# THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 53 May–June 1974 Number 5



# Capacity of the Gaussian Channel With Memory: The Multivariate Case

By L. H. BRANDENBURG and A. D. WYNER

(Manuscript received November 13, 1973)

*A formula is derived for the capacity of a multi-input, multi-output linear channel with memory, and with additive Gaussian noise. The formula is justified by a coding theorem and converse. The channel model under consideration can represent multipair telephone cable including the effect of far-end crosstalk. For such cable under large signal-to-noise conditions, we show that channel capacity and cable length are linearly related; for small signal-to-noise ratio, capacity and length are logarithmically related. Crosstalk tends to reduce the dependence of capacity on cable length. Moreover, for any channel to which our capacity formula applies, and for large signal-to-noise ratio, there is an asymptotic linear relation between capacity and signal-to-noise ratio with slope independent of the channel transfer function. For small signal-to-noise ratio, capacity and signal-to-noise ratio are logarithmically related. Also provided is a numerical evaluation of the channel capacity formula, using measured parameters obtained from an experimental cable.*

## I. INTRODUCTION AND STATEMENT OF RESULTS

Our problem is to calculate the capacity of a multi-input, multi-output linear channel with additive Gaussian noise, and to justify the formula by a coding theorem and converse. Specifically, we consider the following channel. The channel input and output are sequences

$\{x(n)\}_{-\infty}^{\infty}$, $\{y(n)\}_{-\infty}^{\infty}$ of real $s$-vectors* related by

$$y(n) = \sum_{k=-\infty}^{\infty} h(n-k)x(k) + z(n), \tag{1}$$

where $\{h(m)\}_{m=-\infty}^{\infty}$ is a fixed sequence of real $s \times s$ matrices (the indicated operations being ordinary matrix arithmetic), and $\{z(n)\}_{-\infty}^{\infty}$ is a sequence of Gaussian random $s$-vectors for which $Ez(n) = 0$, and

$$Ez(n)[z(n-m)]^t = r(m), \quad -\infty < n, \quad m < \infty, \tag{2}$$

where $r(m)$ is an $s \times s$ matrix. The motivation for this problem is that it is a model for a multipair telephone cable.

The first sections of this paper are highly theoretical; the formula for channel capacity is carefully and precisely established by means of several rather technical theorems. In the final section, Section IV, we discuss some engineering implications of our formula in terms of its asymptotic behavior, and evaluate the capacity numerically with measured parameters obtained from an experimental multipair telephone cable.

A *code* for this channel with parameters $(M, N, S, \lambda)$ is a set of $M$ pairs $\{(\mathbf{x}_i, B_i)\}_{i=1}^{M}$, where $\mathbf{x}_i = (\cdots, x_i^t(-2), x_i^t(-1), x_i^t(0), x_i^t(1), \cdots)^t$ is a sequence of $s$-vectors that satisfy the following:

$$x_i(n) = 0, \quad \text{for} \quad n < 0, \quad \text{and} \quad n \geqq N, \tag{3a}$$

$$\frac{1}{N}\sum_{n=0}^{N-1} \|x_i(n)\|^2 \leqq S, \tag{3b}$$

(where $\|\cdot\|$ denotes Euclidean norm) and the $B_i$ are (measurable) subsets of $\mathcal{R}^{sN}$ with the following property. Let $\mathbf{y}_i = (\cdots, y_i^t(-1), y_i^t(0), y_i^t(1), \cdots)^t$ be the channel output vector which results when the channel input is $\mathbf{x}_i$, i.e.,

$$y_i(n) = \sum_{k=-\infty}^{\infty} h(n-k)x_i(k) + z(n)$$

$$= \sum_{k=0}^{N-1} h(n-k)x_i(k) + z(n).$$

Let $\mathbf{y}_i^{(N)} = [y_i^t(0), \cdots, y_i^t(N-1)]^t \in \mathcal{R}^{sN}$. Then $B_i (1 \leqq i \leqq M)$ must satisfy

$$P_{ei} \triangleq Pr\{\mathbf{y}_i^{(N)} \in B_i\} \leqq \lambda. \tag{4}$$

---

* Vectors will be taken to be column matrices unless otherwise indicated.

Thus the $\mathbf{x}_i$ are code words and $B_i$ is the set of output $N$-vectors which are decoded at the channel output as $\mathbf{x}_i$. Inequality (4) expresses the requirement that the error probability, given that $\mathbf{x}_i$ is transmitted, does not exceed $\lambda$.

A number $\rho \geqq 0$ is said to be "$S$-admissible" ($S \geqq 0$) if for all $\lambda > 0$ there is an $N$ such that there exists a code with parameters $([2^{N\rho}], N, S, \lambda)$. The *channel capacity* $C_S$ is defined as the supremum of $S$-admissible rates. Our problem is the calculation of $C_S$, which we shall solve provided the channel satisfies the following conditions:

(*i*) *The filter* $\{h(m)\}$: We assume that the filter is causal, i.e., $h(m) = 0,\ m < 0$. We also assume that

$$\sum_{m=0}^{\infty} \|h(m)\| < \infty, \tag{5a}$$

and that there exists a $B > 0$ such that for $m > 0$,

$$\|h(m)\| \leqq Bm^{-1}, \tag{5b}$$

where the Euclidean norm "$\|\cdot\|$" of a matrix is the square root of the sum of the squares of its entries. From (5), the (discrete) transfer function,

$$H(\theta) = \sum_{n=0}^{\infty} h(n)e^{in\theta}, \quad -\pi \leqq \theta \leqq \pi, \tag{6}$$

exists and is continuous. $H(\theta)$ is an $s \times s$ matrix. We assume that for $-\pi \leqq \theta \leqq \pi$, $\det H(\theta) \neq 0$.

(*ii*) *The noise covariance*: We assume that the covariance sequence $r(\cdot)$ satisfies

$$\sum_{n=-\infty}^{\infty} \|r(n)\| < \infty, \tag{7}$$

so that the (discrete) power spectral density

$$R(\theta) = \sum_{n=-\infty}^{\infty} r(n)e^{in\theta}, \quad -\pi \leqq \theta \leqq \pi \tag{8}$$

exists and is continuous. $R(\theta)$ is an $s \times s$ matrix. We also assume that for $-\pi \leqq \theta \leqq \pi$, $\det R(\theta) \neq 0$.

We can now give the capacity formula. Let the $s \times s$ matrix† $\Gamma(\theta) = H(\theta)^{-1}R(\theta)H(\theta)^{-*}$, and let $\lambda_1(\theta), \lambda_2(\theta), \cdots, \lambda_s(\theta)$ be the eigen-

† For any nonsingular complex matrix $A$, $A^{-*}$ is the transpose conjugate of $A^{-1}$.

values of $\Gamma(\theta)$, $-\pi \leqq \theta \leqq \pi$. Then $\lambda_j(\theta) > 0$, $1 \leqq j \leqq s$, $-\pi \leqq \theta \leqq \pi$. Let $S \geqq 0$ be given, and let $K_S$ be the (unique) positive number such that

$$\frac{1}{2\pi} \sum_{j=1}^{s} \int_{-\pi}^{\pi} d\theta \max[0, K_S - \lambda_j(\theta)] = S. \tag{9a}$$

Then

$$C_S = \frac{1}{4\pi} \sum_{j=1}^{s} \int_{-\pi}^{\pi} d\theta \max\left(0, \log_2 \frac{K_S}{\lambda_j(\theta)}\right). \tag{9b}$$

Our main result is the following. Consider the channel defined by (1) with $H = H_1(\theta)$ and $R = R_1(\theta)$. Then $C_S$ is calculated from (9) with $\Gamma(\theta) = H_1(\theta)^{-1}R_1(\theta)H_1(\theta)^{-*}$.

*Theorem 1a (Converse): Let $\rho \geqq 0$ be an S-admissible rate for this channel. Then $\rho \leqq C_S$.*

*Theorem 1b (Direct-Half): Let $S \geqq 0$, $\epsilon > 0$ and $\rho(0 \leqq \rho < C_S)$ be arbitrary. Then for N sufficiently large, there exists a code with parameters $(M, N, S, \lambda)$ where*

$$M \geqq e^{\rho N} \quad \textit{and} \quad \lambda \leqq \epsilon.$$

Sections II and III of this paper are concerned with the proof of Theorem 1. Section IV is concerned with the asymptotic behavior and numerical evaluation of the channel capacity formula (9), with specific attention to multipair telephone cable.

Theorem 1 is very similar to the results on continuous-time Gaussian channels due to Holsinger and Gallager.[1] In fact, for the special case in which $H(\theta)$ is the $s \times s$ identity, the theorem follows immediately from the analysis in Ref. 1. We suspect that it might be possible to obtain all of Theorem 1 by paralleling Gallager's techniques for this discrete case, although such an approach is somewhat more cumbersome than the approach followed here. Furthermore, the present approach lends itself immediately to broadening the model to consider the effects of intersymbol interference from previous channel uses, as Gallager's approach does not.[2] In fact, to establish Theorem 1b for the intersymbol interference channel we require only to add in one of our lemmas a term "$\gamma_3$" (representing the effect of previous channel uses), and to show that its norm $|||\gamma_3||| = o(N^{\frac{1}{2}})$.

Careful analysis of the proof of Theorem 1 will indicate that the conditions on the filter $\{h(m)\}$ given in Section I can be replaced by simply requiring that the filter have a causal inverse $\{g(m)\}$, such that $g(m) = 0$, $m \geqq m_0$ (i.e., finite memory). Thus, our results contain a generalization of those given by Toms and Berger.[3]

## II. NOTATION AND MATHEMATICAL PRELIMINARIES

Let $l_2^{(s)}(a, b)$, $s = 1, 2, \cdots$, $-\infty \leqq a < b \leqq \infty$, where $a$, $b$ are integers, be the set of sequences $\{x(n)\}_{n=a}^{b}$ where $x(n)$ is a real $s$-vector. Such sequences will often be written as column matrices $\mathbf{x} = [x^t(a), x^t(a+1), \cdots, x^t(b)]^t$. Let $\|\cdot\|$ denote ordinary Euclidean norm in $s$-space, and for $s \times s$ matrices $A$, let $\|A\|$ be the Euclidean norm (i.e., the square root of the sum of the squares of the components of $A$). For sequences $\mathbf{x} \in l_2^{(s)}(a, b)$, the (Euclidean) norm is

$$|||\mathbf{x}||| = \left[\sum_{n=a}^{b} \|x(n)\|^2\right]^{\frac{1}{2}}. \tag{10}$$

The space $l_2^{(s)}(a, b)$ is a Hilbert space with the obvious inner product, written $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=a}^{b} x^t(n)y(n)$, $\mathbf{x},\mathbf{y} \in l_2^{(s)}(a, b)$. For $\mathbf{x} \in l_2^{(s)}(-\infty, \infty)$, denote by $\mathbf{x}^{(N)} \in l_2^{(s)}(0, N-1)$ the column matrix

$$\mathbf{x}^{(N)} = [x^t(0), x^t(1), \cdots, x^t(N-1)]^t. \tag{11}$$

We denote operators on $l_2^{(s)}(a, b)$ by script letters, e.g., $\mathfrak{F}$. We define the norm of $\mathfrak{F}$ by

$$|\mathfrak{F}| = \sup_{\mathbf{x} \in l_2^{(s)}(a,b)} \frac{|||\mathfrak{F}\mathbf{x}|||}{|||\mathbf{x}|||}. \tag{12}$$

If $|\mathfrak{F}| < \infty$, we say that $\mathfrak{F}$ is bounded. An operator $\mathfrak{F}$ is said to be of a convolution type if, for $\mathbf{x} \in l_2^{(s)}(a, b)$,

$$(\mathfrak{F}\mathbf{x})(n) = \sum_{k=a}^{b} f(n-k)x(k), \quad n = a, \cdots, b, \tag{13}$$

where $\{f(n)\}_{a-b}^{b-a}$ is a fixed sequence of $s \times s$ matrices and the indicated operations are ordinary matrix arithmetic. Let $L$ be the set of convolution-type operators on $l_2^{(s)}(-\infty, \infty)$ for which

$$\sum_{n=-\infty}^{\infty} \|f(n)\| < \infty. \tag{14}$$

For operators $\mathfrak{F}$ in $L$ the transfer matrix

$$F(\theta) = \sum_{n=-\infty}^{\infty} f(n)e^{in\theta}, \quad -\pi \leqq \theta \leqq \pi \tag{15}$$

is well defined. $F(\theta)$ is an $s \times s$ matrix and is continuous (in Euclidean norm) for $-\pi \leqq \theta \leqq \pi$. Concatenation of operators $\mathfrak{F}_1$, $\mathfrak{F}_2 \in L$, defined by sequences $\{f_1(n)\}$ and $\{f_2(n)\}$, results in a convolution-type operator $\mathfrak{F}_3 = \mathfrak{F}_2 \cdot \mathfrak{F}_1$ in $L$ defined by the sequence $\{f_3(n)\}$ where

$f_3(n) = \sum_{k=-\infty}^{\infty} f_2(k) f_1(n-k)$. Further, the corresponding transfer functions satisfy $F_3(\theta) = F_2(\theta)F_1(\theta)$. Other relevant properties of the class $L$ are given in the theorem below, the assertions of which are generalizations of well-known scalar results. The proof is given in the appendix (Section A.1).

