{\frac{1}{2}KT + \log 2}{1 - 2\beta} \qquad \text{and the proof is complete.}$$

Remark: The strong converse for the zero-memory Gaussian channel (Wolfowitz, 1961, Chap. 9) does not seem to be directly applicable here. The strong converse states that for n *sufficiently large*, any code must satisfy $\log M < n(C + \epsilon)$ where C is the channel capacity. The presence of n in the denominator of (40) creates difficulties; the value of n necessary for the strong converse to hold depends on the average power limitation on the input, which in this case depends on n. However, it seems likely that Wolfowitz's proof can be suitably modified so as to apply to the present problem.

APPENDIX

Since a proof of the weak converse for the zero-memory Gaussian channel is apparently not available in the literature, we sketch one here. Consider any code (n, M, β) for the memoryless Gaussian channel with noise variance N and input constraint $(1/n \sum_{j=1}^{n} x_j^2 \leqq P$. We shall prove that the code satisfies

$$\log M < \frac{nC + \log 2}{1 - \beta} \tag{A.1}$$

where $C = \frac{1}{2} \log[1 + (P/N)]$ is the channel capacity.

We use the letter Q, with various affixes, to denote distribution functions. We denote by $I(Q)$ the "information rate" associated with Q, i.e.,

$$I(Q) = H_Q(x) - H_Q(x \mid y) = H_Q(y) - H_Q(y \mid x) \tag{A.2}$$

where H is the entropy function.

Suppose that the code words of a given code (n, M, β) are $(x_{11}, x_{12}, \cdots, x_{1n})$, $(x_{21}, x_{22}, \cdots, x_{2n})$, $\cdots$, $(x_{M1}, x_{M2}, \cdots, x_{Mn})$. For any real number x, let

$$Q_0(x) = \frac{1}{Mn} \text{ (the number of components } x_{ij},\ i = 1, 2, \cdots, M,\ j = 1, 2, \cdots, n, \text{ which are} \leqq x)$$

For each $j = 1, 2, \cdots, n$, let

$$Q^{(j)}(x) = \frac{1}{M} \text{ (the number of components } x_{ij},\ i = 1, 2, \cdots, M, \text{ which are} \leqq x)$$

It follows that

$$Q_0(x) = \frac{1}{n} \sum_{j=1}^{n} Q^{(j)}(x) \tag{A.3}$$

By the concavity of the information rate (Fano, 1961, p. 131)

$$I(Q_0) \geqq \frac{1}{n} \sum_{j=1}^{n} I(Q^{(j)}) \tag{A.4}$$

Since, $(1/Mn) \sum_{i=1}^{M} \sum_{j=1}^{n} x_{ij}^2 \leqq P$, it follows that the variance of Q_0 is $\leqq P$. Thus $I(Q_0)$ cannot exceed the information rate corresponding to a Gaussian input with variance P, i.e.,

$$I(Q_0) \leqq C \tag{A.5}$$

Let $Q_u(\mathbf{x})$ be the n-dimensional distribution function which assigns equal probability to each code word. It follows from Fano (1961, p. 125) that

$$I(Q_u) \leqq \sum_{j=1}^{n} I(Q^{(j)}) \tag{A.6}$$

We may write $I(Q_u)$ as

$$I(Q_u) = H_{Q_u}(\mathbf{x}) - H_{Q_u}(\mathbf{x} \mid \mathbf{y}) = \log M - H_{Q_u}(\mathbf{x} \mid \mathbf{y}) \tag{A.7}$$

By Fano's inequality (Fano, 1961, p. 187)

$$H_{Q_u}(\mathrm{x} \mid \mathrm{y}) < \log 2 + \beta \log M \tag{A.8}$$

The results (A.3) through (A.8) yield the desired result (A.1).

Acknowledgment

The author is indebted to Professor A. J. Thomasian of the University of California at Berkeley for many helpful suggestions, as well as for the proof which appears in the Appendix.

Received: August 27, 1962

References

Blackwell, D., Breiman, L., and Thomasian, A. J. (1958), Proof of Shannon's transmission theorem for finite state indecomposable channels. *Ann. Math. Statist.* **29,** 1209–1220.

Blackwell, D., Breiman, L., and Thomasian, A. J. (1959), The capacity of a class of channels. *Ann. Math. Statist.* **30,** 1229–1241.

Fano, R. M. (1961), "Transmission of Information," M.I.T, Press and Wiley, New York.

Feinstein, A. (1958), "Foundations of Information Theory." McGraw-Hill, New York.

Fortet, R. (1961), Hypothesis testing and estimation for Laplacian functions. "Fourth Berkeley Symposium on Mathematical Statistics and Probability," Vol. 1, pp. 289–305. University of California Press.

Good, I. J. and Caj Doog, K. (1958), A paradox concerning rate of information. *Information and Control* **1,** 113–126.

Kelly, E. J., Reed, I. S., and Root, W. L. (1960), The detection of radar echoes in noise. *J. Soc. Ind. Appl. Math.* **8,** 309–341.

Root, W. L. and Pitcher, T. S. (1955), Some remarks on signal detection. *I.R.E. Trans. Inform. Theory* **IT1,** No. 3, pp. 33–38.

Shannon, C. E. (1948), A mathematical theory of communication. *Bell System Tech. J.* **27,** 379–423, 623–656.

Swerling, P. (1960), Paradoxes related to the rate of transmission of information. *Information and Control* **3,** 351–359.

Thomasian, A. J. (1960), Error bounds for continuous channels. "Fourth London Symposium on Information Theory," pp. 46–60. Butterworth Scientific Publications, London.

Wolfowitz, J. (1961), "Coding Theorems of Information Theory." Springer-Verlag, Berlin.