CAPACITY OF CLASSES OF GAUSSIAN CHANNELS
PART II: CONTINUOUS-TIME

by

W. L. Root

P. P. Varaiya

Memorandum No. ERL-M212

29 May 1967

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Capacity of Classes of Gaussian Channels

Part II: Continuous-time

by

W. L. Root
University of Michigan
Ann Arbor, Michigan

P. P. Varaiya
University of California
Berkeley, California

1. Introduction

A coding theorem and (weak) converse are proved for classes of continuous-time channels with additive white Gaussian noise in which a different time-invariant linear operation is performed on the transmitted signal in each channel. The proof is accomplished by reducing the problem to one involving discrete-time Gaussian channels with matrix operators, which is solved in Ref. 1. The development in this paper rests heavily on Ref. 2-4. The paper of Blackwell, Breiman and Thomasian[2] provides the clue that the capacity of a collection of channels to be considered simultaneously may be defined as $\sup_{p} \inf_{\mathcal{b}}$ [expected

The research reported herein was supported in part by the National Aeronautics and Space Administration under Grant NsG-2-59 to the University of Michigan and by the National Science Foundation under Grant GK-716 to the University of California, Berkeley.

value of the mutual information], where the infimum is over the class of channels, and the supremum is over input probability distributions. The work of Gallagher[3] effectively provides the formula for the mutual information, and the paper of Kac, Murdock and Szego[4] provides an essential asymptotic relation for the eigenvalues of integral operators of an appropriate type.

2. Description of the Problem

We consider communication channels and classes of channels that can be described as follows. By a transmitted signal, or input signal, over the time interval $[-T, T]$ we mean a real-valued function x which is square-integrable with respect to Lebesgue measure on $[-T, T]$. If x is the input signal over $[-T, T]$, the received signal, or output signal, $y(t)$ over an interval $[a, b]$ is to be given by an expression of the form

$$y(t) = \int_{-T}^{T} h(t-\tau)\, x(\tau)\, d\tau + z(t), \quad a \leq t \leq b, \tag{1}$$

where $z(t)$ is white Gaussian noise with average power density N and mean zero.* Unless otherwise stated we shall always assume $a = -T$, $b = T$. Since the communication channel is completely specified once the function h is specified, we may refer to a channel h, and to collections $\mathcal{C}$ of channels h.

Footnote to page 2.

The noise term z(t) in Eq. 1 must be interpreted symbolically since white noise cannot be parametrized with a time variable, but must properly be parametrized with an element of a space of "testing functions." However we deal only with functionals of y(t) of the form

$$\int_a^b y(t)\,\varphi(t)\,dt\,,$$

where $\varphi \in L_2(a,b)$, or with quantities derivable from these functionals. Hence we can define

$$\int_a^b z(t)\,\varphi(t)\,dt$$

to mean

$$\int_a^b \varphi(t)\,d\zeta(t)$$

where $\zeta(t)$ is Brownian motion and the operations to be performed are readily justified.

We say a channel has finite memory δ if $h(t) = 0$, $|t| > \delta$, for some $\delta < \infty$. Obviously this definition distorts the language a little, because it also requires what might be called "finite anticipation." It is mathematically convenient, however, and includes the practical case of a non-anticipative channel with finite memory. All the results to be proved will hold, a fortiori, for non-anticipative channels with finite memory. We shall require in everything that follows that each class $\mathcal{C}$ of channels to be considered has the property that each $h \in \mathcal{C}$ has finite memory δ, where δ is some positive number fixed for the class $\mathcal{C}$; this condition will be referred to by saying that $\mathcal{C}$ has finite memory δ.

Let $\mathcal{C}$ be a collection of channels. By a (G, ϵ, T) code for $\mathcal{C}$ we mean a set $\{x_1, x_2, \ldots, x_G\}$ of distinct signals over $[-T, T]$ and a set $\{B_1, B_2, \ldots, B_G\}$ of G disjoint sets of the output space (of real-valued functions over $[-T, T]$) such that

$$\text{(i)} \quad \int_{-T}^{T} x_i^2(t)\, dt \leq 2T, \qquad i = 1, 2, \ldots, G$$

and

$$\text{(ii)} \quad P_h(y(t) \in B_i^c \mid x_i) \leq \epsilon, \quad i = 1, 2, \ldots, G, \quad \forall h \in \mathcal{C},$$

where $P_h(A \mid x_i)$ denotes the probability of the event A given that the input signal is x_i and the channel is h. Here (i) represents the average input power constraint and (ii) the condition that the probability of error is to be less than ϵ uniformly for all code words x_i and all channels $h \epsilon \mathcal{C}$.

We say that $R \geq 0$ is an attainable rate for $\mathcal{C}$ if there is a sequence of codes $\{(e^{T_n R}, \epsilon_n, T_n)\}$ such that $\lim_{n\to\infty} T_n = +\infty$ and $\lim_{n\to\infty} \epsilon_n = 0$. The supremum of all attainable rates for $\mathcal{C}$ is denoted by $\hat{C}(\mathcal{C})$.

As in Ref. 1 we define the capacity C of a class of channels formally in terms of quantities characterizing the class, and then prove that $C = \hat{C}$. Usually when this is done C is defined first in terms of the mutual information. However it is inconvenient here to talk about the mutual information directly, so we go to an expression that is analogous to that for the expected value of the mutual information for a class of discrete Gaussian channels (see 1). Let $\tilde{h}(\nu)$ be the Fourier transform of h(t), which will always exist because of conditions to be imposed on the channels. Let $\tilde{s}(\nu)$ be the spectral density of a real-valued stationary process with mean zero, variance bounded by one and integrable autocorrelation function (or, alternatively, $\tilde{s}(\nu)$ is a real-valued, non-negative, even function satisfying

$$\int_{-\infty}^{\infty} \tilde{s}(\nu)\, d\nu \leq 1$$

and with integrable inverse Fourier transform); let $\mathcal{A}$ be the set of all such $\tilde{s}(\nu)$. Then the capacity of the class $\mathcal{C}$ is defined to be

$$C(\mathcal{C}) = \sup_{\tilde{s} \in \mathcal{A}} \inf_{h \in \mathcal{C}} \int_{-\infty}^{\infty} \log \left(1 + \frac{|\tilde{h}(\nu)|^2 \tilde{s}(\nu)}{N}\right) d\nu . \tag{2}$$

The object of this paper is to show that under certain integrability conditions on the functions h and certain compactness conditions on the classes $\mathcal{C}$, both of which are enunciated in the next section, $\hat{C}(\mathcal{C}) = C(\mathcal{C})$.

3. Notation and Further Conditions on $\mathcal{C}$

Let L_p denote $L_p(-\infty, \infty)$, for $1 \leq p \leq \infty$, where $L_p(a, b)$ is the L_p space of complex-valued functions p-integrable Lebesgue on the interval (a, b). Let $L_p(T)$ denote $L_p(-T, T)$, for $1 \leq p \leq \infty$. If $f \in L_p$ or $L_p(T)$, then $||f||_p$ denotes the norm of f in that space. If f, $g \in L_2$ or $L_2(T)$, their inner product is written (f, g). An operator on a space $\mathfrak{X}$ is a continuous linear transformation of $\mathfrak{X}$ into itself. P_T is to denote the projection operator on L_p, $1 \leq p \leq \infty$, defined by

$$(P_T x)(t) = x(t), \quad |t| \leq T$$
$$= 0 \quad , \quad |t| > T \tag{3}$$

for all $x \in L_p$.

Let $f \in L_1$. Then its Fourier transform $\tilde{f}$ is given by

$$\tilde{f}(\nu) = \int_{-\infty}^{\infty} f(t)\, e^{i 2\pi \nu t}\, dt ,$$

and $\tilde{f}$ is a continuous, bounded function. If moreover $f \in L_2$, then $\tilde{f} \in L_2$ and the operator $f \to \tilde{f}$ is an isometry of L_2. For each $T < \infty$, f defines a compact (actually Hilbert Schmidt) operator F_T on $L_2(T)$ given by

$$(F_T x)(t) = \int_{-T}^{T} f(t-\tau)\, x(\tau)\, d\tau, \quad -T \leq t \leq T. \tag{4}$$

Also f defines an operator F on L_2 given by the convolution

$$(F x)(t) = \int_{-\infty}^{\infty} f(t-\tau)\, x(\tau)\, d\tau, \quad -\infty < t < \infty. \tag{5}$$

With a slight abuse of notation we identify the operators F_T and $P_T F P_T$. If f has finite memory δ, then

$$P_T F = P_T F P_{T+\delta} \quad \text{and} \quad F P_T = P_{T+\delta} F P_T. \tag{6}$$

If A is an operator on $L_2(T)$ or L_2 then A^* will denote its adjoint. The operator A^* is defined by

$$(x, Ay) = (A^* x, y)$$

for all $x, y \in L_2(T)$ or L_2, since it is required that A be bounded. If A is a compact symmetric operator its trace, $T_r(A)$, is defined if the sum of the eigenvalues of A converges, and is equal to that sum.