*Theorem 2: Given* $\mathfrak{F} \in L$, *defined by* $\{f(n)\}$ or $F(\theta)$, *then*

*a.* $|\mathfrak{F}| \leqq \sum_{n=-\infty}^{\infty} \|f(n)\| < \infty$, *so that* $\mathfrak{F}$ *is bounded on* $l_2^{(s)}(-\infty, \infty)$.

*b.* $|\mathfrak{F}| \leqq \max_{-\pi \leqq \theta \leqq \pi} \|F(\theta)\|$.

*c. If* $\mathfrak{F}$ *is self-adjoint, then for all* $\mathbf{x} \in l_2^{(s)}(-\infty, \infty)$,

$$\left| \sum_{n,m} x^t(n) f(n-m) x(m) \right| \leqq \left( \max_{-\pi \leqq \theta \leqq \pi} \|F(\theta)\| \right) |||\mathbf{x}|||^2.$$

*d.* $\mathfrak{F}$ *has a bounded inverse denoted* $\mathfrak{F}^{-1}$ *if and only if det* $F(\theta) \neq 0$, $-\pi \leqq \theta \leqq \pi$, *in which case the transfer matrix corresponding to* $\mathfrak{F}^{-1}$ *is* $[F(\theta)]^{-1}$, *and* $\mathfrak{F}^{-1} \in L$.

Let $\mathbf{z} = [\cdots, z^t(-1), z^t(0), z^t(1), \cdots,]^t$ be a sequence of random $s$-vectors with covariance

$$Ez(n)z(n-m)^t = r(m),$$

where $r(m)$ is an $s \times s$ matrix. Under the assumption that

$$\sum_{m=-\infty}^{\infty} \|r(m)\| < \infty, \tag{16}$$

the power spectrum

$$R(\theta) = \sum_{m=-\infty}^{\infty} r(m) e^{im\theta}, \quad -\pi \leqq \theta \leqq \pi, \tag{17}$$

is well defined. Let $\mathfrak{F}$ be an operator in $L$ corresponding to the transfer matrix $F(\theta)$. Then $\hat{\mathbf{z}} = \mathfrak{F}\mathbf{z}$ is a sequence of random $s$-vectors with covariance $\{\hat{r}(m)\}$ and corresponding power spectrum $\hat{R}(\theta) = F(\theta)\, R(\theta) F(\theta)^*$.

Let $\mathbf{z}$ be a sequence of zero mean Gaussian $n$-vectors with covariance $r(m)$ satisfying (16) and power spectrum $R(\theta)$. Let $\lambda_{\min}(\theta)$ be the minimum eigenvalue of the matrix $R(\theta)$. Let $\mathbf{z}^{(N)} = [z^t(0), z^t(1), \cdots, z^t(N-1)]^t$ be a segment of $\mathbf{z}$ of duration $N$. Then there exists an $(N \cdot s) \times (N \cdot s)$ matrix $T_N$ such that

$$\mathbf{w} = T_N \mathbf{z}^{(N)}, \tag{18}$$

is "white," i.e., $E\mathbf{w}\mathbf{w}^t = I_{N \cdot s}$. The indicated operation in (18) is matrix multiplication. The only property of $T_N$ which we need here is that for any $N \cdot s$-vector $\mathbf{u}$

$$|||T_N \mathbf{u}||| \leqq \left[ \frac{1}{\min_{-\pi \leqq \theta \leqq \pi} \lambda_{\min}(\theta)} \right]^{\frac{1}{2}} |||\mathbf{u}|||. \tag{19}$$

Finally, let $\mathcal{H}$ be the convolution-type operator on $l_2^{(s)}(-\infty, \infty)$ defined by $\{h(m)\}$ (Section I). Note that by (5) $\mathcal{H} \in L$, so that by the assumption following (6) and by Theorem 2d it has an inverse, say $\mathcal{G} = \mathcal{H}^{-1}$, of the convolution type in $L$. Let $\{g(n)\}_{-\infty}^{\infty}$ be the sequence which defines $\mathcal{G}$. Let the operator $\mathcal{G}_N (N = 0, 1, 2, \cdots)$ be defined by

$$(\mathcal{G}_N \mathbf{x})(n) = \sum_{k=0}^{N-1} g(n-k) x(k), \qquad -\infty < n < \infty. \tag{20}$$

We conclude this section by stating as lemmas two known results. We explain in the appendix (Section A.2) how to obtain these results from published material.

*Lemma 3: For the special case when $H(\theta) = I_s$ (the $s \times s$ identity) and $R(\theta) = R_2(\theta)$ [so that $\Gamma = H^{-1}RH^{-*} = R_2(\theta)$], say that $\mathbf{x}$ is a random channel input sequence for which $x(n) = 0$, $n < 0$, $n \geqq N$, and $E|||\mathbf{x}|||^2 \leqq NS (S > 0,\ N = 0, 1, 2, \cdots)$. Let $\mathbf{y}$ be the corresponding channel output sequence. Then, the mutual information*

$$I\{\mathbf{x}, \mathbf{y}\} \leqq NC_S$$

*(where $C_S$ is calculated with $\Gamma(\theta) = R_2(\theta)$).*

*Lemma 4: For the special case where $H(\theta) = I_s$ and $R(\theta) = R_2(\theta)$, then we have a stronger version of Theorem 1b: Let $S \geqq 0$ and $\rho$ $(0 \leqq \rho < C_S)$ be arbitrary, where $C_S$ is calculated with $\Gamma(\theta) = R_2(\theta)$. Then, for $N = 1, 2, \cdots$, there exists a code with parameters $(M, N, S, \lambda)$ where*

$$M \geqq 2^{\rho N} \quad \textit{and} \quad \lambda \leqq Ae^{-BN}, \quad A, B > 0.$$

## III. PROOF OF THEOREM 1

### 3.1 Converse

Let $\{(\mathbf{x}_i, B_i)\}_1^M$ be a code with parameters $(M, N, S, \lambda)$ for the channel of (1) with $H = H_1(\theta)$ and $R = R_1(\theta)$. Let $\mathbf{x}$ be the random sequence which results when $\mathbf{x} = \mathbf{x}_i$ with probability $1/M$ $(1 \leqq i \leqq M)$. Let $\mathbf{y} = \mathcal{H}\mathbf{x} + \mathbf{z}$ be the corresponding output sequence, and $\mathbf{y}^{(N)} = (y(0)^t, \cdots, y(N-1)^t)^t$. The theorem will follow in the standard way from the Fano inequality (see, for example, Ref. 1) if we can show that

$$I\{\mathbf{x}, \mathbf{y}^{(N)}\} \leqq NC_S. \tag{21}$$

But

$$I\{\mathbf{x}, \mathbf{y}^{(N)}\} \leqq I\{\mathbf{x}, \mathbf{y}\} = I\{\mathbf{x}, \mathcal{H}^{-1}\mathbf{y}\} = I\{\mathbf{x}, \mathbf{x} + \hat{\mathbf{z}}\}, \tag{22}$$

where $\hat{\mathbf{z}} = \mathcal{H}^{-1}\mathbf{z}$ is a stationary Gaussian random process with power spectrum $\Gamma(\theta) = H_1^{-1}(\theta)R_1(\theta)H_1^{-*}(\theta)$. Thus, we can apply Lemma 3 (since $\mathbf{x}$ satisfies the required hypotheses) with $R_2(\theta) = \Gamma(\theta)$ to obtain (21). Hence, the theorem follows.

### 3.2 Direct-half

Consider again the special case of our channel where $H(\theta) = I_s$ and $R(\theta) = R_2(\theta)$. The idea behind the proof is to construct codes for the general case ($H$, $R$ arbitrary) by modifying codes (whose existence is guaranteed by Lemma 4) which are known to be good for this special case. We proceed as follows. Let $\{(\mathbf{x}_i, B_i)\}_1^M$ be a code for this channel with parameters $N = N_1$ and $S = S_1$. Then, for $1 \leqq i \leqq M$, we have

$$\mathbf{y}_i^{(N)} = \mathbf{x}_i^{(N)} + \mathbf{z}^{(N)},$$

where the superscript operation is defined by (11). Let $T_N$ be the whitening filter (discussed in Section II) for which $T_N\mathbf{z}^{(N)} = \mathbf{w}$ and $E\mathbf{w}\mathbf{w}^t = I_{Ns}$. Letting $\mathbf{v}_i = T_N\mathbf{y}_i^{(N)}$ and $\mathbf{u}_i = T_N\mathbf{x}_i^{(N)}$, we have

$$\mathbf{v}_i = \mathbf{u}_i + \mathbf{w}. \tag{23}$$

Let us assume that the $\{B_i\}$ correspond to the minimum distance decoder, i.e., $\mathbf{y}^{(N)} \in B_i$ if $|||\mathbf{v} - \mathbf{u}_i||| < |||\mathbf{v} - \mathbf{u}_j|||$ for all $j \neq i$, where $\mathbf{v} = T_N\mathbf{y}^{(N)}$. Then

$$\begin{aligned} P_{ei} = Pr\{\mathbf{y}_i^{(N)} \notin B_i\} &= Pr \bigcup_{j\neq i} \{|||\mathbf{v}_i - \mathbf{u}_i||| \geqq |||\mathbf{v}_i - \mathbf{u}_j|||\} \\ &= Pr \bigcup_{j\neq i} \{|||\mathbf{w}||| > |||\mathbf{w} - (\mathbf{u}_j - \mathbf{u}_i)|||\} \\ &= Pr \bigcup_{j\neq i} \{\langle \mathbf{w}, \mathbf{u}_j - \mathbf{u}_i\rangle \geqq \tfrac{1}{2}|||\mathbf{u}_j - \mathbf{u}_i|||^2\}. \end{aligned} \tag{24}$$

Thus, in particular, for all $j \neq i$,

$$\begin{aligned} P_{ei} &\geqq Pr\{\langle \mathbf{w}, \mathbf{u}_j - \mathbf{u}_i\rangle \geqq \tfrac{1}{2}|||\mathbf{u}_j - \mathbf{u}_i|||^2\} \\ &= \Phi_c(\tfrac{1}{2}|||\mathbf{u}_j - \mathbf{u}_i|||) = \Phi_c[\tfrac{1}{2}|||T_N(\mathbf{x}_j^{(N)} - \mathbf{x}_i^{(N)})|||], \end{aligned} \tag{25}$$

where

$$\Phi_c(u) = \frac{1}{\sqrt{2\pi}} \int_u^\infty e^{-u^2/2} du$$

is the complementary error function.

Let $H_1(\theta)$ and $R_1(\theta)$ be arbitrary, and suppose that we are given a code $\{(\mathbf{x}_i, B_i)\}_1^M$ with parameters $(M, N, S, \lambda_1)$ for use on the special channel with $H = I_s$ and $R(\theta) = R_2 = \Gamma(\theta) = H_1^{-1}R_1H_1^{-*}$. Assume that the $B_i$ corresponds to the minimum distance decoder so that $P_{ei}$ is given by (24). We now construct a new code $\{(\mathbf{x}_i^*, B_i^*)\}_{i=1}^M$ with parameters $N = N_2 = (1 + \delta)N_1$ and $S = S_2 = \alpha^2 S_1/(1 + \delta)$ for use on the general channel with $H_1(\theta)$, $R_1(\theta)$ arbitrary. We set

$$x_i^*(n) = \begin{cases} \alpha x_i(n), & 0 \leqq n \leqq N_1 - 1, \\ 0, & \text{otherwise}, \end{cases} \tag{26}$$

where $\alpha > 1$ and $\delta > 0$ are arbitrary. Note that we have allowed a guard-band or dead-space or width $\delta N_1$ following the channel input signal. The decoding sets $B_i^*$ $(1 \leqq i \leqq M)$ are described below.

The channel output is as in (1)

$$\mathbf{y} = \mathcal{H}\mathbf{x} + \mathbf{z},$$

where $\mathbf{x}$ is the channel input, $\mathbf{z}$ is the noise, and $\mathcal{H}$ is the operator corresponding to $H_1(\theta)$. Let $\mathcal{G}_{N_2}$ be the operator defined in (20), and let

$$\begin{aligned} \hat{\mathbf{y}} = \mathcal{G}_{N_2}\mathbf{y} &= \mathcal{G}_{N_2}\mathcal{H}\mathbf{x} + \mathcal{G}_{N_2}\mathbf{z} \\ &= \mathbf{x} + \hat{\mathbf{z}} + \boldsymbol{\xi}_1 + \boldsymbol{\xi}_2, \end{aligned}$$

where $\boldsymbol{\xi}_1 = \mathcal{G}_{N_2}\mathcal{H}\mathbf{x} - \mathbf{x}$, $\hat{\mathbf{z}} = \mathcal{G}\mathbf{z}$, and $\boldsymbol{\xi}_2 = \mathcal{G}_{N_2}\mathbf{z} - \hat{\mathbf{z}}$. Let us note that $\hat{\mathbf{y}}$ is calculable from $\mathbf{y}^{(N_2)} = (y^t(0), \cdots, y^t(N_2 - 1))^t$. Further, the noise $\hat{\mathbf{z}}$ has power spectrum $\Gamma(\theta) = H_1^{-1}(\theta)R_1(\theta)H_1^{-*}(\theta)$. (In fact, if $\boldsymbol{\xi}_1 = \boldsymbol{\xi}_2 = 0$, the channel would be equivalent to the special case, and the direct-half of the coding theorem would follow from Lemma 4. Although this is not the case, of course, we will show that $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are sufficiently small so that Lemma 4 can be applied anyway.) The decoding sets $B_i^*$ are defined by: $\mathbf{y}^{(N_2)} \in B_i^*$ if $\hat{\mathbf{y}}^{(N_1)} \in B_i$, $1 \leqq i \leqq M$.

Letting $\mathbf{y}_i^* = \mathbf{x}_i^* + \hat{\mathbf{z}} + \boldsymbol{\xi}_1 + \boldsymbol{\xi}_2$ $(1 \leqq i \leqq M)$, and letting $T_{N_1}$ be as above, we define

$$\begin{aligned} \mathbf{v}_i^* \triangleq T_{N_1}\mathbf{y}_i^{*(N_1)} &= T_{N_1}\mathbf{x}_i^{*(N_1)} + T_{N_1}\hat{\mathbf{z}}^{(N_1)} + T_{N_1}\boldsymbol{\xi}_1^{(N_1)} + T_{N_1}\boldsymbol{\xi}_2^{(N_1)} \\ &= \alpha\mathbf{u}_i + \mathbf{w} + \boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2, \end{aligned} \tag{27}$$

where $\mathbf{u}_i$ and $\mathbf{w}$ are exactly as in (23) and $\boldsymbol{\gamma}_i = T_{N_1}\boldsymbol{\xi}_i^{(N_1)}$ $(i = 1, 2)$. The decoder for the derived code is the minimum distance decoder for the $\mathbf{v}^*$'s. Now, following the same steps as in (24), we have

$$\begin{aligned} P_{ei}^* \triangleq Pr\{\mathbf{y}_i^{(N_2)} \notin B_i^*\} &= Pr \bigcup_{j \neq i} \{|||\mathbf{v}_i^* - \alpha\mathbf{u}_i||| \geqq |||\mathbf{v}_i^* - \alpha\mathbf{u}_j|||\} \\ &= Pr \bigcup_{j \neq i} \{\langle \mathbf{w} + \boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2, \mathbf{u}_j - \mathbf{u}_i \rangle \geqq \frac{\alpha}{2} |||\mathbf{u}_j - \mathbf{u}_i|||^2\}. \end{aligned} \tag{28}$$

Now, according to (9), the channel of Section I [with arbitrary $H_1(\theta)$ and $R_1(\theta)$] and the special channel with $H(\theta) = I_s$ and $R(\theta) = H_1(\theta)^{-1}R_1(\theta)H_1(\theta)^{-*} = \Gamma(\theta)$ have the same $C_S$. Let $\epsilon$, $S > 0$ and $\rho$ ($0 \leqq \rho < C_S$) given, and let $\{(\mathbf{x}_i, B_i)\}_1^M$ be a code with parameters $(M, N_1, S_1, \lambda_1)$ (as guaranteed by Lemma 4) such that

$$M \geqq 2^{\rho N_1} \quad \text{and} \quad \lambda_1 \leqq \epsilon.$$

We will show that with $N_1$ sufficiently large, the derived code has parameter $\lambda \leqq \lambda_1 + \epsilon$. Thus, we will have found a set of codes with parameters $(M, N_2, S_2, \lambda)$ with

$$S_2 = \frac{\alpha^2 S_1}{(1+\delta)}, \quad M \geqq \exp_2\left\{\frac{\rho N_2}{1+\delta}\right\}$$

and

$$\lambda \leqq 2\epsilon.$$

Since $C_S$ is continuous in its arguments, and $\alpha$ may be chosen arbitrarily close to 1, and $\delta$ arbitrarily close to 0, the direct-half of the coding theorem (Theorem 1b) will have been established.

To show that $\lambda$ for the derived code $\leqq \lambda_1 + \epsilon$, we must show that for each $i = 1, 2, \cdots$,

$$Pr\{\mathbf{y}_i^{*(N_2)} \notin B_i^*\} \leqq Pr\{\mathbf{y}_i^{(N_1)} \notin B_i\} + \epsilon. \tag{29}$$

Inequality (29) will follow directly from the following lemmas, the proofs of which are given at the end of this section.

*Lemma 5: Inequality* (29) *is satisfied if*

$$Pr\{|||\boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2||| \leqq \frac{(\alpha - 1)}{2} \min_{i \neq j} |||\mathbf{u}_i - \mathbf{u}_j|||\} \geqq 1 - \epsilon. \tag{30}$$

*Lemma 6: For the codes* $\{(\mathbf{x}_i, \mathbf{B}_i)\}_1^M$, *as* $N \to \infty$,

$$\min_{i \neq j} |||\mathbf{u}_i - \mathbf{u}_j|||^2 \geqq 0(N_1).$$

*Lemma 7: For arbitrary* $a > 0$,

$$Pr\{|||\boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2|||^2 \leqq aN_1\} \to 1, \quad \textit{as} \quad N_1 \to \infty.$$

Now, from Lemmas 6 and 7, condition (30) in Lemma 4 will be satisfied for $N_1$ sufficiently large. This establishes Theorem 1b.

*Proof of Lemma 5:* Let

$$S = \left\{|||\boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2||| \leqq \frac{(\alpha - 1)}{2} \min_{i \neq j} |||\mathbf{u}_i - \mathbf{u}_j|||\right\}. \tag{31}$$

By hypothesis, $Pr\{S\} \geqq 1 - \epsilon$. Since $B_i$ and $B_i^*$ correspond to the minimum distance decoder, we have from (28)

$$Pr\{\mathbf{y}_i^{*(N_2)} \notin B_i^*\} \leqq Pr\{S \cap \{\mathbf{y}_i^{*(N_2)} \notin B_i^*\}\} + Pr\{S^c\}$$

$$\leqq Pr \bigcup_{j \neq i} S \cap \left\{\langle \mathbf{w} + \gamma_1 + \gamma_2, \mathbf{u}_j - \mathbf{u}_i \rangle \geqq \frac{\alpha}{2} |||\mathbf{u}_j - \mathbf{u}_i|||^2 \right\} + \epsilon. \quad (32)$$

Now if $S$ occurs,

$$|\langle \gamma_1 + \gamma_2, \mathbf{u}_j - \mathbf{u}_i \rangle| \leqq |||\gamma_1 + \gamma_2||| \cdot |||\mathbf{u}_j - \mathbf{u}_i|||$$
$$\leqq \left(\frac{\alpha - 1}{2}\right) |||\mathbf{u}_j - \mathbf{u}_i|||^2.$$

Thus, the event in the right member of (32) satisfies

$$S \cap \left\{\langle \mathbf{w} + \gamma_1 + \gamma_2, \mathbf{u}_j - \mathbf{u}_i \rangle \geqq \frac{\alpha}{2} |||\mathbf{u}_i - \mathbf{u}_i|||^2 \right\}$$
$$\subseteq \left\{\langle \mathbf{w}, \mathbf{u}_j - \mathbf{u}_i \rangle \geqq \frac{\alpha}{2} |||\mathbf{u}_j - \mathbf{u}_i|||^2 - \left(\frac{\alpha - 1}{2}\right) |||\mathbf{u}_j - \mathbf{u}_i|||^2 \right\}$$
$$\subseteq \{\langle \mathbf{w}, \mathbf{u}_j - \mathbf{u}_i \rangle \geqq \tfrac{1}{2} |||\mathbf{u}_j - \mathbf{u}_i|||^2\},$$

and (32) becomes

$$Pr\{\mathbf{y}_i^{*(N_2)} \notin B_i^*\} \leqq Pr \bigcup_{j \neq i} \{\langle \mathbf{w}, \mathbf{u}_j - \mathbf{u}_i \rangle \geqq \tfrac{1}{2} |||\mathbf{u}_j - \mathbf{u}_i|||^2\} + \epsilon$$
$$= Pr\{\mathbf{y}_i^{(N_1)} \notin B_i\} + \epsilon,$$

where the last equality follows from (24). This is (29) so that we have proved Lemma 5.

*Proof of Lemma 6:* For the codes $\{(\mathbf{x}_i, B_i)\}_1^M$, and $N_1 \to \infty$,

$$P_{ei} \leqq Ae^{-BN_1},$$

so that from (25),

$$\Phi_c(\tfrac{1}{2} |||\mathbf{u}_j - \mathbf{u}_i|||) \leqq Ae^{-BN_1}.$$

Since, as $\eta \to \infty$, $\Phi_c(\eta) = e^{-(\eta^2/2)[1+o(1)]}$, we have

$$|||\mathbf{u}_j - \mathbf{u}_i|||^2 \geqq 8BN_1[1 + o(1)],$$

which implies Lemma 6.

*Proof of Lemma 7*: First note that by (19)

$$|||\gamma_1 + \gamma_2|||$$
$$= |||T_{N_1}(\xi_1^{(N_1)} + \xi_2^{(N_1)})||| \leqq \left[\frac{1}{\min \lambda_{\min}(\theta)}\right]^{\frac{1}{2}} |||\xi_1^{(N_1)} + \xi_2^{(N_1)}|||$$
$$\leqq \left[\frac{1}{\min \lambda_{\min}(\theta)}\right]^{\frac{1}{2}} [|||\xi_1^{(N_1)}||| + |||\xi_2^{(N_1)}|||].$$

Thus, it will suffice to show first, $|||\xi_1^{(N_1)}||| = o(N_1^{\frac{1}{2}})$ as $N_1 \to \infty$, and second, for arbitrary $a > 0$, $Pr\{|||\xi_2^{(N_1)}|||^2 \geqq aN_1\} \to 0$, as $N_1 \to \infty$.

1. Let $\mathbf{x}$ be one of the code vectors in $\{(\mathbf{x}_i^*, B_i^*)\}$. Then $\xi_1 = \mathcal{G}_{N_2}\mathcal{H}\mathbf{x} - \mathbf{x}$, so that for $-\infty < n < \infty$,

$$\xi_1(n) = \sum_{\substack{k<0 \\ k \geqq N_2}} g(n-k)(\mathcal{H}\mathbf{x})(k). \tag{33}$$

Now, since $\mathbf{x}$ is one of the code vectors $\{\mathbf{x}_i^*\}_1^M$, $x(n) = 0$ for $n \notin [0, N_1 - 1]$. Also since $\mathcal{H}$ is causal, we have $(\mathcal{H}\mathbf{x})(k) = 0$, $k < 0$, and

$$\xi_1(n) = \sum_{k=N_2}^{\infty} g(n-k)(\mathcal{H}\mathbf{x})(k)$$

$$= \sum_{k=N_2}^{\infty} g(n-k) \sum_{j=0}^{N_1-1} h(k-j)x(j). \tag{34}$$

Next, define the sequence $\psi$ by

$$\psi(k) = \begin{cases} \sum_{j=0}^{N_1-1} h(k-j)x(j), & k \geqq N_2. \\ 0, & k < N_2 \end{cases}$$

Then (34) is

$$\xi_1(n) = \sum_{k=-\infty}^{\infty} g(n-k)\psi(k),$$

i.e., $\xi_1 = \mathcal{G}\psi$, and

$$|||\xi_1||| \leqq |\mathcal{G}| \cdot |||\psi|||. \tag{35}$$

Now $|\mathcal{G}| \leqq \sum_{n=-\infty}^{\infty} \|g(n)\| < \infty$ from Theorem 2d, and

$$|||\psi|||^2 = \sum_{k=N_2}^{\infty} \|\psi(k)\|^2 = \sum_{k=N_2}^{\infty} \left\| \sum_{j=0}^{N_1-1} h(k-j)x(j) \right\|^2$$

$$\leqq \sum_{k=N_2}^{\infty} \left( \sum_{j=0}^{N_1-1} \|h(k-j)\|^2 \right) \left( \sum_{j=0}^{N_1-1} \|x(j)\|^2 \right),$$

where in the final step we have used the following form of the Schwarz inequality: if $\mathbf{a}$ is a sequence of $s \times s$ matrices, and $\mathbf{x}$ is a sequence of $s$-vectors, then

$$\|\sum_n a(n)x(n)\|^2 \leqq \sum_n \|a(n)\|^2 \sum_n \|x(n)\|^2.$$

But since $\sum_{j=0}^{N_1-1} \|x^2(j)\|^2 = |||\mathbf{x}|||^2 \leqq \alpha^2 S_1 N_1$, we have

$$|||\boldsymbol{\psi}|||^2 \leqq \alpha^2 S_1 N_1 \sum_{k=N_2}^{\infty} \sum_{j=0}^{N_1-1} \|h(k-j)\|^2. \tag{36}$$

We will now show that, as $N_1 \to \infty$,

$$\sum_{k=N_2}^{\infty} \sum_{j=0}^{N_1-1} \|h(k-j)\|^2 = \sum_{k=N_2}^{\infty} \sum_{j=0}^{N_1-1} f^2(k-j) \to 0, \tag{37}$$

where $f(k) = \|h(k)\|$. Expressions (35), (36), and (37) together imply that

$$\|\boldsymbol{\xi}_1^{(N_1)}\|^2 \leqq |||\boldsymbol{\xi}_1|||^2 = o(N_1), \quad \text{as} \quad N_1 \to \infty,$$

which is what we set out to establish. It remains to establish (37).