There are certain conditions required on the classes $\mathscr{C}$ of channel weighting functions h which will be needed in the proof of the coding theorem. These are enumerated in the definition to follow and some immediate implications of them are noted. We shall say $\mathscr{C}$ is an admissible class of channels if

(i) $\mathscr{C}$ has finite memory δ.

(ii) Each $h \in L_2$, and $||h||_2 \leq 1$

for all $h \in \mathscr{C}$ (It would be sufficient to take any bound, but there is no loss of generality in taking the bound to be 1.)

(iii) If $\tilde{h}$ is the Fourier transform of h, then

$$\int_{-\infty}^{-A} + \int_{A}^{\infty} |\tilde{h}(\nu)|^2 d\nu \to 0 \quad \text{as} \quad A \to \infty$$

uniformly for all $h \in \mathscr{C}$.

Now, since each h vanishes outside the interval $[-\delta, \delta]$ by (i), it follows from (ii) that $h \in L_1$ and $||h||_1 \leq \sqrt{2\delta}$ for all $h \in \mathscr{C}$. Therefore the Fourier transform $\tilde{h}$ exists not only in the sense of the Plancherel theorem, but also as a bounded continuous function on R^1. It also follows from (i), (ii) and (iii) that $\mathscr{C}$ is a conditionally compact subset of L_2. In fact we show that the functions $\tilde{h}(\nu)$, $h \in \mathscr{C}$, form a conditionally compact subset of L_2. Necessary and sufficient conditions for this are (Ref. 5, p. 298) that the set $\{\tilde{h}(\nu) \mid h \in \mathcal{C}\}$ is bounded, that the condition (iii) stated above is satisfied, and that

$$\int_{-\infty}^{\infty} |\tilde{h}(\nu+\mu) - \tilde{h}(\nu)|^2 \, d\nu \rightarrow 0$$

as $\mu \rightarrow 0$, uniformly in $\mathcal{C}$. But

$$\int_{-\infty}^{\infty} |\tilde{h}(\nu+\mu) - \tilde{h}(\nu)|^2 \, d\nu = \int_{-\infty}^{\infty} |h(t)\, e^{i2\pi\mu t} - h(t)|^2 \, dt$$

$$= \int_{-\delta}^{\delta} |h(t)|^2 \, |e^{i2\pi\mu t} - 1|^2 \, dt \leq 4\pi^2 \mu^2 \delta^2 \, ||h||^2 \leq 4\pi^2 \mu^2 \delta^2$$

We shall also have occasion to consider the set of functions

$$\mathcal{H}(\mathcal{C}) = \{|\tilde{h}(\nu)|^2 \mid \tilde{h} = \text{Fourier transform of } h \in \mathcal{C}\}$$

Since each $\tilde{h} \in L_2$, $\mathcal{H}(\mathcal{C})$ is a subset of L_1. As a subset of L_1, $\mathcal{H}(\mathcal{C})$ is conditionally compact. In fact, the necessary and sufficient conditions that this be so are (Ref. 5., p. 295): $\mathcal{H}(\mathcal{C})$ is a bounded subset of L_1; condition (iii) is satisfied, and

$$\int_{-\infty}^{\infty} \left| |\tilde{h}(\nu)|^2 - |\tilde{h}(\nu+\mu)|^2 \right| d\nu \rightarrow 0$$

as $\mu \rightarrow 0$, uniformly in the class. But

$$\int_{-\infty}^{\infty} \left| |\tilde{h}(\nu)|^2 - |\tilde{h}(\nu+\mu)|^2 \right| d\nu \leq \int_{-\infty}^{\infty} |\tilde{h}^2(\nu) - \tilde{h}^2(\nu+\mu)| \, d\nu$$

$$\leq 2||h|| \left[\int_{-\infty}^{\infty} |\tilde{h}(\nu) - \tilde{h}(\nu+\mu)|^2 \, d\nu \right]^{1/2}$$

which approaches zero as above.

4. Preliminary Lemmas

In this section we obtain certain results that allow us to extend the application of the Kac, Murdock, Szego (KMS)[4] theorem on the asymptotic behavior of the eigenvalues of a type of integral operator on $L_2(T)$ as $T \to \infty$. We need to apply the KMS theorem to compact (in an appropriate topology) classes of operators, instead of single operators, and we need to apply it to certain truncated operators which do not meet the conditions of that theorem. For convenience we state the theorem we are referring to:

KMS Theorem. Let $\rho \in L_1$ be an even function and suppose that its Fourier transform $\tilde{\rho}$ also belongs to L_1. Let ρ define the self-adjoint operators R_T and R on $L_2(T)$, L_2 respectively ($R_T = P_T R P_T$). Let a, b, $a < b$, be real numbers and let $N(R_T, a, b)$ denote the number of eigenvalues of R_T which lie in the interval (a, b). If

(i) $0 \notin (a, b)$

and

$$\text{(ii)} \quad \mu\{\nu \mid \tilde{\rho}(\nu) = a \quad \text{or} \quad \tilde{\rho}(\nu) = b\} = 0 ,$$

then

$$\lim_{T \to \infty} \frac{1}{2T} N(R_T, a, b) = \mu\{\nu \mid \tilde{\rho}(\nu) \in (a, b)\} .$$

$\mu\{E\}$ denotes the Lebesgue measure of E.

Throughout this section $\mathcal{C}$ will denote an admissible class of channels as defined in Section 3. Also throughout this section s will denote the covariance function of a stationary stochastic process with mean zero with the additional property that $s \in L_1$. By known properties (Ref. 6, Thm. 9) of positive semi-definite functions it follows that $\tilde{s} \in L_1$.

For each $T < \infty$, $h \in \mathcal{C}$ and s as above let us define the operators $H_T = P_T H P_T$ and $S_T = P_T S P_T$ where H and S are defined in terms of h and s as in Eq. (5). H_T is then a compact operator, so the positive semi-definite operator

$$W_T = P_T H P_T S P_T H^* P_T = H_T S H_T^* \tag{7}$$

is also compact. Finally we define

$$Q_T = P_T H S H^* P_T = P_T Q P_T . \tag{8}$$

Q and Q_T are positive semi-definite, and Q_T is compact by virtue of the fact that $Q = HSH^*$ is a convolution operator with kernel in L_2. Indeed, let $q = h * s * \underset{\sim}{h}$, where $*$ means convolution, and $\underset{\sim}{h}(t) = h(-t)$. Then $q \in L_1 \cap L_2$ (since $h \in L_1 \cap L_2$, $s \in L_1$), and its Fourier transform is

$$\tilde{q}(\nu) = |\tilde{h}(\nu)|^2 \tilde{s}(\nu), \tag{9}$$

which also belongs to L_1 since $\tilde{s}(\nu)$ is bounded. We note, therefore, that the KMS theorem applies to Q_T. Lemmas 1, 2 and 3 extend the application to W_T. Lemma 1 is included as a reminder of an essentially well-known fact.

Lemma 1.

Let A be a positive semi-definite compact self-adjoint operator, and P a projection operator on a Hilbert space. Let $B = PAP$. Let $a_1 \geq a_2 \geq \ldots$, and $b_1 \geq b_2 \geq \ldots$ be the eigenvalues of A and B respectively. Then $a_i \geq b_i$, $i = 1, 2, \ldots$.

Proof: Let $A^{1/2}$ be the positive square root of A; $A^{1/2}$ is compact. Let $C = PA^{1/2}$, so that $CC^* = B$. Now the eigenvalues pf $B = CC^* = PAP$ are the same as the eigenvalues of $C^*C = A^{1/2}PA^{1/2}$, although the invariant subspaces are different. Since $A = A^{1/2}A^{1/2}$ dominates $A^{1/2}PA^{1/2}$ the conclusion follows by a standard theorem (Ref. 7, p. 239).

Lemma 2.

Let the operators W_T and Q_T be as defined in Eqs. 7 and 8 and $\tilde{q}(\omega)$ be as defined in Eq. 9. Then

(i) $\frac{1}{T} \mathrm{Tr}(Q_T) = 2 \int_{-\infty}^{\infty} \tilde{q}(\nu)\, d\nu$ for all $T < \infty$

(ii) $\lim_{T\to\infty} \frac{1}{T} \mathrm{Tr}(W_T) = 2 \int_{-\infty}^{\infty} \tilde{q}(\nu)\, d\nu = 2 \int_{-\infty}^{\infty} |\tilde{h}(\nu)|^2 \tilde{s}(\nu)\, d\nu$

uniformly for all $h \in \mathcal{B}$ and all $s \in \mathcal{S}$.