Now, from (5), $\sum_0^\infty f(k) < \infty$, and $f(k) \leqq B/k$. Setting

$$F(k) = \sum_{j=k}^{\infty} f^2(j),$$

we have

$$Q \triangleq \sum_{k=N_2}^{\infty} \sum_{j=0}^{N_1-1} f^2(k-j)$$

$$= \sum_{k=N_2}^{\infty} \sum_{i=k-N_1+1}^{k} f^2(i) = \sum_{k=N_2}^{\infty} [F(k-N_1+1) - F(k)]$$

$$= \sum_{k=N_2-N_1+1}^{N_2-1} F(k) \leqq \sum_{\delta N_1+1}^{N_2} F(k).$$

Now*

$$\sum_{\delta N_1+1}^{N_2} F(k) = kF(k)\Big|_{\delta N_1}^{N_2} + \sum_{\delta N_1+1}^{N_2} (k-1) f^2(k-1).$$

But

$$kF(k)\Big|_{\delta N_1}^{N_2} \leqq N_2 F(\delta N_1) = N_2 \sum_{\delta N_1}^{\infty} f^2(k) \leqq \frac{(1+\delta)}{\delta} \sum_{\delta N_1}^{\infty} k f^2(k),$$

---

* We have made use of the formula (summation by parts)

$$\sum_a^b v(k)\Delta u(k) = v(k)u(k)\Big|_{a-1}^{b} - \sum_a^b u(k-1)\Delta v(k),$$

where $\Delta u(k) = u(k) - u(k-1)$. Here $v(k) = F(k)$ and $u(k) = k$.

and

$$\sum_{\delta N_1+1}^{N_2} (k-1)f^2(k-1) \leqq \sum_{\delta N_1}^{\infty} kf^2(k).$$

Thus,

$$Q \leqq \left(\frac{1+\delta}{\delta}+1\right)\sum_{\delta N_1}^{\infty} kf^2(k) \leqq \left(\frac{1+2\delta}{\delta}\right)B\sum_{\delta N_1}^{\infty} f(k) \to 0,$$

since $\sum_0^\infty f(k) < \infty$. Thus, (37) is established and we have finished the first part.

2. Since for any $a > 0$,

$$Pr\{|||\xi_2^{(N_1)}|||^2 \geqq aN_1\} \leqq \frac{E|||\xi_2^{(N_1)}|||^2}{aN_1},$$

and since $N_2 = (1+\delta)N_1$, it suffices to show that

$$E|||\xi_2^{(N_2)}|||^2 = o(N_2), \quad \text{as} \quad N_2 \to \infty.$$

Now $\boldsymbol{\xi}_2 = \mathcal{G}_{N_2}\mathbf{z} - \hat{\mathbf{z}}$, where $\hat{\mathbf{z}} = \mathcal{H}^{-1}\mathbf{z}$. Hence,

$$E[\xi_2(n)\xi_2^t(n)] = \sum_{\substack{i,j<0\\ i,j\geqq N_2}} g(n-i)r(i-j)g^t(n-j).$$

Let $\beta$ denote any fixed $s$-vector and define a sequence $\boldsymbol{\psi}$ of $s$-vectors by

$$\psi(n-i) = \begin{cases} g^t(n-i)\beta, & i<0 \quad \text{and} \quad i \geqq N_2. \\ 0, & 0 \leqq i < N_2. \end{cases}$$

Then

$$\beta^t\, E[\xi_2(n)\xi_2^t(n)]\,\beta = \sum_{i,j=\infty}^{\infty} \psi^t(n-i)r(i-j)\psi(n-j).$$

Since $r(\cdot)$ is a covariance, $r(k) = r^t(-k)$ for all $k$. Application of Theorem 2c shows that the double summation above is bounded by

$$\max_{-\pi\leqq\theta\leqq\pi} \|R(\theta)\| \sum_{i=-\infty}^{\infty} \|\psi(n-i)\|^2.$$

Since

$$|||\xi_2^{(N_2)}|||^2 = \sum_{n=0}^{N_2-1} \xi_2^t(n)\xi_2(n) = \sum_{n=0}^{N_2-1}\sum_{\nu=1}^{s} e_\nu^t\xi_2(n)\xi_2^t(n)e_\nu,$$

where $e_\nu$ is the $s$-vector with $j$th entry $\delta_{\nu j}$ ($\nu$, $j = 1, \cdots, s$), the desired result will follow if we show that

$$\frac{1}{N_2}\beta^t \sum_{n=0}^{N_2-1} E[\xi_2(n)\xi_2^t(n)]\beta = o(1)$$

for any $\beta$. Thus, it suffices to show that

$$\frac{1}{N_2}\sum_{n=0}^{N_2-1}\sum_{i=-\infty}^{\infty} \|\psi(n-i)\|^2 = o(1).$$

But from the definition of $\psi$, the last expression can be rewritten as

$$\frac{1}{N_2}\sum_{n=0}^{N_2-1}\sum_{k=n+1}^{\infty} [\|\psi(k)\|^2+\|\psi(-k)\|^2].$$

Since $\mathcal{G}$ is in $L$, Theorem $2a$ implies $\psi$ is in $l_2^{(s)}$ $(-\infty, \infty)$. In particular, as $n \to \infty$

$$\sum_{k=n+1}^{\infty} [\|\psi(k)\|^2+\|\psi(-k)\|^2] = o(1)$$

and the desired conclusion follows immediately.

## IV. ASYMPTOTIC BEHAVIOR AND NUMERICAL EVALUATION OF THE CAPACITY FORMULA

In this final section, we discuss some implications of the channel capacity formula given in (9). As in the prior sections, the channel is assumed to be a multi-input, multi-output channel with memory, and with additive Gaussian noise. But in contrast to the previous case, instead of a discrete time channel, we consider an equivalent continuous time bandlimited channel.

Specifically, the channel inputs or code words (in a $T$ second block coding interval) are vector-valued functions $x(\cdot)$ of dimensions $s$, bandlimited in frequency interval $[-W, W]$ such that the samples of $x(\cdot)$ satisfy

$$x\left(\frac{n}{2W}\right) = 0, \quad \text{for} \quad \frac{n}{2W} < 0 \quad \text{or} \quad \frac{n}{2W} \geqq T.$$

We also have the average power constraint

$$\int_{-\infty}^{\infty} \|x(t)\|^2 dt \leqq ST,$$

where $\|\cdot\|$ denotes Euclidean norm.

The channel has $s$ inputs and $s$ outputs and has transfer function matrix $H(f)$ for $f \in [-W, W]$. The additive noise is also vector-valued and has power spectral density matrix $R(f)$. Let $\{\lambda_i(f)\}$ denote the set of eigenvalues of $H^{-1}(f)R(f)[H^{-1}(f)]^*$.

The capacity of this channel is determined as follows. Let $S > 0$ be given and let $K$ be the unique positive number satisfying

$$\sum_{i=1}^{s} \int_{-W}^{W} \max[0, K - \lambda_i(f)]df = S. \tag{38a}$$

Then

$$C = \frac{1}{2} \sum_{i=1}^{s} \int_{-W}^{W} \max\left(0, \log_2 \frac{K}{\lambda_i(f)}\right) df \tag{38b}$$

with $C$ in bits per second.

Formula (38) can be obtained from the analogous formula (9) via application of the sampling theorem. The somewhat tedious derivation is carried out for the scalar case ($s = 1$) in the appendix (Section A.3).

We consider several implications of (38). Specifically, for large signal-to-noise ratio, $C$ is linearly related to signal-to-noise ratio; a change in $C$ is proportional to a change in signal-to-noise ratio (in dB). Furthermore, the constant of proportionality depends only on the product $sW$ and is *independent of* any other characteristic of the channel. For small signal-to-noise ratio, $C$ is logarithmically related to signal-to-noise ratio; a change in $\log_{10} C$ is proportional to a change in signal-to-noise ratio (in dB). The constant of proportionality is 0.1 for any channel. In the case in which the channel represents multi-pair telephone cable with small far-end crosstalk, we show that for large signal-to-noise ratio, $C$ is linearly related to the length of the cable, and for small signal-to-noise ratio, $C$ is logarithmically related to length. Furthermore, the effect of the crosstalk is to reduce the dependence of $C$ on cable length. Finally, we present a numerical evaluation of (38) using realistic parameters obtained from an experimental cable consisting of two twisted pairs of wire.

#### *4.1 Dependence of channel capacity on signal-to-noise ratio*

Define a number $N_o$ as

$$N_o = \frac{1}{sW} \int_{-W}^{W} \text{trace } R(f) df.$$

Then $N_o$ represents the noise power per hertz, per dimension, and $sWN_o$ represents the total noise power. We define the following normalized quantities:

$$\begin{aligned} \lambda_i^*(f) &= 2\lambda_i(f)/N_o \\ K^* &= 2K/N_o \\ P &= S/sWN_o \\ P^* &= 10\,(\log_{10} P). \end{aligned}$$

Now $P^*$ is a measure of signal-to-noise ratio in dB. By substituting the above quantities into (38), and using the fact that each $\lambda_i^*$ is a symmetric function, we have

$$P = \frac{1}{sW} \sum_{i=1}^{s} \int_0^W \max(0, K^* - \lambda_i^*(f))\,df \tag{39a}$$

and

$$C = \sum_{i=1}^{s} \int_0^W \max\left(0, \log_2 \frac{K^*}{\lambda_i^*(f)}\right) df. \tag{39b}$$

We will determine the asymptotic behavior of $C$ for both very large and very small $P$. For this purpose we define for every number $K^*$ sets $\Delta_i$, $i = 1, \cdots s$ as

$$\Delta_i = \{f\colon \lambda_i^*(f) \leqq K^*;\ f \geqq 0\}.$$

Let $\delta_i$ be the measure of $\Delta_i$ and define $\delta = (1/s) \sum \delta_i$. In addition, we require the definition of two average channel characteristics, $\bar{\lambda}$ and $\overline{\log \lambda}$. Let

$$\bar{\lambda} = \frac{1}{s\delta} \sum_{i=1}^{s} \int_{\Delta_i} \lambda_i^*(f)\,df,$$

and

$$\overline{\log \lambda} = \frac{1}{s\delta} \sum_{i=1}^{s} \int_{\Delta_i} \log_2 \lambda_i^*(f)\,df.$$

Note that $\delta$, $\bar{\lambda}$ and $\overline{\log \lambda}$ are all functions of $P$. Let $\lambda_{\min} = \min\{\lambda_i^*(f)\colon 0 \leqq f \leqq W;\ 1 \leqq i \leqq s\}$. Recall from Section I that $\lambda_{\min} > 0$.

Now, from (39),

$$\frac{WP}{\delta} = K^* - \bar{\lambda},$$

and

$$\frac{C}{s\delta} = \log_2 K^* - \overline{\log \lambda}.$$

These equations combine to yield

$$\frac{C}{s\delta} = \log_2\left(\frac{WP}{\delta\bar{\lambda}} + 1\right) + \log_2 \bar{\lambda} - \overline{\log \lambda}. \tag{40a}$$

We investigate (40a) for large $P$. Assume all the $\lambda_i^*$'s are bounded. (Actually, boundedness follows from the hypotheses in Section I.) From (39a), $K^*$ is an increasing function of $P$. Let $P$ be sufficiently large so that $\lambda_i(f) \leqq K^*$ for all $f \in [0, W]$ and $i = 1, \cdots s$. Then $\delta = W$ and (40a) yields

$$\frac{C}{sW} = \log_2 \left( \frac{P}{\bar{\lambda}} + 1 \right) + (\log_2 \bar{\lambda} - \overline{\log \lambda}). \tag{40b}$$

Note that for given $\bar{\lambda}$, $s$, $W$, and $P$, $C$ is minimized when all the $\lambda_i^*$'s are equal and constant. Now, for $P \gg \bar{\lambda}$, we have from (40b),

$$C \approx sW(\log_2 P - \overline{\log \lambda}),$$

or

$$C \approx 0.3322\, sWP^* - sW\, \overline{\log \lambda}. \tag{41}$$

Now (41) represents a line in the $C - P^*$ plane with intercept $-sW\, \overline{\log \lambda}$ and slope 0.3322 $sW$, and the region of validity of (41) is $P \gg \bar{\lambda}$. Note that the slope is independent of the $\lambda_i^*$'s; the intercept and region of validity are determined by the average channel characteristics $\bar{\lambda}$ and $\overline{\log \lambda}$ evaluated over the whole interval $[0, W]$.