Proof. Since $\tilde{s}(\nu) \geq 0$ one can take $\tilde{r}(\nu) = \sqrt{\tilde{s}(\nu)} \geq 0$; then $\tilde{r}(\nu) = \tilde{r}(-\nu)$, and $\tilde{r} \in L_2$. Let R be the convolution operator determined by r, the inverse Fourier transform of $\tilde{r}$, as in Eq. 5. R is positive semi-definite and $R^2 = S$, so

$$Q_T = P_T H R R H^* P_T = (P_T HR)(P_T HR)^*$$

and

$$W_T = (P_T H P_T R)(P_T H P_T R)^*. \tag{10}$$

One has,

$$(P_T HR\,x)(t) = \int_{-\infty}^{\infty} \left\{ I_T(t) \int_{-\infty}^{\infty} h(t-u)r(u-v)\, du \right\} x(v)\, dv$$

$$(RH^*P_T x)(t) = \int_{-T}^{T} \left\{ \int_{-\infty}^{\infty} r(t-u)\, h(v-u)\, du \right\} x(v)\, dv \quad .$$

and

$$(P_T H P_T R x)(t) = \int_{-\infty}^{\infty} \left\{ I_T(t) \int_{-T}^{T} h(t-u)\, r(u-v)\, du \right\} x(v)\, dv$$

$$(R P_T H^* P_T x)(t) = \int_{-T}^{T} \left\{ \int_{-T}^{T} r(t-u)\, h(v-u)\, du \right\} x(v)\, dv$$

where I_T is the indicator function of the interval $[-T, T]$.

If we let $k(t, w)$ be the kernel of the operator Q_T, it follows that

$$k(t, w) = I_T(t)\, I_T(w) \int \int_{-\infty}^{\infty} \int h(t-u)\, r(u-v)\, h(w-u')\, r(v-u')\, du\, du'\, dv \; ,$$

whence, using the fact that r is even,

$$\frac{1}{T} \operatorname{Tr}(Q_T) = \frac{1}{T} \int_{-T}^{T} k(t, t)\, dt$$

$$= \frac{1}{T} \int_{-T}^{T} || h * r ||^2\, dt = 2 \int_{-\infty}^{\infty} |\tilde{h}(\nu)|^2\, |\tilde{r}(\nu)|^2\, d\nu \; .$$

The existence of the integrals and interchanges of order of integration all follow from the conditions, $h \in L_1 \cap L_2$, $\tilde{r} \in L_2$, and $\tilde{r}(\nu)$ bounded. This proves (i).

We actually prove (ii) under weaker conditions on $\mathcal{b}$, which will be stated below. Observe first that

$$\frac{1}{T}\,\mathrm{Tr}(W_T) = \frac{1}{T}\int_{-T}^{T} dt \int_{-\infty}^{\infty}\left[\int_{-T}^{T} h(t-u)\, r(u-v)\, du\right]^2 dv .$$

For convenience, put

$$A = \int_{-\infty}^{\infty} h(t-u)\, r(u-v)\, du, \quad B = \int_{-T}^{T} h(t-u)\, r(u-v)\, du .$$

Then

$$\frac{1}{T}(\mathrm{Tr}\, Q_T - \mathrm{Tr}\, W_T) = \frac{1}{T}\int_{-T}^{T} dt \int_{-\infty}^{\infty} (A^2 - B^2)\, dv$$

and

$$\frac{1}{T}\left|\mathrm{Tr}\, Q_T - \mathrm{Tr}\, W_T\right| \leq \frac{1}{T}\int_{-T}^{T} dt \int_{-\infty}^{\infty} |A+B| \cdot |A-B|\, dv \tag{10}$$

One has,

$$|A+B| \leq 2\int_{-\infty}^{\infty} |h(t-u)\, r(u-v)| du = \varphi(t-v) \tag{11}$$

where φ, which is defined by this equation, belongs to L_2, and in fact satisfies

$$||\varphi||_2 \leq ||h||_1 ||r||_2 = ||h||_1 ||\sqrt{s}\,||_2 \leq ||h||_1$$

Also,

$$|A - B| \leq \int_{|u|>T} |h(t-u)\, r(u-v)|\, du$$

$$\leq \left[\int_{|u|>T} |h(t-u)|\, du\right]^{1/2} \left[\int_{|u|>T} |h(t-u)|\; |r(u-v)|^2\, du\right]^{1/2} . \tag{12}$$

The second factor on the right side of Eq. 12 is dominated by $|\alpha(t-v)|^{1/2}$ where

$$\alpha(t-v) = \int_{-\infty}^{\infty} |h(t-u)|\; |r(u-v)|^2\, du$$

and

$$||\alpha^{1/2}||_2^2 = ||\alpha||_1 \leq ||h||_1 ||r^2||_1 = ||h||_1 ||s||_1 \leq ||h||_1 .$$

Hence, from Eqs. (10), (11) and (12),

$$\frac{1}{T}\,|\mathrm{Tr}\, Q_T - \mathrm{Tr}\, W_T| \leq \frac{1}{T}\int_{-T}^{T} dt \int_{-\infty}^{\infty} dv\, \varphi(t-v)\, \alpha^{1/2}(t-v)\left[\int_{|u|>T} |h(t-u)|\, du\right]^{1/2}$$

$$\leq \frac{1}{T}\int_{-T}^{T} dt \left\{ ||\varphi||_2\, ||\alpha^{1/2}||_2 \left[\int_{|u|>T} |h(t-u)|\, du\right]^{1/2}\right\}$$ (cont'd)

cont'd.

$$\leq \|h\|_1^{3/2} \frac{1}{T} \int_{-T}^{T} dt \left[\int_{|u|>T} |h(t-u)|\, du \right]^{1/2} . \tag{13}$$

Now, given $\epsilon > 0$, suppose there is a number $A(\epsilon) > 0$, not depending on h such that

$$\int_{|\tau| > A(\epsilon)} |h(\tau)|\, d\tau \leq \epsilon^2 \tag{14}$$

for all $h \in \mathcal{C}$. If $\mathcal{C}$ is admissible this condition is satisfied a fortiori since the integral vanishes on the set $|\tau| > \delta$. Then, for $|t| < T - A(\epsilon)$,

$$\int_{|u|>T} |h(t-u)|\, du \leq \epsilon^2 ,$$

and we have from Eq. 13 and 14

$$\frac{1}{T} |\mathrm{Tr}\, Q_T - \mathrm{Tr}\, W_T| \leq \|h\|_1^{3/2} \left\{ \int_{-(T-A)}^{T-A} \frac{\epsilon}{T} dt + \int_{-T}^{-(T-A)} \left[\|h\|_1^{1/2} \frac{dt}{T} \right] \right.$$

$$\left. + \int_{T-A}^{T} \left[\|h\|_1^{1/2} \frac{dt}{T} \right] \right\} = \|h\|_1^{3/2} \left\{ \frac{2(T-A(\epsilon))\,\epsilon}{T} + \frac{2A(\epsilon)}{T} \|h\|_1^{1/2} \right\} .$$

If T is taken greater than $\frac{A(\epsilon)}{\epsilon}$, then

$$\frac{1}{T} | \mathrm{Tr}\, Q_T - T_r\, W_T | \leq ||h||_1^{3/2} (2 + 2\, ||h||_1^{1/2}) \epsilon .$$

Since $||h||_1$ is less than a fixed constant, this proves the lemma. We have proved in fact the stronger result:

Lemma 2a.

Let $\mathcal{B}'$ be a set of functions $h(t)$, $-\infty < t < \infty$, such that $\mathcal{B}'$ is a bounded subset of L_1 and such that

$$\lim_{A\to\infty} \int_{|\tau| > A} |h(\tau)|\, d\tau = 0$$

uniformly in $\mathcal{B}'$. Then, with the notations of Lemma 2,

$$\lim_{T\to\infty} \frac{1}{T} \mathrm{Tr}\, (W_T) = 2 \int_{-\infty}^{\infty} \tilde{q}(\nu)\, d\nu$$

uniformly for $h \in \mathcal{B}'$ and $s \in \mathcal{A}$.

Lemma 3.

Let $0 < a < b < \infty$ and suppose that

$$\mu\{\nu\,|\tilde{q}(\nu) = a\} = \mu\{\nu\,|\tilde{q}(\nu) = b\} = 0.$$

Then

(i) $\lim_{T\to\infty} \frac{1}{T} N(Q_T, a, b) = \mu\{\nu|\tilde{q}(\nu) \in (a, b)\}$

(ii) $$\lim_{T\to\infty} \frac{1}{T}[N(W_T, a, b) - N(Q_T, a, b)] = 0$$

(iii) $$\lim_{T\to\infty} \frac{1}{T} N(W_T, a, b) = \mu\{\nu | \tilde{q}(\nu) \in (a, b)\}$$

Proof: The assertion (i) is given by the KMS theorem, and (iii) follows from (i) and (ii), so it is sufficient to prove (ii).

Since h has finite memory δ,

$$Q_T = P_T H S H^* P_T = P_T H P_{T+\delta} S P_{T+\delta} H^* P_T$$

$$= P_T [P_{T+\delta} H P_{T+\delta} S P_{T+\delta} H^* P_{T+\delta}] P_T$$

$$= P_T W_{T+\delta} P_T \tag{15}$$

If $w_1 \geq w_2 \geq w_3 \geq \cdots$ are the eigenvalues of $W_{T+\delta}$ and $q_1 \geq q_2 \geq q_3 \geq \cdots$ are the eigenvalues of Q_T, then by Lemma 1,

$$\omega_i \geq q_i \tag{16}$$

It is evident that

$$\lim_{T\to\infty} \frac{1}{T} Tr(W_T) = \lim_{T\to\infty} \frac{1}{T} Tr(W_{T+\delta})$$

so that from Lemma 2 we obtain

$$\lim_{T\to\infty} \frac{1}{T} \mathrm{Tr}(Q_T) = \lim_{T\to\infty} \frac{1}{T} \mathrm{Tr}(W_{T+\delta}) . \tag{17}$$

From Eqs. 16 and 17 it follows [see Gallager[3], Lemma 8.5.3] that

$$\lim_{T\to\infty} \frac{1}{T} N(Q_T, a, b) = \lim_{T\to\infty} \frac{1}{T} N(W_{T+\delta}, a, b) . \tag{18}$$

But evidently,

$$\lim_{T\to\infty} \frac{1}{T} N(W_T, a, b) = \lim_{T\to\infty} \frac{1}{T} N(W_{T+\delta}, a, b) . \tag{19}$$

Combining Eqs. 18 and 19 gives Part (ii) of the lemma.

The next two lemmas build on Lemmas 2 and 3 to give the results needed for obtaining a limiting expression for the average mutual information for the class $\mathcal{C}$.

Lemma 4.

Let $h \in \mathcal{C}$, and H be the corresponding operator as given by Eq. 5. Let Q_T, W_T then be given as in Eq. 7 and 8 and denote their eigenvalues by $q_1(T) \geq q_2(T) \geq \dots$, and $w_1(T) \geq w_2(T) \geq \dots$, respectively. Let f be a continuous monotone increasing real-valued function on the real numbers which satisfies $f(0) = 0$, $f(x) \geq k_1 x$ in some neighborhood of 0 and $|f(x) - f(y)| \leq k|x - y|$ for all $x, y \in R$ for some $k < \infty$. Then,

$$\lim_{T\to\infty} \frac{1}{2T} \sum_{i=1}^{\infty} f(q_i(T)) = \lim_{T\to\infty} \frac{1}{2T} \sum_{i=1}^{\infty} f(w_i(T))$$

$$= \int_{-\infty}^{\infty} f(|\tilde{h}(\nu)|^2 \tilde{s}(\nu))\, d\nu \ .$$

Proof. For any $\epsilon > 0$, let

$$S(T,\epsilon) = \frac{1}{2T} \sum_{\{f(q_i) > \epsilon\}} f(q_i(T))$$

where the summation is over all i for which $f(q_i(T)) > \epsilon$. From a standard argument that involves bounding both the sums and the integral from above and below by the integrals of simple functions and using the KMS theorem, it follows that

$$\lim_{T\to\infty} S(T,\epsilon) = \int_{E_\epsilon} f(|\tilde{h}(\nu)|^2 \tilde{s}(\nu))\, d\nu \tag{20}$$

where

$$E_\epsilon = \{ \nu \mid f(|\tilde{h}(\nu)|^2 \tilde{s}(\nu) > \epsilon \} .$$

(This argument is given in detail in Ref. 8 for a special case, and the extension is straightforward). One then has that

$$\lim_{\epsilon \to 0} \lim_{T\to\infty} S(T,\epsilon) = \lim_{\epsilon\to 0} \int_{E_\epsilon} f(|\tilde{h}(\nu)|^2 \tilde{s}(\nu))\, d\nu$$

$$= \int_{-\infty}^{\infty} f(|\tilde{h}(\nu)|^2 \tilde{s}(\nu))\, d\nu$$

since the conditions on f guarantee that $f(|\tilde{h}(\nu)|^2 \tilde{s}(\nu)) \epsilon L_1$. To get the limit relation asserted in the Lemma requires some additional argument, however.

First, since Q_T has finite trace, we have for all $T > 0$,

$$\frac{1}{2T} \sum_{q_i > \epsilon} q_i(T) + \frac{1}{2T} \sum_{q_i \leq \epsilon} q_i(T) = \frac{1}{2T} \sum_i q_i(T)$$

By an argument like the one indicated just above,

$$\lim_{T\to\infty} \frac{1}{2T} \sum_{q_i > \epsilon} q_i(T) = \int_{F_\epsilon} |\tilde{h}(\nu)|^2 \tilde{s}(\nu)\, d\nu$$

where

$$F_\epsilon = \{\nu \mid |\tilde{h}(\nu)|^2 \tilde{s}(\nu) > \epsilon\}.$$

By Lemma 2,

$$\lim_{T\to\infty} \frac{1}{2T} \sum_i q_i(T) = \int_{-\infty}^{\infty} |\tilde{h}(\nu)|^2 \tilde{s}(\nu)\, d\nu$$

(actually the left side equals the integral for all T). Hence,

$$\lim_{T\to\infty} \frac{1}{2T} \sum_{q_i < \epsilon} q_i(T)$$

exists. Furthermore, since

$$\lim_{\epsilon \to 0} \int_{F_\epsilon} |\tilde{h}(\nu)|^2 \tilde{s}(\nu)\, d\nu = \int_{-\infty}^{\infty} |\tilde{h}(\nu)|^2 \tilde{s}(\nu)\, d\nu ,$$

it follows that given arbitrary $\epsilon_1 > 0$, for all sufficiently small ϵ,

$$\lim_{T\to\infty} \frac{1}{2T} \sum_{q_i \leq \epsilon} q_i(T) \leq \epsilon_1 .$$

Hence, by the conditions on f,

$$\overline{\lim_{T\to\infty}} \frac{1}{2T} \sum_{q_i \leq \epsilon} f(q_i(T)) \leq k\epsilon_1$$

for all sufficiently small ϵ ; and finally

$$\overline{\lim_{T\to\infty}} \frac{1}{2T} \sum_{f(q_i) \leq \epsilon} f(q_i(T)) \leq k\epsilon_1 \tag{21}$$

for all sufficiently small ϵ . Then,

$$\overline{\lim_{T\to\infty}} \frac{1}{2T} \sum_{i=1}^{\infty} f(q_i(T)) \leq \overline{\lim_{T\to\infty}} S(T,\epsilon) + \overline{\lim_{T\to\infty}} \frac{1}{2T} \sum_{\{f(q_i)\leq\epsilon\}} f(q_i(T))$$

for any $\epsilon > 0$; hence by Eqs. 20 and 21

$$\overline{\lim_{T\to\infty}} \frac{1}{2T} \sum_{i=1}^{\infty} f(q_i(T)) \leq \int_{-\infty}^{\infty} f(|\tilde{h}(\nu)|^2 \tilde{s}(\nu))\, d\nu + k\epsilon_1$$

for arbitrary $\epsilon_1 > 0$. But

$$\underline{\lim_{T\to\infty}} \frac{1}{2T} \sum_{i=1}^{\infty} f(q_i(T)) \geq \int_{E_\epsilon} f(|\tilde{h}(\nu)|^2 \tilde{s}(\nu))\, d\nu \qquad \forall\ \epsilon > 0$$

Hence,

$$\lim_{T\to\infty} \frac{1}{2T} \sum_{i=1} f(q_i(T)) = \int_{-\infty}^{\infty} f(|\tilde{h}(\nu)|^2 \tilde{s}(\nu))\, d\nu .$$

The proof that

$$\lim_{T\to\infty} \frac{1}{2T} \sum_{i=1}^{\infty} f(w_i(T)) = \int_{-\infty}^{\infty} f(|\tilde{h}(\nu)|^2 \tilde{s}(\nu))\, d\nu$$

is identical, if one uses the extension of the KMS theorem given by Lemma 3.

Lemma 5.

Let $\mathcal{H}(\mathcal{C})$ be the class of $|\tilde{h}(\nu)|^2$, $h \in \mathcal{C}$, as defined in Section 3. Then it is a conditionally compact subset of L_1. Let f be a real-valued function satisfying the conditions imposed in Lemma 4. Since Q_T, and W_T depend on h, we can define functions $q_T : \mathcal{C} \to R$ and $w_T : \mathcal{C} \to R$ by

$$q_T(h) = \frac{1}{2T} \sum_{i=1}^{\infty} f(q_i(T)) \qquad (22)$$

and

$$w_T(h) = \frac{1}{2T} \sum_{i=1}^{\infty} f(w_i(T)) ,$$

where $q_i(T)$ and $w_i(T)$ are as defined in Lemma 4. Then,

(i) $$\lim_{T \to \infty} q_T(h) = \int_{-\infty}^{\infty} f(|\tilde{h}(\nu)|^2 \tilde{s}(\nu))\, d\nu$$

uniformly for $h \in \mathcal{C}$.

(ii) $$\lim_{T \to \infty} (w_T(h) - q_T(h)) = 0$$

uniformly for $h \in \mathcal{C}$.

(iii) $\lim_{T\to\infty} w_T(h) = \int_{-\infty}^{\infty} f(|\tilde{h}(\nu)|^2 \tilde{s}(\nu))d\nu$

uniformly for $h \in \mathcal{b}$.

<u>Proof:</u> Clearly (iii) follows from (i) and (ii).