We now investigate (40a) for small $P$. Observe that $(WP/\delta\bar{\lambda}) + 1 = K^*/\bar{\lambda}$, and as $P$ approaches zero, both $K^*$ and $\bar{\lambda}$ approach $\lambda_{\min}$. Hence, $WP/\delta\bar{\lambda} \to 0$, as $P \to 0$. Then (40a) is approximately

$$\frac{C}{s\delta} \approx \frac{WP}{\delta\bar{\lambda}} \log_2 e + \log_2 \bar{\lambda} - \overline{\log \lambda},$$

which can be rewritten as

$$\frac{C}{sW} \approx \left[ \log_2 e + \frac{\bar{\lambda}(\log_2 \bar{\lambda} - \overline{\log \lambda})}{K^* - \bar{\lambda}} \right] \frac{P}{\bar{\lambda}}. \tag{40c}$$

We show in appendix Section A.4 that for any channel characteristic with $\lambda_{\min} > 0$,

$$\lim_{P \to 0} \frac{\log_2 \bar{\lambda} - \overline{\log \lambda}}{K^* - \bar{\lambda}} = 0. \tag{42}$$

Hence, for small $P$,

$$C \approx \frac{(\log_2 e) sWP}{\lambda_{\min}},$$

or in logarithmic terms,

$$\log_{10} C \approx \frac{P^*}{10} + \log_{10} (sW \log_2 e) - \log_{10} \lambda_{\min}. \tag{43}$$

Now (43) represents a line in the $\log_{10} C - P^*$ plane with intercept $\log_{10} (sW \log_2 e) - \log_{10} \lambda_{min}$, and slope 1/10. However, the region of validity of (43) is difficult to specify because the location of this region depends not just on $\lambda_{min}$ but on the shape of the channel characteristic in a neighborhood of $\lambda_{min}$.

### *4.2 Dependence of channel capacity on cable length*

Suppose that the channel characteristic is a function of a length parameter $l$. Let $l_1$ and $l_2$ be two values of $l$ and suppose that $P^*$ is large and the channel capacity vs $P^*$ characteristic is in the linear region for $l_1$ and $l_2$. Then, from (41),

$$C(l_2) - C(l_1) \approx sW[\overline{\log \lambda(l_1)} - \overline{\log \lambda(l_2)}],$$

or

$$C(l_2) - C(l_1) \approx \sum_{i=1}^{s} \int_0^W \log_2 \left( \frac{\lambda_i^*(l_1; f)}{\lambda_i^*(l_2; f)} \right) df, \tag{44}$$

where in these relations we have explicitly shown the dependence of $\lambda_i^*$ on length. If $P^*$ is very small so that (43) is valid, then

$$\log_{10} C(l_2) - \log_{10} C(l_1) \approx \log_{10}\left( \frac{\lambda_{min}(l_1)}{\lambda_{min}(l_2)} \right). \tag{45}$$

Now consider a multipair cable of length $l$, with $s$ twisted pairs, small far-end crosstalk, and additive white noise. We assume that the crosstalk voltage on a single pair due to all disturbers is proportional to $l^{\frac{1}{2}} f$. Assume also that the attenuation on any pair is proportional to $l f^{\frac{1}{2}}$. If the crosstalk is very small, then a reasonable form for $\lambda_i^*$ is

$$\lambda_i^*(l; f) = \frac{e^{b_i l f^{\frac{1}{2}}}}{1 + c_i l f^2},$$

where $b_i$ and $c_i$ are constants related to attenuation and crosstalk coupling.[4] Define the averages $b$ and $c$ as $b = \sum b_i/s$ and $c = \sum c_i/s$. Now $\lambda_i^*$ can be expressed as

$$\lambda_i^* = \exp[b_i l f^{\frac{1}{2}} - \ln(1 + c_i f^2 l)],$$

and for small crosstalk, we have $c_i f^2 l \ll 1$ for all $i$ and all $f$ and $l$ in a range of interest. Then we have approximately

$$\lambda_i^* \approx e^{l(b_i f^{\frac{1}{2}} - c_i f^2)},$$

and, from (44),

$$\frac{\Delta C}{\Delta l} \approx - \log_2(e) \sum_{i=1}^{s} \int_0^W (b_i f^{\frac{1}{2}} - c_i f^2) df,$$

which can be evaluated as

$$\frac{\Delta C}{\Delta l} \approx -\ 0.96\ bsW^{\frac{1}{2}}\left(1 - \frac{cW^{\frac{3}{2}}}{2b}\right) \tag{46}$$

(for large $P$). Thus, $C$ and $l$ are linearly related. Note that the effect of the crosstalk is to reduce $\Delta C/\Delta l$. If $cW^{\frac{3}{2}}/2b \approx 1$, then $C$ is effectively independent of length. It is theoretically possible to have $cW^{\frac{3}{2}}/2b \approx 1$ and yet have $cW^2 l \ll 1$ as required by our analysis. These relations imply that $2W^{\frac{1}{2}}lb \ll 1$ is a necessary condition that very small crosstalk significantly reduce $\Delta C/\Delta l$. However, we expect that for realistic cable parameters, the reduction in $\Delta C/\Delta l$ due to small crosstalk will not be significant.

Now assume that the channel does not pass dc; i.e., the channel characteristic is that given above for $f \in [f_0, f_1]$, a band of strictly positive frequencies, and is infinite for frequencies outside this band. Let $b_k = \min_i b_i$. Then, for small crosstalk,

$$\lambda_{\min} \approx \exp l(b_k f_0^{\frac{1}{2}} - c_k f_0^2),$$

and

$$\log_{10} \lambda_{\min} \approx 0.434 b_k f_0^{\frac{1}{2}}\left(1 - \frac{c_k f_0^{\frac{3}{2}}}{b_k}\right) l.$$

Thus, for small $P$, we have from (45),

$$\frac{\Delta \log_{10} C}{\Delta l} \approx -\ 0.434 b_k f_0^{\frac{1}{2}}\left(1 - \frac{c_k f_0^{\frac{3}{2}}}{b_k}\right). \tag{47}$$

As in (46), the effect of the crosstalk is to reduce the dependence of channel capacity on cable length.

### 4.3 Numerical example

We consider a two-twisted-pair cable with white additive noise. The transfer function matrix $H(f)$ is given by

$$H(f) = e^{-\gamma l}\begin{bmatrix} 1 & i2\pi k l^{\frac{1}{2}} f \\ i2\pi k l^{\frac{1}{2}} f & 1 \end{bmatrix}$$

with $l$ in feet and $f$ in hertz and

$$\gamma = a\sqrt{2\pi} f^{\frac{1}{2}} + ib2\pi f.$$

The off-diagonal terms in the matrix $H(f)$ represent far-end crosstalk. This model is an approximate representation of an experimental

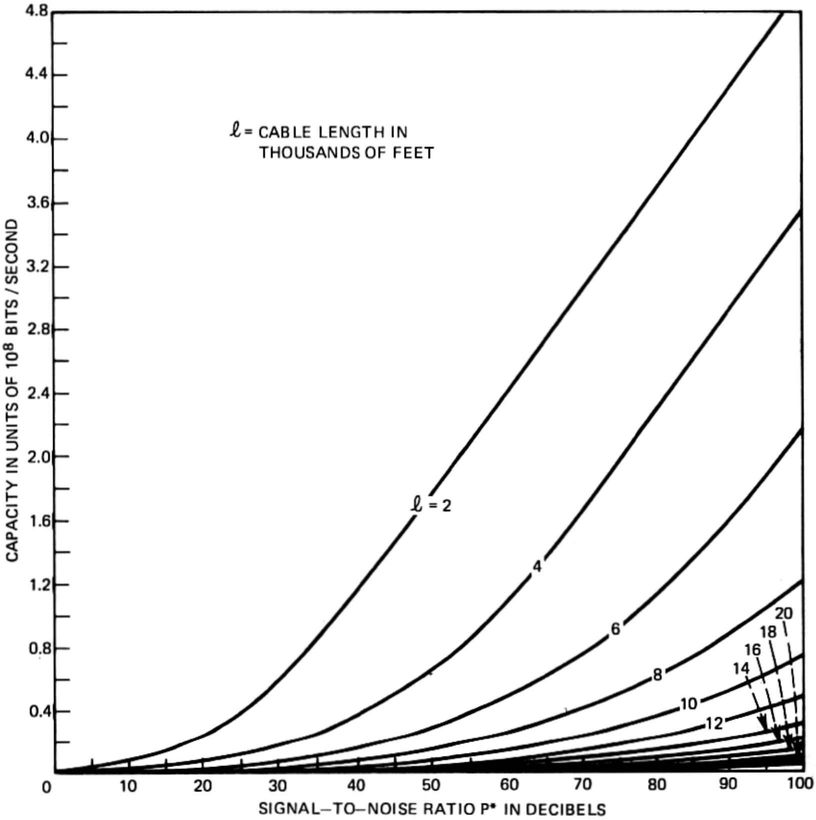


Fig. 1—Channel capacity for experimental cable. Capacity C in units of $10^8$ bits per second is plotted as a function of signal-to-noise ratio $P^*$ for various values of cable length $l$.

two-pair cable. Parameters obtained from measurement are

$$k = 1.26 \times 10^{-12},$$
$$a = 0.23 \times 10^{-6},$$
$$b = 1.48 \times 10^{-9}.$$

This model is valid in the range $10^3 \leqq l \leqq 50 \times 10^3$ feet, and $10^6/2 \leqq f \leqq 10^7$ Hz.

Since the noise is assumed white, the $\lambda_i^*$'s are the eigenvalues of $(H^*H)^{-1}$ and are given by

$$\lambda_1^* = \lambda_2^* = \lambda^* = \frac{\exp(2\sqrt{2\pi}\; af^{\frac{1}{2}})}{1 + (2\pi)^2 f^2 k^2 l}.$$

The capacity equations (39) become

$$\frac{1}{W}\int_{f_0}^{f_1} \max[0, K^* - \lambda^*(f)]df = P \tag{48a}$$

and

$$C = 2\int_{f_0}^{f_1} \max\left(0, \log_2 \frac{K^*}{\lambda^*}\right)df, \tag{48b}$$

where $f_0 = 10^6/2$ and $f_1 = 10^7$.

Numerical evaluation of (48) for various values of $P$ and $l$ has been performed and the results are given in Figs. 1 and 2 and Table I. The figures show $C$ vs $P^*$ for various values of $l$. The $C$ axis is linear in

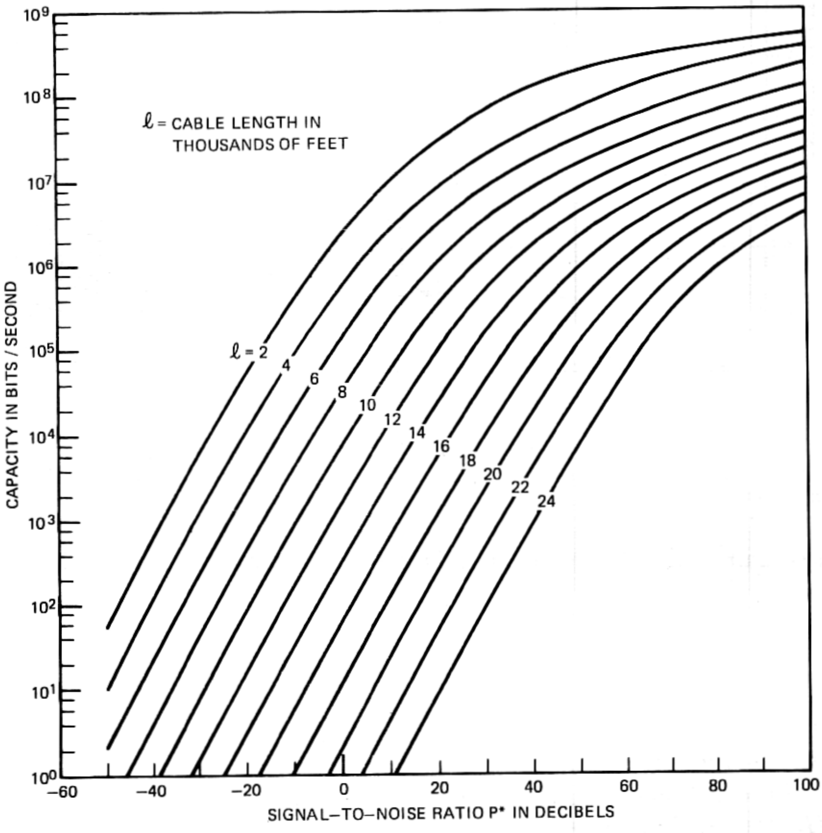


Fig. 2—Channel capacity for experimental cable. Capacity C in bits per second is plotted as a function of signal-to-noise ratio $P^*$ for various values of cable length $l$. The C axis is logarithmic.