We first prove (i). It is pointed out above that HSH^* is a convolution operator whose kernel has Fourier transform $\tilde{q}(\nu) = |\tilde{h}(\nu)|^2 \tilde{s}(\nu)$. Therefore Q_T, and consequently q_T, may be regarded as a function of $|\tilde{h}(\nu)|^2 \epsilon\, \mathcal{H}(\mathcal{b})$, and we sometimes write $q_T(|\tilde{h}(\nu)|^2)$, although it is an abuse of notation. Since it is already known from Lemma 4 that $q_T(h)$ converges to the indicated limit, and since $\mathcal{H}(\mathcal{b})$ is a conditionally compact subset of L_1, it suffices to show (by the Arzelà-Ascoli theorem) that the family of functions $\{q_T\}$, $0 < T < \infty$, each q_T considered as a mapping from $\mathcal{H}(\mathcal{b})$ to R, is equicontinuous.

Let $|\tilde{h}_1(\nu)|^2, |\tilde{h}_2(\nu)|^2 \epsilon\, \mathcal{H}(\mathcal{b})$ and let $T < \infty$. Define an L_1 function $\tilde{\varphi}$ by

$$\tilde{\varphi}(\nu) = \min\{|\tilde{h}_1(\nu)|^2 \tilde{s}(\nu), |\tilde{h}_2(\nu)|^2 \tilde{s}(\nu)\}$$

Let $\varphi(t)$ be the inverse Fourier transform of $\tilde{\varphi}(\nu)$. Let Φ be the convolution operator on L_2 defined by φ. Put $\Phi_T = P_T \Phi P_T$, and denote the (real, nonnegative) eigenvalues of Φ_T by $\varphi_1(T) \geq \varphi_2(T) \geq \cdots$. The function q_T is defined for $\varphi(t)$ through Eq. 22, even though φ may not belong to $\mathcal{b}$.

Then, for $j = 1, 2$, $\Phi_T \leq Q_{j,T}$ so that $\varphi_i(T) \leq q_{j,i}(T)$, $i = 1, 2, \ldots,$ where $Q_{j,T}$ is the operator defined by $|\tilde{h}_j(\omega)|^2$ and $q_{j,i}(T)$ are its eigenvalues. Therefore, for $j = 1, 2$

$$|q_T(|\tilde{h}_j|^2) - q_T(\tilde{\varphi}/\tilde{s})|$$

$$= \frac{1}{T} \left| \sum_{i=1}^{\infty} f(q_{j,i}(T)) - f(\varphi_i(T)) \right|$$

$$\leq \frac{1}{T} \sum_{i=1}^{\infty} | f(q_{j,i}(T)) - f(\varphi_i(T))|$$

$$\leq \frac{k}{T} \sum_{i=1}^{\infty} |q_{j,i}(T) - \varphi_i(T)|$$

$$= \frac{k}{T} \sum_{i=1}^{\infty} [q_{j,i}(T) - \varphi_i(T)] = \frac{k}{T} [\operatorname{Tr}(Q_{j,T}) - \operatorname{Tr}(\Phi_T)]$$

$$= 2k \int_{-\infty}^{\infty} [|\tilde{h}(\nu)|^2 \tilde{s}(\nu) - \tilde{\varphi}(\nu)] \, d\nu$$

$$\leq 2k \int_{-\infty}^{\infty} \left| |\tilde{h}_1(\nu)|^2 - |\tilde{h}_2(\nu)|^2 \right| \tilde{s}(\nu) \, d\nu \, .$$

An application of the triangle inequality then gives

$$\left|q_T(|\tilde{h}_1|^2) - q_T(|\tilde{h}_2|^2)\right| \leq 4k \int_{-\infty}^{\infty} \left| |\tilde{h}_1(\nu)|^2 - |\tilde{h}_2(\nu)|^2 \right| \tilde{s}(\nu) \, d\nu$$
$$\leq 4k \max(\tilde{s}(\nu)) \| \tilde{h}_1^2 - \tilde{h}_2^2 \|_1$$

so the family of functions q_T on $\mathcal{H}(\mathcal{C})$ is equicontinuous for $0 < T < \infty$.

To prove (ii) we recall that $Q_T = P_T W_{T+\delta} P_T$ and that therefore, by Lemma 1, $q_i(T) \leq w_i(T+\delta)$ for $i = 1, 2, \ldots$. Then

$$|w_T(h) - q_T(h)| = \frac{1}{2T} \left| \sum_{i=1}^{\infty} [f(w_i(T) - f(q_i(T))] \right|$$

$$\leq \frac{1}{2T} \left| \sum_{i=1}^{\infty} [f(w_i(T+\delta)) - f(q_i(T))] \right|$$

$$+ \frac{1}{2T} \left| \sum_{i=1}^{\infty} [f(w_i(T+\delta)) - f(w_i(T))] \right| .$$

The first term on the right side of this inequality is dominated by

$$\frac{k}{2T} \sum_{i=1}^{\infty} [w_i(T+\delta) - q_i(T)] = \frac{k}{2T} [\mathrm{Tr}(W_{T+\delta}) - \mathrm{Tr}(Q_T)]$$

which converges to 0 uniformly over $\mathcal{C}$ as $T \to \infty$ by Lemma 2. The second term obviously converges to 0 uniformly over $\mathcal{C}$, so the lemma is proved.

As a corollary to Lemma 5 we have :

<u>Theorem 1</u>

Let $\mathcal{C}$ be a collection of channels h with finite memory δ, such that $\mathcal{H}(\mathcal{b})$ is a conditionally compact subset of L_1. Let $s(t)$ be a covariance function belonging to L_1 with Fourier transform $\tilde{s}(\nu)$. For each $T < \infty$, $h \in \mathcal{b}$, let W_T be the self-adjoint positive semi-definite operator

$$W_T = P_T H P_T S P_T H^* P_T$$

with eigenvalues $w_1(T) \geq w_2(T) \geq \cdots$. Let N be a fixed positive number. Then,

$$\lim_{T\to\infty} \frac{1}{2T} \sum_{i=1}^{\infty} \log\left(1 + \frac{w_i(T)}{N}\right) = \int_{-\infty}^{\infty} \log\left(1 + \frac{|\tilde{h}(\nu)|^2 \tilde{s}(\nu)}{N}\right) d\nu \tag{23}$$

uniformly over $\mathcal{b}$.

Proof: The function $f(x) = \log(1 + \frac{x}{N})$ satisfies the conditions required for Lemma 5.

5. The Coding Theorem

In this section we use Theorem 1 of the previous section and the results of [1] to prove that $\hat{C}(\mathcal{b}) \geq C(\mathcal{b})$ for an admissible class of channels $\mathcal{b}$. Let $R < C(\mathcal{b})$ be fixed. Then it follows from Theorem 1 that there exists $T < \infty$ and $\tilde{s} \in \mathcal{S}$ such that

$$\frac{1}{2(T+\delta)} \sum_{i=1}^{\infty} \log\left(1 + \frac{w_i(T)}{N}\right) > R \tag{24}$$

uniformly over $\mathcal{C}$, where $\delta < \infty$ is the memory of $\mathcal{C}$ and $w_1(T) \geq w_2(T) \ldots$ are the eigenvalues of the operator

$$W_T = P_T H P_T S P_T H^* P_T = (P_T H P_T)(P_T S P_T)(P_T H^* P_T) .$$

Let $\{\varphi_i | 1 \leq i < \infty\}$ be a complete orthonormal basis in $L_2(T)$. Relative to this basis the operators $P_T H P_T$, $P_T S P_T$, and $P_T H^* P_T$ have a representation as infinite-dimensional matrices which we denote by H_T, S_T, and H_T^* respectively. We note that H_T^* is the transpose of H_T and the collection $\mathcal{C}_T$ of matrices H_T form a conditionally compact set in the Hilbert-Schmidt norm. Furthermore, S_T can be considered to be the covariance matrix of an infinite-dimensional random Gaussian vector, and the trace of S_T is less than or equal to $2T$. Finally the additive white noise $z(t)$, $-T \leq t \leq T$, will have the representation

$$z(t) = \sum_{i=1}^{\infty} z^i \varphi_i(t)$$

where $z^1, z^2, \ldots$ are independent identically distributed Gaussian random variables with zero mean and variance N.

Now consider the class of discrete, memoryless, infinite-dimensional Gaussian channels $\mathcal{C}_T = \{H_T\}$. The input vectors to these channels are infinite-dimensional vectors $x = (x^1, x^2, \ldots)$ and the output $y = (y^1, y^2, \ldots)$ corresponding to the channel H_T and input x is given by

$$y = H_T x + z$$

where

$z = (z^1, z^2, \ldots)$, and $z^1, z^2, \ldots$ are independent identically distributed Gaussian random variables with zero mean and variance N. The n-extension of the channel H_T is defined in the usual manner so that it carries an n-sequence of input vectors $u = (x_1, \ldots, x_n)$ into an n-sequence of output vectors $v = (y_1, \ldots, y_n)$ with

$$y_i = H_T x_i + z_i \qquad i = 1, \ldots, n$$

where the z_i are mutually independent. If $x = (x^1, x^2, \ldots)$, we define

$$||x||^2 = \sum_{i=1}^{\infty} |x^i|^2$$

and we impose the average input power constraint on an n-sequence $u = (x_1, \ldots, x_n)$ by requiring that

$$||u||^2 = \sum_{i=1}^{n} ||x_i||^2 \leq n(2T) = 2nT .$$

We define the capacity $C_T(\mathcal{C}_T)$ of the class $\mathcal{C}_T$ of channels by the formula

$$C_T(\mathcal{C}_T) = \sup_{S_T \in \mathcal{A}_T} \inf_{H_T \in \mathcal{C}_T} \frac{1}{2} \sum_{i=1}^{\infty} \log \left(1 + \frac{w_i(T)}{N} \right)$$

where $\mathcal{A}_T$ is the set of all covariance matrices S_T whose trace is dominated by $2T$ and where $w_1(T) \geq w_2(T) \ldots$ are the eigenvalues of the matrix $H_T S_T H_T^*$. From (24) it follows that

$$C_T(\mathcal{C}_T) > (T+\delta)R \,. \tag{25}$$

If $\hat{C}_T(\mathcal{C}_T)$ denotes the supremum of the attainable rates for $\mathcal{C}_T$ (for a precise definition of $\hat{C}_T$ see Ref. (1)), then by Theorem 4 of Ref. (1) we have $\hat{C}_T(\mathcal{C}_T) \geq C_T(\mathcal{C}_T)$ so that from Eq. 25 we obtain

$$\hat{C}_T(\mathcal{C}_T) > (T+\delta)R \,.$$

Therefore there exists $\{e^{(T+\delta)Rn}, \epsilon_n, n\}$ codes for $\mathcal{C}_T$ with $\epsilon_n \to 0$ as $n \to \infty$.

<u>Lemma 6.</u>

If there exists a $\{e^{(T+\delta)Rn}, \epsilon_n, n\}$ code for $\mathcal{C}_T$, then there exists a $\{e^{(T+\delta)Rn}, \epsilon_n, (T+\delta)n\}$ code for $\mathcal{C}$.

<u>Proof.</u> Let $G = e^{(T+\delta)Rn}$, and let the code words and decoding sets for $\mathcal{C}_T$ be $u_1 = (x_{11}, \ldots, x_{1n}), \ldots, u_G = (x_{G1}, \ldots, x_{Gn})$ and $B_1, \ldots, B_G$ respectively. We have $||u_i||^2 \leq 2nT$ for each i and

$$P_{H_T} \{(y_1, \ldots, y_n) \notin B_i | u_i\} \leq \epsilon_n$$

for all $i = 1, \ldots, n$ and all $H_T \in \mathcal{C}_T$. Here $P_{H_T} \{A | u_i\}$ is the probability that the event A occurs when the sequence u_i is the input to the channel H_T. We now proceed to construct a code for $\mathcal{C}$.

Corresponding to each vector $x_{ij} = (x_{ij}^1, x_{ij}^2, \ldots)$ define the function $x_{ij}(t)$, $-T \leq t \leq T$ by,

$$x_{ij}(t) = \sum_{k=1}^{\infty} x_{ij}^k \varphi_k(t) \qquad i = 1, \ldots, G. \quad j = 1, \ldots, n.$$

Now for $i = 1, \ldots, G$ define the function $\hat{u}_i(t)$, $0 \leq t \leq 2n(T+\delta)$ by

$$\hat{u}_i(t) = x_{ij}(t - T - 2(j-1)(T+\delta) - \delta) \text{ for } 2(j-1)(T+\delta) + \delta \leq t \leq 2j(T+\delta) - \delta$$

$$j = 1, \ldots, n.$$

$$= 0 \quad \text{elsewhere.}$$

Next define the functions $u_1(t), \ldots, u_G(t)$ on the interval $-n(T+\delta) \leq t \leq n(T+\delta)$ by,

$$u_i(t) = \hat{u}_i(t + n(T+\delta)).$$

From the construction of the function $u_i(t)$ (see Fig. 1) and the fact that $\mathcal{C}$ has memory δ it is evident that $u_1(t), \ldots, u_G(t)$ are the codewords

Theorem 2.

$$\hat{C}(\mathcal{C}) \geq C(\mathcal{C}) .$$

Proof. Let $R < C(\mathcal{C})$. Then by Lemma 6 there exists a $T < \infty$ and a sequence of $\{e^{(T+\delta)Rn}, \epsilon_n, (T+\delta)n\}$ codes for $\mathcal{C}$ with $\epsilon_n \to 0$ as $n \to \infty$. Therefore $R \leq \hat{C}(\mathcal{C})$.

6. Weak Converse of the Coding Theorem.

In this section we will prove that $\hat{C}(\mathcal{C}) \leq C(\mathcal{C})$. The usual way of proving the weak converse to a coding theorem is the following: One starts with a sequence of codes which yield an attainable rate R. Then, a stochastic process is constructed from the codes with the probability measures determined by the empirical distribution. Next the logarithm of the number of codeword is essentially dominated by the value of the mutual information determined by this stochastic process and the channel. Finally it is shown that the rate of mutual information is dominated by the channel capacity so that R is less than or equal to the channel capacity. Unfortunately, in our case we cannot follow this program completely, because the empirical distribution obtained in the above manner gives rise to a non-stationary covariance function, whereas the capacity was defined by taking the supremum over stationary covariance functions only, so that one cannot immediately assert that the rate of mutual information corresponding to the empirical distribution is dominated by the capacity.

We now proceed to prove the converse in a sequence of steps.

(i) Suppose we are given a $(G, \frac{1}{2}\epsilon, T)$ code for $\mathcal{C}$ i.e., G distinct functions $x_1(t), \ldots, x_G(t)$ defined on the interval $[-T, T]$, satisfying the power constraint

$$\int_{-T}^{T} x_i^2(t)\,dt \leq 2T$$

for all i, and G disjoint Borel subsets $B_1, \ldots, B_G$ of the output space of real-valued functions on $[-T, T]$ such that

$$P_h\{y(t) \in B_i^c \mid x_i(t)\} \leq \frac{1}{2}\epsilon \qquad i = 1, \ldots, G, \quad \forall\, h \in \mathcal{C}.$$

Now let $\mathcal{C}_f$ be a fixed finite subset of $\mathcal{C}$. Then it is clear that if $x_i^n(t)$, $n = 1, 2, \ldots$ is a sequence of <u>continuous</u> functions which converges to x_i in $L_2(T)$ then

$$\lim_{n \to \infty} P_h\{y(t) \in B_i^c \mid x_i^n(t)\} = P_h\{y(t) \in B_i^c \mid x_i(t)\}$$

for each $h \in \mathcal{C}$ and hence the convergence is uniform over the finite set $\mathcal{C}_f$. Thus we can assume that there is a $\{G, \epsilon, T\}$ code for $\mathcal{C}_f$ whose codewords, which we again denote by $x_1, \ldots, x_G$, are continuous functions.

(ii) Next we construct a stochastic process $\xi(t)$, $-T \leq t \leq T$ from the continuous codewords $x_1, \ldots, x_G$ by:

$$P\{\xi(t_1) \leq a_1, \ldots, \xi(t_n) \leq a_n\} = \frac{1}{G} \{\text{number of codewords } x_i \text{ such that } x_i(t_1) \leq a_1, \ldots, x_i(t_n) \leq a_n\}$$

for every finite subset $\{t_1, \ldots, t_n\} \subset [-T, T]$.

The ξ process is extended to the interval $[-T+\delta, T+\delta]$ by defining $\xi(t) = 0$ for $T \leq |t| \leq T + \delta$. Finally the ξ process is extended to a periodic process on the real line, with period $2(T+\delta)$, as follows: we regard ξ as a random function, i.e., as a random variable whose values are real-valued functions defined on the interval $[-T+\delta, T+\delta]$. Let ξ_n, $n = 0, \pm 1, \pm 2, \ldots$ be a sequence of independent random functions each of them having the same distribution as the random function ξ. Now define the ξ process on the line by (here ω is an element of the underlying probability space),

$$\xi(\omega, t) = \xi_0(\omega)(t), \qquad -(T+\delta) \leq t \leq T + \delta$$

$$\xi(\omega, t) = \xi_n(\omega)(t - 2n(T+\delta)), \quad (2n-1)(T+\delta) \leq t \leq (2n+1)(T+\delta)$$

$$n = 1, 2, 3 \ldots$$

and

$$\xi(\omega,t) = \xi_n(\omega)(t+2n(T+\delta)), \ -(2n-1)(T+\delta) \geq t \geq -(2n+1)(T+\delta)$$

$$n = -1, -2, \ldots$$

(iii) For convenience, let $\hat{T} = T + \delta$. For each integer $k \geq 1$ let $x^k(t)$ be the process defined on the interval $[-k\hat{T}, k\hat{T}]$ by,

$$\begin{aligned} x^k(t) &= \xi(t), & |t| &\leq (k-1)\hat{T} \\ &= 0, & (k-1)\hat{T} &\leq |t| \leq k\hat{T}, \end{aligned}$$

and for each $\tau \in [-\hat{T}, \hat{T}]$ let x_τ^k be the translations of the x^k process defined on $[-k\hat{T}, k\hat{T}]$ by,

$$\begin{aligned} x_\tau^k(t) &= x^k(t-\tau) & \text{for} \quad |t-\tau| &\leq k\hat{T} \\ &= 0 & \text{elsewhere, see Fig. 2.} \end{aligned}$$

Finally for each positive integer n let $z_n^k(t)$, $t \in [-k\hat{T}, k\hat{T}]$ be the process defined by

$$z_n^k(t) = \sum_{i=0}^{n} \alpha^i x_{\tau_i}^k(t) \tag{26}$$

where $\tau_i = -\hat{T} + \frac{i}{n} 2\hat{T}$ and $\alpha_n = (\alpha^0, \ldots, \alpha^n)$ is a random vector, independent of the x^k process, and taking on values $(1, 0, \ldots, 0)$, $(0, 1, 0, \ldots, 0)$, $\ldots$, $(0, 0, \ldots, 0, 1)$, each with probability $\frac{1}{n+1}$. Thus

z_n^k is the process obtained from x^k by n random, uniform shifts.