Table I — Values of channel capacity $C$ in bits per second for various values of signal-to-noise ratio $P^*$ and cable length. Exponential notation is employed for $C$ ($aEb \equiv a \times 10^b$).

| Signal-to-Noise Ratio $P^*$ in dB | Cable Length $l$ in Thousands of Feet | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 |
| −46 | 1.33E2 | 2.60E1 | 5.07E0 | — | — | — | — | — | — | — | — | — |
| −44 | 2.10E2 | 4.12E1 | 8.18E0 | 1.57E0 | — | — | — | — | — | — | — | — |
| −42 | 3.32E2 | 6.53E1 | 1.28E1 | 2.53E0 | — | — | — | — | — | — | — | — |
| −40 | 5.25E2 | 1.03E2 | 2.03E1 | 3.95E0 | — | — | — | — | — | — | — | — |
| −38 | 8.29E2 | 1.63E2 | 3.21E1 | 6.29E0 | 1.21E0 | — | — | — | — | — | — | — |
| −36 | 1.31E3 | 2.58E2 | 5.11E1 | 9.92E0 | 1.97E0 | — | — | — | — | — | — | — |
| −34 | 2.06E3 | 4.07E2 | 8.06E1 | 1.58E1 | 3.04E0 | — | — | — | — | — | — | — |
| −32 | 3.23E3 | 6.43E2 | 1.27E2 | 2.51E1 | 4.94E0 | — | — | — | — | — | — | — |
| −30 | 5.07E3 | 1.01E3 | 2.01E2 | 3.96E1 | 7.86E0 | 1.52E0 | — | — | — | — | — | — |
| −28 | 7.94E3 | 1.59E3 | 3.17E2 | 6.27E1 | 1.24E1 | 2.36E0 | — | — | — | — | — | — |
| −26 | 1.24E4 | 2.50E3 | 4.99E2 | 9.92E1 | 1.95E1 | 3.79E0 | — | — | — | — | — | — |
| −24 | 1.93E4 | 3.91E3 | 7.87E2 | 1.57E2 | 3.10E1 | 6.11E0 | 1.19E0 | — | — | — | — | — |
| −22 | 2.98E4 | 6.09E3 | 1.26E3 | 2.47E2 | 4.91E1 | 9.66E0 | 1.89E0 | — | — | — | — | — |
| −20 | 4.59E4 | 9.47E3 | 1.94E3 | 3.90E2 | 7.72E1 | 1.52E1 | 3.03E0 | — | — | — | — | — |
| −18 | 7.02E4 | 1.47E4 | 3.03E3 | 6.14E2 | 1.22E2 | 2.41E1 | 4.77E0 | — | — | — | — | — |
| −16 | 1.07E5 | 2.26E4 | 4.72E3 | 9.63E2 | 1.92E2 | 3.82E1 | 7.57E0 | 1.47E0 | — | — | — | — |
| −14 | 1.60E5 | 3.45E4 | 7.33E3 | 1.51E3 | 3.03E2 | 6.00E1 | 1.18E1 | 2.29E0 | — | — | — | — |
| −12 | 2.39E5 | 5.24E4 | 1.13E4 | 2.36E3 | 4.78E2 | 9.50E1 | 1.87E1 | 3.62E0 | — | — | — | — |
| −10 | 3.53E5 | 7.88E4 | 1.74E4 | 3.68E3 | 7.50E2 | 1.51E2 | 2.98E1 | 5.87E0 | 1.11E0 | — | — | — |
| −8 | 5.14E5 | 1.17E5 | 2.65E4 | 5.70E3 | 1.18E3 | 2.36E2 | 4.73E1 | 9.16E0 | 1.77E0 | — | — | — |
| −6 | 7.40E5 | 1.73E5 | 4.01E4 | 8.80E3 | 1.84E3 | 3.72E2 | 7.42E1 | 1.48E1 | 2.88E0 | — | — | — |
| −4 | 1.05E6 | 2.51E5 | 6.01E4 | 1.35E4 | 2.87E3 | 5.87E2 | 1.17E2 | 2.30E1 | 4.45E0 | — | — | — |
| −2 | 1.47E6 | 3.60E5 | 8.91E4 | 2.06E4 | 4.45E3 | 9.18E2 | 1.84E2 | 3.67E1 | 7.32E0 | 1.35E0 | — | — |
| 0 | 2.03E6 | 5.10E5 | 1.31E5 | 3.11E4 | 6.87E3 | 1.44E3 | 2.91E2 | 5.97E1 | 1.15E1 | 2.25E0 | — | — |

## Table I — continued

| Signal-to-Noise Ratio $P^*$ in dB | Cable Length $l$ in Thousands of Feet | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 |
| 2 | 2.76E6 | 7.10E5 | 1.89E5 | 4.66E4 | 1.05E4 | 2.24E3 | 4.58E2 | 9.12E1 | 1.81E1 | 3.56E0 | — | — |
| 4 | 3.71E6 | 9.74E5 | 2.69E5 | 6.90E4 | 1.61E4 | 3.48E3 | 7.19E2 | 1.44E2 | 2.88E1 | 5.61E0 | 1.11E0 | — |
| 6 | 4.90E6 | 1.32E6 | 3.78E5 | 1.01E5 | 2.43E4 | 5.38E3 | 1.12E3 | 2.27E2 | 4.55E1 | 8.99E0 | 1.74E0 | — |
| 8 | 6.40E6 | 1.75E6 | 5.23E5 | 1.46E5 | 3.64E4 | 8.26E3 | 1.75E3 | 3.58E2 | 7.12E1 | 1.42E1 | 2.80E0 | — |
| 10 | 8.24E6 | 2.29E6 | 7.11E5 | 2.07E5 | 5.39E4 | 1.26E4 | 2.73E3 | 5.62E2 | 1.13E2 | 2.24E1 | 4.32E0 | — |
| 12 | 1.05E7 | 2.96E6 | 9.53E5 | 2.90E5 | 7.89E4 | 1.91E4 | 4.21E3 | 8.79E2 | 1.78E2 | 3.54E1 | 6.95E0 | 1.36E0 |
| 14 | 1.32E7 | 3.78E6 | 1.26E6 | 4.00E5 | 1.14E5 | 2.86E4 | 6.48E3 | 1.37E3 | 2.80E2 | 5.54E1 | 1.09E1 | 2.20E0 |
| 16 | 1.64E7 | 4.76E6 | 1.63E6 | 5.43E5 | 1.62E5 | 4.25E4 | 9.91E3 | 2.13E3 | 4.38E2 | 8.79E1 | 1.73E1 | 3.44E0 |
| 18 | 2.02E7 | 5.92E6 | 2.09E6 | 7.26E5 | 2.27E5 | 6.23E4 | 1.50E4 | 3.31E3 | 6.88E2 | 1.39E2 | 2.73E1 | 5.43E0 |
| 20 | 2.46E7 | 7.27E6 | 2.64E6 | 9.53E5 | 3.14E5 | 9.01E4 | 2.26E4 | 5.09E3 | 1.07E3 | 2.19E2 | 4.35E1 | 8.49E0 |
| 22 | 2.98E7 | 8.85E6 | 3.29E6 | 1.23E6 | 4.26E5 | 1.28E5 | 3.36E4 | 7.80E3 | 1.67E3 | 3.42E2 | 6.83E1 | 1.37E1 |
| 24 | 3.57E7 | 1.07E7 | 4.06E6 | 1.58E6 | 5.68E5 | 1.80E5 | 4.94E4 | 1.19E4 | 2.60E3 | 5.38E2 | 1.08E2 | 2.12E1 |
| 26 | 4.25E7 | 1.27E7 | 4.95E6 | 1.98E6 | 7.47E5 | 2.49E5 | 7.16E4 | 1.79E4 | 4.00E3 | 8.41E2 | 1.71E2 | 3.40E1 |
| 28 | 5.02E7 | 1.51E7 | 5.97E6 | 2.46E6 | 9.66E5 | 3.39E5 | 1.02E5 | 2.66E4 | 6.14E3 | 1.31E3 | 2.68E2 | 5.34E1 |
| 30 | 5.88E7 | 1.78E7 | 7.13E6 | 3.02E6 | 1.23E6 | 4.53E5 | 1.44E5 | 3.92E4 | 9.36E3 | 2.04E3 | 4.21E2 | 6.49E1 |
| 32 | 6.86E7 | 2.07E7 | 8.44E6 | 3.67E6 | 1.55E6 | 5.96E5 | 2.00E5 | 5.71E4 | 1.41E4 | 3.15E3 | 6.59E2 | 1.33E2 |
| 34 | 7.92E7 | 2.40E7 | 9.92E6 | 4.41E6 | 1.92E6 | 7.72E5 | 2.72E5 | 8.19E4 | 2.11E4 | 4.84E3 | 1.03E3 | 2.10E2 |
| 36 | 9.04E7 | 2.77E7 | 1.16E7 | 5.25E6 | 2.36E6 | 9.86E5 | 3.65E5 | 1.16E5 | 3.13E4 | 7.39E3 | 1.60E3 | 3.29E2 |
| 38 | 1.02E8 | 3.18E7 | 1.34E7 | 6.19E6 | 2.86E6 | 1.24E6 | 4.82E5 | 1.61E5 | 4.57E4 | 1.12E4 | 2.48E3 | 5.15E2 |
| 40 | 1.14E8 | 3.62E7 | 1.52E7 | 7.25E6 | 3.43E6 | 1.54E6 | 6.26E5 | 2.21E5 | 6.58E4 | 1.68E4 | 3.82E3 | 8.06E2 |
| 42 | 1.26E8 | 4.11E7 | 1.76E7 | 8.42E6 | 4.07E6 | 1.89E6 | 8.01E5 | 2.97E5 | 9.34E4 | 2.49E4 | 5.84E3 | 1.26E3 |
| 44 | 1.39E8 | 4.64E7 | 2.01E7 | 9.71E6 | 4.79E6 | 2.29E6 | 1.01E6 | 3.93E5 | 1.31E5 | 3.66E4 | 8.88E3 | 1.95E3 |
| 46 | 1.51E8 | 5.22E7 | 2.27E7 | 1.11E7 | 5.60E6 | 2.75E6 | 1.26E6 | 5.12E5 | 1.79E5 | 5.29E4 | 1.33E4 | 3.01E3 |
| 48 | 1.63E8 | 5.85E7 | 2.56E7 | 1.27E7 | 6.50E6 | 3.26E6 | 1.54E6 | 6.58E5 | 2.43E5 | 7.54E4 | 1.99E4 | 4.61E3 |

## Table I — continued

| Signal-to-Noise Ratio $P^*$ in dB | Cable Length $l$ in Thousands of Feet | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 |
| 50 | 1.76E8 | 6.52E7 | 2.88E7 | 1.44E7 | 7.48E6 | 3.84E6 | 1.87E6 | 8.32E5 | 3.23E5 | 1.06E5 | 2.93E4 | 7.02E3 |
| 52 | 1.88E8 | 7.25E7 | 3.21E7 | 1.62E7 | 8.56E6 | 4.49E6 | 2.25E6 | 1.04E6 | 4.22E5 | 1.46E5 | 4.26E4 | 1.06E4 |
| 54 | 2.01E8 | 8.04E7 | 3.58E7 | 1.82E7 | 9.74E6 | 5.20E6 | 2.68E6 | 1.28E6 | 5.44E5 | 1.99E5 | 6.10E4 | 1.59E4 |
| 56 | 2.14E8 | 8.88E7 | 3.97E7 | 2.04E7 | 1.10E7 | 5.99E6 | 3.15E6 | 1.55E6 | 6.90E5 | 2.66E5 | 8.61E4 | 2.35E4 |
| 58 | 2.26E8 | 9.79E7 | 4.39E7 | 2.27E7 | 1.24E7 | 6.85E6 | 3.68E6 | 1.87E6 | 8.64E5 | 3.50E5 | 1.20E5 | 3.43E4 |
| 60 | 2.39E8 | 1.08E8 | 4.84E7 | 2.52E7 | 1.39E7 | 7.78E6 | 4.27E6 | 2.23E6 | 1.07E6 | 4.52E5 | 1.63E5 | 4.93E4 |
| 62 | 2.51E8 | 1.18E8 | 5.31E7 | 2.79E7 | 1.55E7 | 8.80E6 | 4.91E6 | 2.63E6 | 1.30E6 | 5.76E5 | 2.20E5 | 7.00E4 |
| 64 | 2.64E8 | 1.29E8 | 5.82E7 | 3.07E7 | 1.73E7 | 9.90E6 | 5.62E6 | 3.07E6 | 1 57E6 | 7.24E5 | 2.90E5 | 9.78E4 |
| 66 | 2.76E8 | 1.40E8 | 6.37E7 | 3.37E7 | 1.91E7 | 1.10E7 | 6.35E6 | 3.56E6 | 1.87E6 | 8.98E5 | 3.77E5 | 1.34E5 |
| 68 | 2.89E8 | 1.52E8 | 6.94E7 | 3.70E7 | 2.11E7 | 1.24E7 | 7.22E6 | 4 11E6 | 2.21E6 | 1.10E6 | 4.83E5 | 1.82E5 |
| 70 | 3.01E8 | 1.64E8 | 7.55E7 | 4.04E7 | 2.32E7 | 1.37E7 | 8.12E6 | 4.70E6 | 2.59E6 | 1.33E6 | 6.10E5 | 2.41E5 |
| 72 | 3.14E8 | 1.76E8 | 8.19E7 | 4.40E7 | 2.55E7 | 1.52E7 | 9.09E6 | 5.35E6 | 3.02E6 | 1.59E6 | 7.59E5 | 3.17E5 |
| 74 | 3.26E8 | 1.89E8 | 8.88E7 | 4.78E7 | 2.78E7 | 1.67E7 | 1.01E7 | 6.05E6 | 3.48E6 | 1.88E6 | 9.32E5 | 4.06E5 |
| 76 | 3.39E8 | 2.01E8 | 9.59E7 | 5.19E7 | 3.04E7 | 1.84E7 | 1.13E7 | 6.80E6 | 3.98E6 | 2.21E6 | 1.13E6 | 5.15E5 |
| 78 | 3.52E8 | 2.14E8 | 1.04E8 | 5.62E7 | 3.30E7 | 2.02E7 | 1.24E7 | 7.62E6 | 4.54E6 | 2.58E6 | 1.36E6 | 6.44E5 |
| 80 | 3.64E8 | 2.26E8 | 1.11E8 | 6.07E7 | 3.58E7 | 2.20E7 | 1.37E7 | 8.49E6 | 5.13E6 | 2.98E6 | 1.61E6 | 7.94E5 |
| 82 | 3.77E8 | 2.39E8 | 1.20E8 | 6.54E7 | 3.88E7 | 2.40E7 | 1.51E7 | 9.42E6 | 5.78E6 | 3.41E6 | 1.90E6 | 9.67E5 |
| 84 | 3.89E8 | 2.51E8 | 1.29E8 | 7.04E7 | 4.19E7 | 2.61E7 | 1.65E7 | 1.04E7 | 6.48E6 | 3.89E6 | 2.22E6 | 1.17E6 |
| 86 | 4.02E8 | 2.64E8 | 1.39E8 | 7.56E7 | 4.52E7 | 2.83E7 | 1.80E7 | 1.15E7 | 7.22E6 | 4.41E6 | 2.57E6 | 1.39E6 |
| 88 | 4.14E8 | 2.76E8 | 1.47E8 | 8.11E7 | 4.87E7 | 3.06E7 | 1.96E7 | 1.26E7 | 8.02E6 | 4.97E6 | 2.95E6 | 1.64E6 |
| 90 | 4.27E8 | 2.89E8 | 1.58E8 | 8.68E7 | 5.23E7 | 3.30E7 | 2.13E7 | 1.38E7 | 8.87E6 | 5.57E6 | 3.37E6 | 1.92E6 |
| 92 | 4.39E8 | 3.01E8 | 1.68E8 | 9.28E7 | 5.61E7 | 3.56E7 | 2.31E7 | 1.51E7 | 9.78E6 | 6.22E6 | 3.82E6 | 2.23E6 |
| 94 | 4.52E8 | 3.14E8 | 1.79E8 | 9.91E7 | 6.01E7 | 3.82E7 | 2.49E7 | 1.64E7 | 1.07E7 | 6.91E6 | 4.31E6 | 2.56E6 |
| 96 | 4.64E8 | 3.27E8 | 1.91E8 | 1.06E8 | 6.42E7 | 4.10E7 | 2.69E7 | 1.78E7 | 1.18E7 | 7.65E6 | 4.84E6 | 2.93E6 |
| 98 | 4.77E8 | 3.39E8 | 2.02E8 | 1.13E8 | 6.86E7 | 4.40E7 | 2.90E7 | 1.93E7 | 1.28E7 | 8.43E6 | 5.41E6 | 3.33E6 |
| 100 | 4.90E8 | 3.52E8 | 2.14E8 | 1.20E8 | 7.31E7 | 4.70E7 | 3.11E7 | 2.09E7 | 1.40E7 | 9.27E6 | 6.01E6 | 3.77E6 |

Fig. 1, and is logarithmic in Fig. 2. The linear regions discussed above are evident in these figures. The asymptotic estimates for large $P$ given in (41) and (46) are $\Delta C/\Delta P^* \approx 6.3 \times 10^6$ (b/s/dB), and $\Delta C/\Delta l \approx -70 \times 10^3$ (b/s/ft). For constant $C$, $\Delta P^*/\Delta l \approx 11 \times 10^{-3}$ (dB/ft) or about 58 dB/mile. This is the amount of increase of signal-to-noise ratio necessary to maintain a fixed level of $C$ as length is increased. The asymptotic estimates for small $P$ given (43) and (47) are $\Delta \log_{10} C/\Delta P^* \approx 1/10$, and $\Delta \log_{10} C/\Delta l \approx -0.353 \times 10^{-3}$. For constant $C$, $\Delta P^*/\Delta l \approx 3.53 \times 10^{-3}$(dB/ft), or about 19 dB/mile. These asymptotic estimates are borne out in the numerical evaluation.

## V. ACKNOWLEDGMENT

The authors are indebted to M. I. Schwartz for several helpful discussions regarding Section IV of this work and for providing the data used in the numerical results, and to Ms. M. A. Gatto for programming assistance.

## APPENDIX

### A.1 *Proof of Theorem 2*

a. Let $y(n) = (\mathfrak{F}\mathbf{x})(n) = \sum_k f(n-k)x(k)$, and let $\sum_n \|f(n)\| = C < \infty$. From the triangle and Schwarz inequalities,

$$\|y(n)\|^2 \leqq (\sum_k \|f(n-k)\| \cdot \|x(k)\|)^2 \leqq C \sum_k \|f(n-k)\| \cdot \|x(k)\|^2.$$

Hence,

$$|||\mathbf{y}|||^2 = \sum_n \|y(n)\|^2 \leqq C \sum_n \sum_k \|f(n-k)\| \cdot \|x(k)\|^2 \leqq C^2 |||\mathbf{x}|||^2,$$

and $|\mathfrak{F}| \leqq C$.

b. From Parseval's theorem,

$$\frac{|||\mathfrak{F}\mathbf{x}|||^2}{|||\mathbf{x}|||^2} = \frac{\int_{-\pi}^{\pi} \|F(\theta)X(\theta)\|^2 d\theta}{\int_{-\pi}^{\pi} \|X(\theta)\|^2 d\theta} \leqq \max_\theta \|F(\theta)\|^2.$$

c. If $\mathfrak{F}$ is self-adjoint, then

$$|\langle \mathbf{x}, \mathfrak{F}\mathbf{x} \rangle| \leqq |\mathfrak{F}| \cdot |||\mathbf{x}|||^2 \leqq \max_{-\pi \leqq \theta \leqq \pi} \|F(\theta)\| \cdot |||\mathbf{x}|||^2.$$

d. The essence of this result is a matricized version of Wiener's well-known theorem on the reciprocal of an absolutely convergent Fourier series (see Ref. 5, p. 430).

Suppose $\det F(\theta) \neq 0$. We show that $\mathfrak{F}$ has a bounded inverse in $L$. Now $\det F(\theta)$ is a scalar function consisting of a sum of products of functions each possessing an absolutely convergent Fourier series.* Hence, $\det F(\theta)$ and, by Wiener's theorem, $[\det F(\theta)]^{-1}$ have absolutely convergent Fourier series. Each element of $F^{-1}(\theta)$ is the ratio of a minor determinant to $\det F(\theta)$ so each element has an absolutely convergent Fourier series. Hence, $F^{-1}(\theta)$ has a Fourier series $\sum_n g(n)e^{in\theta}$ with $\sum_n \|g(n)\| < \infty$. Consequently, $\mathfrak{F}$ has a bounded inverse $\mathfrak{F}^{-1}$ in $L$.

Conversely, let $\mathfrak{F}$ have a bounded inverse $\mathfrak{F}^{-1}$. Then there exists an $\alpha > 0$ such that for all $\mathbf{x}$ in $l_2^{(s)}(-\infty, \infty)$, $|||\mathfrak{F}\mathbf{x}||| \geqq \alpha|||\mathbf{x}|||$. Let $X(\theta) = \sum_n x(n)e^{in\theta}$. From Parseval's theorem,

$$0 < \alpha^2 \leqq \inf_{X(\theta)} \frac{\int_{-\pi}^{\pi} \|F(\theta)X(\theta)\|^2 d\theta}{\int_{-\pi}^{\pi} \|X(\theta)\|^2 d\theta},$$

which implies, since $F(\theta)$ is continuous, that $F(\theta)$ is one-to-one; i.e., $\det F(\theta) \neq 0$ for $-\pi \leqq \theta \leqq \pi$. This completes the proof of Theorem 2.

There are other interesting properties of the class $L$, which are not directly relevant to the main results of this paper. For the sake of completeness, we mention two generalizations of Theorem 2d, that also have well-known scalar counterparts:

(*i*) Let $\mathfrak{F} \in L$ and let $\sigma(\mathfrak{F})$ denote the set of all eigenvalues of $F(\theta)$, $-\pi \leqq \theta \leqq \pi$. Let $I$ be the identity on $l_2^{(s)}$. For $\lambda$ any complex number, $\lambda I - \mathfrak{F}$ has a bounded inverse in $L$ if and only if $\lambda \notin \sigma(\mathfrak{F})$.

(*ii*) Let $g(\cdot)$ be any function analytic in a neighborhood containing $\sigma(\mathfrak{F})$. Then there is an operator $g(\mathfrak{F})$ in $L$ which has as its transfer matrix the function $g[F(\theta)]$.

### A.2 *Lemmas 3 and 4*

These lemmas apply to the special case where $H(\theta) = I_s$, and $R(\theta) = R_2(\theta)$. Let $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ be the channel input, output, and noise sequences respectively, and let $\mathbf{x}^{(N)}$, $\mathbf{y}^{(N)}$, $\mathbf{z}^{(N)}$ be the corresponding finite sequences. Thus,

$$\mathbf{y}^{(N)} = \mathbf{x}^{(N)} + \mathbf{z}^{(N)}.$$

---

* Note that $\sum_n \|f(n)\| < \infty$ if and only if $\sum_n |f_{ij}(n)| < \infty$ for $1 \leqq i, j \leqq s$.

Letting $T_N$ be the whitening matrix defined in (18), we have

$$\mathbf{v} \triangleq T_N \mathbf{y}^{(N)} = T_N \mathbf{x}^{(N)} + \mathbf{w},$$

where $E\mathbf{w}\mathbf{w}^t = I_{N \cdot s}$. This is precisely the discrete-time version of the problem treated in Chapter 8 of Gallager.[1] The results obtained there apply here exactly when we use, instead of his Lemma 8.5.2 (the Kac-Murdock-Szego theorem), the following discrete-time version:

*Theorem. Let $\{c_i\}$ $i = 0, \pm 1, \cdots$ be a sequence of $s \times s$ matrices such that the $Ns \times Ns$ matrix $C_N = \{c_{i-j}\}$, $i, j = 0, \cdots, N-1$, is Hermitian, and $\sum_k \|c_k\| < \infty$. Let $v_1^{(N)}, v_2^{(N)}, \cdots, v_{sN}^{(N)}$ be the eigenvalues of $C_N$ (each counted according to its multiplicity) and let $\lambda_1(\theta), \lambda_2(\theta), \cdots, \lambda_s(\theta)$ be the eigenvalues of $C(\theta) \equiv \sum_k c_k e^{ik\theta}$. Let $g(\cdot)$ be any continuous function defined on an interval containing the values $\{\lambda_k(\theta): -\pi \leqq \theta \leqq \pi, k = 1, 2, \cdots, s\}$. Then*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{sN} g(v_k^{(N)}) = \frac{1}{2\pi} \sum_{k=1}^{s} \int_{-\pi}^{\pi} g[\lambda_k(\theta)] d\theta.$$

*Furthermore, let $D_N(x) = 1/N$ (number of eigenvalues $v_k^{(N)} \leqq x$). Then*

$$\lim_{N \to \infty} D_N(x) = \frac{1}{2\pi} \sum_{k=1}^{s} \int_{\lambda_k(\theta) \leqq x} d\theta.$$

In the scalar case, $s = 1$, this theorem represents well-known results,[6] a simple account of which can also be found in Ref. 7. The validity of the theorem for $s > 1$ follows on verifying that the arguments employed in Ref. 7 are valid in general for $s \geqq 1$.

### A.3 *Derivation of (38)*

We show how to obtain the capacity formula given in (38). We will do this for the scalar case; the result for vectors follows similarly. The capacity formula justified in Theorem 1 can be stated as follows.