(iv) Now we let $h \epsilon \mathcal{L}_f$ be any fixed channel. If $x^k(t)$, $x_{\tau_i}^k(t)$, $z_n^k(t)$ are the inputs to h, let the corresponding outputs (over the same interval $[-k\hat{T}, k\hat{T}]$) be denoted by $y^k(t)+n^k(t)$, $y_{\tau_i}^k(t) + n^k(t)$, $w_n^k(t)+n^k(t)$ respectively. In these expressions $n^k(t)$ denotes the additive white noise on the interval $[-k\hat{T}, k\hat{T}]$. It is important to note that since the channel in time-invariant, $y_\tau^k(t) = y^k(t-\tau)$; also

$$w_n^k(t) = \sum_{i=0}^{n} \alpha^i y_{\tau_i}^k(t)$$

where $\alpha_n = (\alpha^0, \dots, \alpha^n)$ is the same random vector as in Eq. 26. Furthermore, with probability one, each of the processes $x^k(t)$, $x_{\tau_i}^k(t)$, $y^k(t)$ and $y_{\tau_i}^k(t)$ have exactly $N = (G)^{k-1}$ equiprobable sample functions; whereas with probability one, the processes $z_n^k(t)$, $w_n^k(t)$ have $(n+1)N$ equiprobable sample functions. In the following, when we refer to the sample functions of these processes we mean those which have nonzero probability.

(v) Let the sample functions of $x^k = x_{\tau_0}^k, x_{\tau_1}^k, \dots, x_{\tau_n}^k$ be $\{\varphi_1, \dots, \varphi_N\}$, $\{\varphi_{N+1}, \dots, \varphi_{2N}\}$ $\dots$, $\{\varphi_{nN+1}, \dots, \varphi_{(n+1)N}\}$ respectively. Similarly let the sample functions of $y^k = y_{\tau_0}^k, y_{\tau_1}^k, \dots, y_{\tau_n}^k$ be $\{\psi_1, \dots, \psi_N\}$, $\{\psi_{N+1}, \dots, \psi_{2N}\}, \dots, \{\psi_{nN+1}, \dots, \psi_{(n+1)N}\}$, respectively.

Finally let $\{\varphi_{(n+1)N+i}\}$ $i = 1, 2, \ldots$ be an orthonormal set of functions which span the orthogonal complement in $L_2[-k\hat{T}, k\hat{T}]$ of the space spanned by $\{\varphi_1, \ldots, \varphi_{(n+1)N}\}$, and similarly let the orthonormal set $\{\psi_{(n+1)N+i};\ i = 1, 2, \ldots\}$ span the orthogonal complement in $L_2[-k\hat{T}, k\hat{T}]$ of the space spanned by $\{\psi_1, \ldots, \psi_{(n+1)N}\}$. Relative to these bases we have the following representations:

$$x^k(t) = x^k_{\tau_0}(t) = \sum_{i=1}^{\infty} (x^k)_i\, \varphi_i(t)\ ,$$

$$x^k_{\tau_j}(t) = \sum_{i=1}^{\infty} (x^k_{\tau_j})_i\, \varphi_i(t) \qquad j = 0, \ldots, n,$$

$$z^k_n(t) = \sum_{i=1}^{\infty} (z^k_n)_i\, \varphi_i(t)\ ;$$

$$y^k(t) = y^k_{\tau_0}(t) = \sum_{i=1}^{\infty} (y^k)_i\, \psi_i(t)\ ,$$

$$y^k_{\tau_j}(t) = \sum_{i=1}^{\infty} (y^k_{\tau_j})_i\, \psi_i(t) \qquad j = 0, \ldots, n\ ,$$

$$w^k_n(t) = \sum_{i=1}^{\infty} (w^k_n)_i\, \psi_i(t)\ ;$$

and finally,

$$n^k(t) = \sum_{i=1}^{\infty} (n^k)_i \ \psi_i(t)$$

Remark

In these representations, the coefficients of the basis functions are random variables. For future reference, we note that with probability one $(x^k_{\tau_i})_j = (y^k_{\tau_i})_j = 0$ for $j \notin \{iN+1, \ldots, (i+1)N\}$ and $(z^k_n)j = (w^k_n) = 0$ for $j \notin \{1, \ldots, (n+1)N\}$. We also note that the random vectors $((n^k)_1, \ldots, (n^k)_{(n+1)N})$ and $((n^k)_{(n+1)N+1}, \ldots .)$ are independent.

Now we define

$$I(x^k_{\tau_i}; y^k_{\tau_i} + n^k) = \lim_{\ell \to \infty} I((x^k_{\tau_i})_1, \ldots, (x^k_{\tau_i})_\ell ; (y^k_{\tau_i})_1 + (n^k)_1, \ldots, (y^k_{\tau_i})_\ell + (n^k)_\ell)$$

and

$$I(z^k_n; w^k_n + n^k) = \lim_{\ell \to \infty} I((z^k_n)_1, \ldots, (z^k_n)_\ell ; (w^k_n)_1 + (n^k)_1, \ldots, (w^k_n)_\ell + (n^k)_\ell)$$

where for finite-dimensional random vectors ζ and η $I(\zeta;\eta)$ is the average mutual information between ζ and η; thus $I(\zeta;\eta) = H(\eta) - H(\eta \mid \zeta)$

where H is the entropy function.

Lemma 7

(a) For each $\ell \geq (n+1)N$

$$I((z_n^k)_1, \ldots, (z_n^k)_\ell ; (w_n^k)_1 + (n^k)_1, \ldots, (w_n^k)_\ell + (n^k)_\ell)$$

$$\geq I((x^k)_1, \ldots, (x^k)_\ell ; (y^k)_1 + (n^k)_1, \ldots, (y^k)_\ell + (n^k)_\ell)$$

so that

(b) $I(z_n^k ; w_n^k + n^k) \geq I(x^k ; y^k + n^k)$.

Proof. We first prove that

$$H((w_n^k)_1 + (n^k)_1, \ldots, (w_n^k)_\ell + (n^k)_\ell)$$

$$\geq \frac{1}{n+1} \sum_{i=0}^{n} H((y_{\tau_i}^k)_1 + (n^k)_1, \ldots, (y_{\tau_i}^k)_\ell + (n^k)_\ell) \tag{27}$$

$$P\{(y_{\tau_i}^k)_1 + (n^k)_1 \leq a_1, \ldots, (y_{\tau_i}^k)_\ell + (n^k)_\ell \leq a_\ell\}$$

$$= P\{(y_{\tau_i}^k)_1 + (n^k)_1 \leq a_1, \ldots, (y_{\tau_i}^k)_{(n+1)N} + (n^k)_{(n+1)N} \leq a_{(n+1)N},$$

$$(n^k)_{(n+1)N+1} \leq a_{(n+1)N+1}, \ldots, (n^k)_\ell \leq a_\ell\}$$

(cont'd.)

$$= P\{(y^k_{\tau_i})_1 + (n^k)_1 \le a_1, \ldots, (y^k_{\tau_i})_{(n+1)N} + (n^k)_{(n+1)N} \le a_{(n+1)N}\} \times$$

$$P\{(n^k)_{(n+1)N+1} \le a_{(n+1)N+1}, \ldots, (n^k)_\ell \le a_\ell\}$$

by the remark preceding this lemma. Therefore,

$$H((y^k_{\tau_i})_1 + (n^k)_1, \ldots, (y^k_{\tau_i})_\ell + (n^k)_\ell)$$

$$= H((y^k_{\tau_i})_1 + (n^k)_1, \ldots, (y^k_{\tau_i})_{(n+1)N} + (n^k)_{(n+1)N})$$

$$+ H((n^k)_{(n+1)N+1}, \ldots, (n^k)_\ell). \tag{28}$$

Similarly,

$$H((w^k_n)_1 + (n^k)_1, \ldots, (w^k_n)_\ell + (n^k)_\ell)$$

$$= H((w^k_n)_1 + (n^k)_1, \ldots, (w^k_n)_{(n+1)N} + (n^k)_{(n+1)N})$$

$$+ H((n^k)_{(n+1)N+1}, \ldots, (n^k)_\ell) \tag{29}$$

Also,

$$P\{(y^k_{\tau_i})_1 + (n^k)_1 \le a_1, \ldots, (y^k_{\tau_i})_{(n+1)N} + (n^k)_{(n+1)N} \le a_{(n+1)N}\}$$