The channel input and output are sequences $\mathbf{x} = \{x_n\}_{-\infty}^{\infty}$ and $\mathbf{y} = \{y_n\}_{-\infty}^{\infty}$ respectively, related by

$$y_n = \sum_{k=-\infty}^{\infty} h_{n-k} x_k + z_n, \tag{49}$$

where $\mathbf{h} = \{h_n\}_{n=-\infty}^{\infty}$ is a fixed sequence and $\mathbf{z} = \{z_n\}_{-\infty}^{\infty}$ is a stationary sequence of Gaussian random variables for which $Ez_n = 0$,

$$Ez_m z_{m-n} = \hat{r}_n, \quad -\infty < n, m < \infty. \tag{50}$$

We write (49) symbolically as

$$\mathbf{y} = \mathbf{h}*\mathbf{x} + \mathbf{z}, \tag{51}$$

where "*" denotes vector convolution. The capacity of this channel is given as follows.

For codes of block length $N$, say that the code vectors $\mathbf{x}$ must satisfy

$$x_n = 0, \quad n \notin [0, N-1]. \tag{52a}$$

$$\frac{1}{N}\sum_{n=0}^{N-1} x_n^2 \leqq S_D. \tag{52b}$$

The capacity is then given by

$$C = \frac{1}{4W}\int_{-W}^{W} df \max\left(0, \log_2 \frac{K|H_D(f)|^2}{\hat{R}_D(f)}\right), \tag{53a}$$

where

$$H_D(f) = \sum_{n=-\infty}^{\infty} h_n e^{(i\pi/W)nf}, \quad |f| \leqq W, \tag{53b}$$

and

$$\hat{R}_D(f) = \sum_{n=-\infty}^{\infty} \hat{r}_n e^{(i\pi/W)nf}, \quad |f| \leqq W, \tag{53c}$$

are the discrete Fourier transforms of $\{h_n\}$ and $\{\hat{r}_n\}$ respectively, and $K$ is the unique solution of

$$S_D = \frac{1}{2W}\int_{-W}^{W} df \max\left(0, K - \frac{\hat{R}_D(f)}{|H_D(f)|^2}\right). \tag{53d}$$

#### A.3.1 *Some facts about band-limited functions*

Before discussing the continuous-time channel, we digress to mention several facts about band-limited functions.

We denote time functions by lower-case letters, e.g., $u(t)$, and the corresponding Fourier transform by upper-case letters, e.g., $U(f)$. Thus

$$U(f) = \int_{-\infty}^{\infty} u(t)e^{i2\pi ft}dt, \tag{54a}$$

and

$$u(t) = \int_{-\infty}^{\infty} U(f)e^{-i2\pi ft}df. \tag{54b}$$

We shall assume that all functions are square-integrable, and all integrals and infinite sums are limits in the mean. We say that $u$ is band-limited to $W$ Hz if $U(f) = 0, \ |f| > W$. Let

$$g_n(t) = \frac{\sin 2\pi W\left(t - \dfrac{n}{2W}\right)}{\pi\left(t - \dfrac{n}{2W}\right)}, \quad n = 0, \pm 1, \pm 2, \cdots, \tag{55}$$

be the sampling functions. Note that for $k, n = 0, \pm 1, \cdots,$

$$g_n\left(\frac{k}{2W}\right) = \begin{cases} 0 & k \neq n, \\ 2W & k = n, \end{cases} \tag{56a}$$

and

$$G_n(f) = \int_{-\infty}^{\infty} g_n(t) e^{i2\pi f t} dt = \begin{cases} e^{(in\pi/W)f} & |f| < W \\ 0 & |f| > W \end{cases} \tag{56b}$$

(so that $g_n$ is band-limited), and for $k, n = 0, \pm 1, \cdots$

$$\int_{-\infty}^{\infty} g_n(t) g_k(t) dt = \int_{-\infty}^{\infty} G_n(f) G_k^*(f) df = \int_{-W}^{W} e^{(i\pi f/W)(n-k)} df = \begin{cases} 0, & n \neq k, \\ 2W, & n = k. \end{cases} \tag{56c}$$

Further, the well-known Sampling Theorem implies that any band-limited function, $u(t)$, can be written as

$$u(t) = \sum_{n=-\infty}^{\infty} u_n g_n(t), \tag{57a}$$

where

$$u_n = \frac{1}{2W} u\left(\frac{n}{2W}\right). \tag{57b}$$

Further, from (57), and (56c),

$$\int_{-\infty}^{\infty} u^2(t) dt = 2W \sum_{n=-\infty}^{\infty} u_n^2. \tag{57c}$$

Let $u(t) = \sum u_n g_n(t)$ and $v(t) = \sum v_n g_n(t)$ be band-limited functions. Then their convolution is

$$w(t) = \int_{-\infty}^{\infty} u(t - \lambda) v(t) dt = \sum w_n g_n(t), \tag{58a}$$

where

$$w_n = \sum_{m=-\infty}^{\infty} u_{n-m} v_m, \quad -\infty < n < \infty, \tag{58b}$$

i.e., $\mathbf{w} = \mathbf{u} * \mathbf{v}$.

Finally, let $z(t)$ $(-\infty < t < \infty)$ be a random process with $Ez(t) = 0$, and covariance

$$Ez(t)z(t-\tau) = r(\tau) \quad -\infty < t, \tau < \infty.$$

Let

$$R(f) = \int_{-\infty}^{\infty} r(t) e^{i2\pi ft} dt$$

satisfy $R(f) = 0$, $|f| > W$, so that $r(t)$ is band-limited. Then, from (57),

$$r(t) = \sum_n \frac{1}{2W} r\left(\frac{n}{2W}\right) g_n(t),$$

and from (56b),

$$R(f) = \sum_n \frac{1}{2W} r\left(\frac{n}{2W}\right) G_n(f) = \frac{1}{2W} \sum_n r\left(\frac{n}{2W}\right) e^{(i\pi/W)nf}. \tag{59}$$

Further, the random process $z(t)$ is a band-limited function so that by (57) we can write

$$z(t) = \sum_n z_n g_n(t) = \sum_n \frac{1}{2W} z\left(\frac{n}{2W}\right) g_n(t).$$

Thus,

$$\hat{r}_n \triangleq E z_m z_{m-n} = \frac{1}{(2W)^2} E z\left(\frac{m}{2W}\right) z\left(\frac{m-n}{2W}\right) = \frac{1}{(2W)^2} r\left(\frac{n}{2W}\right).$$

Thus, the discrete Fourier transform of $\{\hat{r}_m\}$ is, using (59),

$$\hat{R}_D(f) = \sum_n \hat{r}_n e^{(i\pi/W)nf} = \frac{1}{(2W)^2} \sum_n r\left(\frac{n}{2W}\right) e^{(i\pi/W)nf} = \frac{1}{2W} R(f). \tag{60}$$

### A.3.2 *The continuous-time channel*

The continuous-time channel is defined as follows. The channel input and output are functions $x(t)$ and $y(t)$, respectively, where

$$y(t) = \int_{-\infty}^{\infty} h(t-\lambda) x(\lambda)\, d\lambda + z(t), \quad -\infty < t < \infty, \tag{61}$$

where $h(t)$ is a fixed function and $z(t)$ is a Gaussian random process with covariance as described above. We assume that $x(t)$, $h(t)$, and, therefore, $y(t)$ are band-limited to $W$ Hz. Let us expand $x$, $h$, $z$, and $y$ into series in $g_n(t)$ as in (57). Using (58) we obtain

$$y_n = \sum_{m=-\infty}^{\infty} h_{n-m}x_m + z_m. \tag{62}$$

Since knowledge of the sequences of coefficients $\{x_n\}$, $\{y_n\}$, etc. is equivalent to knowledge of the time functions $x(t)$, $y(t)$, etc., the continuous-time channel is equivalent to the discrete-time channel discussed at the beginning of this appendix. It remains to find the corresponding parameters.

Now the code words (in a $T$-second block-coding interval) are taken to be band-limited functions $x(t)$ such that the samples $x(n/2W) = 0$, for $(n/2W) < 0$ or $(n/2W) \geqq T$, i.e., $x(n/2W) = 0$, $n \notin [0, N-1]$, where $N = 2WT$. Thus, $x_n = (1/2W)x(n/2W) = 0$, $n \notin [0, N-1]$. The condition

$$\int_{-\infty}^{\infty} x^2(t)dt \leqq ST$$

is, in the light of (57c),

$$\frac{1}{N}\sum_{n=0}^{N-1} x_n^2 \leqq \frac{S}{(2W)^2}. \tag{63}$$

The quantity

$$H_D(f) = \sum_{n=-\infty}^{\infty} h_n e^{(i\pi/W)nf} = \sum_n h_n G_n(f) = H(f),$$

and by (60), $\hat{R}_D(f) = (1/2W)R(f)$. Thus, the continuous-time channel is equivalent to a discrete-time channel with $S_D = S/2W$, $H_D(f) = H(f)$, and $\hat{R}_D(f) = (1/2W)R(f)$. Thus, from (53)

$$C = \frac{1}{4W}\int_{-W}^{W} df \max\left(0, \log_2 \frac{(2WK)|H(f)|^2}{R(f)}\right),$$

where $K$ is the solution to

$$\frac{S}{(2W)^2} = \frac{1}{2W}\int_{-W}^{W} df \max\left(0, K - \frac{1}{2W}\frac{R(f)}{|H(f)|^2}\right).$$

Letting $K^* = 2WK$, we have

$$S = \int_{-W}^{W} df \max\left(0, K^* - \frac{R(f)}{|H(f)|^2}\right)$$

and

$$C = \frac{1}{4W} \int_{-W}^{W} df \max\left(0, \log_2 \frac{K^*|H(f)|^2}{R(f)}\right).$$

The expression for $C$ is in bits per sampling time. To obtain $C$ in bits per second, multiply by $2W$.

### A.4 *Proof of equation (42)*

Equation (42) follows at once from the theorem below if the functions $\lambda_i^*$, $i = 1, 2, \cdots s$, are replaced by a single function $f$ representing a concatenation of the functions $\lambda_i^*$; $f$ will be bounded away from zero since the same is true of each $\lambda_i^*$.

*Theorem: Let $f(\cdot)$ be a measurable function on a finite interval and* ess inf $f(x) = f_0 > 0$. *For any $K > f_0$ define $\Delta = \{x; f(x) \leqq K\}$ and let $\delta$ be the measure of $\Delta$. Define*

$$I_f(K) = \frac{\log \frac{1}{\delta} \int_\Delta f(x)dx - \frac{1}{\delta} \int_\Delta \log f(x)dx}{K - \frac{1}{\delta} \int_\Delta f(x)dx}.$$

*Then*

$$\lim_{K \to f_0} I_f(K) = 0.$$

*Proof:* Without loss of generality we can take the log to be the natural log and can assume $f_0 = 1$. Let $f = 1 + g$ and $K = 1 + k$. For each $n \geqq 1$ define

$$\overline{g^n} = \frac{1}{\delta} \int_\Delta g^n(x)dx.$$

For all $x \in \Delta$, $g(x) \leqq k$ and $\bar{g} \leqq k$. For $k < 1$, the log may be expanded in a power series. After some rearrangement of terms, we have

$$I_f(K) = \frac{\sum_{n=1}^{\infty} \left[\frac{1}{2n}(\overline{g^{2n}} - \bar{g}^{2n}) - \frac{1}{2n+1}(\overline{g^{2n+1}} - \bar{g}^{2n+1})\right]}{k - \bar{g}}.$$

But $\bar{g} \leqq k$ and by Jensen's inequality $\bar{g}^n \leqq \overline{g^n}$ for all $n \geqq 1$. Then

$$I_f(K) \leqq \frac{\sum_{n=1}^{\infty} \frac{1}{2n}(k^{2n} - \bar{g}^{2n})}{k - \bar{g}}.$$

Now for all $n \geqq 1$,

$$k^n - \bar{g}^n = (k - \bar{g}) \sum_{i=0}^{n-1} k^{n-i-1}\bar{g}^i \leqq (k - \bar{g})nk^{n-1},$$

and

$$I_f(K) \leqq \sum_{n=1}^{\infty} \frac{k^{n-1}}{2} (k^n + \bar{g}^n) \leqq \frac{1}{k} \sum_{n=1}^{\infty} k^{2n} = \frac{k}{1 - k^2},$$

or, since $K = 1 + k$,

$$I_f(K) \leqq \frac{K - 1}{1 - (K - 1)^2},$$

and the result is proved.

## REFERENCES


1. R. G. Gallager, *Information Theory and Reliable Communication,* New York: John Wiley, 1968.
2. A. D. Wyner, "On the Intersymbol Interference Problem for the Gaussian Channel," B.S.T.J., *50,* No. 7 (September 1971), pp. 2355–2363.
3. W. Toms and T. Berger, "Capacity and Error Exponents of a Channel Modeled as a Linear Dynamical System," IEEE Trans. Info. Theory, *IT-19,* No. 1 (January 1973), pp. 124–126.
4. M. I. Schwartz, private communication.
5. F. Riesz and B. Sz.-Nagy, *Functional Analysis,* New York: Frederick Ungar, 1955.
6. U. Grenander and G. Szego, *Toeplitz Forms and their Applications,* Berkeley: University of California, 1958.
7. R. M. Gray, "On the Asymptotic Eigenvalue Distribution of Toeplitz Matrices," IEEE Trans. Info. Theory, *IT-18,* No. 6 (November 1972), pp. 725–730.