(cont'd.)

$$= \frac{1}{N}\sum_{j=1}^{N} P\{(n^k)_1 \le a_1, \ldots, (n^k)_{iN+j} \le a_{iN+j} - 1, \ldots, (n^k)_{(n+1)N} \le a_{(n+1)N}\}$$

because $(y^k_{\tau_i})_j = 0$ for $j \notin \{iN+1, \ldots, (i+1)N\}$ and the process $y^k_{\tau_i}(t)$ has exactly N equiprobable sample functions. Similarly,

$$P\{(w^k_n)_1 + (n^k)_1 \le a_1, \ldots, (w^k_n)_{(n+1)N} + (n^k)_{(n+1)N} \le a_{(n+1)N}\}$$

$$= \frac{1}{(n+1)N}\sum_{j=1}^{(n+1)N} P\{(n^k)_1 \le a_1, \ldots, (n^k)_j \le a_j - 1, \ldots, (n^k)_{(n+1)N} \le a_{(n+1)N}\}$$

$$= \frac{1}{(n+1)}\sum_{i=0}^{n} P\{(y^k_{\tau_i})_1 + (n^k)_1 \le a_1, \ldots, (y^k_{\tau_i})_{(n+1)N} + (n^k)_{(n+1)N} \le a_{(n+1)N}\}$$

From the concavity of the entropy function we therefore have

$$H((w^k_n)_1 + (n^k)_1, \ldots, (w^k_n)_{(n+1)N} + (n^k)_{(n+1)N})$$

$$\ge \frac{1}{n+1}\sum_{i=0}^{n} H((y^k_{\tau_i})_1 + (n^k)_1, \ldots, (y^k_{\tau_i})_{(n+1)N} + (n^k)_{(n+1)N}).$$

Combining the above inequality with (28) and (29) we obtain (27). Now since the noise is additive we also have that

$$H((w_n^k)_1 + (n^k)_1, \ldots, (w_n^k)_\ell + (n^k)_\ell \,\Big|\, (z_k^n)_1, \ldots, (z_k^n)_\ell)$$

$$= H((n^k)_1, \ldots, (n^k)_\ell)$$

$$= H((y_{\tau_i}^k)_1 + (n^k)_1, \ldots, (y_{\tau_i}^k)_\ell + (n^k)_\ell \,\Big|\, (x_{\tau_i}^k)_1, \ldots, (x_{\tau_i}^k)_\ell)$$

Combining this result with Eq. 27 and using the fact that $I(\zeta;\eta) = H(\eta) - H(\eta|\zeta)$ we see that

$$I((z_n^k)_1, \ldots, (z_n^k)_\ell \,;\, (w_n^k)_1 + (n^k)_1, \ldots, (w_n^k)_\ell + (n^k)_\ell)$$

$$\geq \frac{1}{n+1} \sum_{i=0}^{n} I((x_{\tau_i}^k)_1, \ldots, (x_{\tau_i}^k)_\ell \,;\, (y_{\tau_i}^k)_1 + (n^k)_1, \ldots, (y_{\tau_i}^k)_\ell + (n^k)_\ell)$$

But the processes $x_{\tau_i}^k(t)$, $x_{\tau_j}^k(t)$ are identical except for a translation. Furthermore the channel is time-invariant and n is stationary so that

$$I((x^k)_1, \ldots, (x^k)_\ell \,;\, (y^k)_1 + (n^k)_1, \ldots, (y^k)_\ell + (n^k)_\ell)$$

$$= I((x_{\tau_i}^k)_1, \ldots, (x_{\tau_i}^k)_\ell \,;\, (y_{\tau_i}^k)_1 + (n^k)_1, \ldots, (y_{\tau_i}^k)_\ell + (n^k)_\ell)$$

and the lemma is proved.

(vi) Using a standard identity we have

$$I((x^k)_1, \dots, (x^k)_\ell \; ; \; (y^k)_1 + (n^k)_1, \dots, (y^k)_\ell + (n^k)_\ell)$$

$$= H((x^k)_1, \dots, (x^k)_\ell) - H((x^k)_1, \dots, (x^k)_\ell \,\Big|\, (y^k)_1 + (n^k)_1, \dots, (y^k)_\ell + (n^k)_\ell)$$

$$= \log \; G^{k-1} - H((x)_1^k, \dots, (x^k)_\ell \,\Big|\, (y^k)_1 + (n^k)_1, \dots, (y^k)_\ell + (n^k)_\ell)$$

$$\geq (k-1) \log G \; - \; (k-1)\, [\epsilon \log G + \log 2]$$

since x^k has G^{k-1} equiprobable samples and since the error probability is bounded by ϵ (see Ref. 9, p. 187).

(vii) We now obtain a lower bound for $I((x^k)_1, \dots, (x^k)_\ell \; ; (y^k)_1 + (n^k)_1, \dots, (y^k)_\ell + (n^k)_\ell)$. An examination of the construction of the x^k process shows that the x^k process is obtained by transmitting a sequence (of length (k-1)) of codewords each obtained independently and with uniform probability so that

$$I((x^k)_1, \dots, (x^k)_\ell \; ; \; (y^k)_1 + (n^k)_1, \dots, (y^k)_\ell + (n^k)_\ell)$$

$$= (k-1)\, I((x^2)_1, \dots, (x^2)_\ell \; ; (y^2)_1 + (n^2)_1, \dots, (y^2)_1 + (n^2)_\ell)$$

$$= (k-1)\{H((x^2)_1, \dots, (x^2)_\ell) - H((x^2)_1, \dots, (x^2)_\ell) \,\Big|\, (y^2)_1 + (n^2)_1, \dots, (y^2)_\ell + (n^2)_\ell\}$$

$$= (k-1) \;\; \Delta, \text{ say.}$$

Now the x^2 process consists of G equiprobable sample functions so that $H((x^2)_1, \ldots, (x^2)_\ell) = \log G$. An application of two results of Fano (Ref. 9, p. 185, 187) shows that

$$H((x^2)_1, \ldots, (x^2)_\ell \,\Big|\, (y^2)_1 + (n^2)_1, \ldots, (y^2)_\ell + (n^2)_\ell)$$

$$\geq \epsilon \log G + \log 2 \qquad \text{so that}$$

$$\Delta \geq (1-\epsilon) \log G - \log 2$$

and therefore

$$I((x^k)_1, \ldots, (x^k)_\ell ; (y^k)_1 + (n^k)_1, \ldots, (y^k)_\ell + (n^2)_\ell)$$

$$\geq (k-1)\{(1-\epsilon)\log G - \log 2\}$$

Combining the above inequality with Lemma 7 yields

$$I(z_n^k; w_n^k + n^k) \geq (k-1)[(1-\epsilon)\log G - \log 2]$$

If the process z_n^k is replaced by a zero-mean Gaussian process $\bar{z}_n^k$ with the same covariance function as z_n^k and if $\bar{w}_n^k + n^k$ denotes the corresponding output of the channel h then (see Ref. 1, Corrollary to Lemma 9)

$$I(\bar{z}_n^k; \bar{w}_n^k + n^k) \geq (k-1)[(1-\epsilon)\log G - \log 2] \tag{30}$$

Let $Z_n^k(t,s) = E\{(z_k^n(t) - E\,z_k^n(t))(z_k^n(s) - E\,z_k^n(s))\} = E(\bar{z}_k^n(t)\,\bar{z}_k^n(s))$ be the covariance function of z_k^n, and similarly let $X^k(t,s)$, $X_{\tau_i}^k(t,s)$ and $\Xi(t,s)$ be the covariance functions of x^k, $x_{\tau_i}^k$ and ξ respectively. Then

$$X^k(t,s) = \Xi(t,s) \quad \text{for} \quad -(k-1)\hat{T} \le t,s \le (k-1)\hat{T}$$

$$= 0 \qquad \text{for} \quad |t| \ge (k-1)\hat{T} \text{ or } |s| \ge (k-1)\hat{T}$$

and

$$X_{\tau_i}^k(t,s) = X(t-\tau_i,\, s-\tau_i).$$

Consequently,

$$Z_n^k(t,s) = \frac{1}{n+1} \sum_{i=0}^{n} X(t-\tau_i,\, s-\tau_i)$$

$$= \frac{1}{n+1} \sum_{i=0}^{n} X(t - \frac{i}{n}\hat{T},\; s - \frac{i}{n}\hat{T}).$$

Since X is piecewise continuous, the Z_n^k are a convergent sequence of Riemann sums, and we may write

$$Z^k(t,s) = \lim_{n\to\infty} Z_n^k(t,s) = \frac{1}{2\hat{T}} \int_{-\hat{T}}^{\hat{T}} X(t-\tau,\, s-\tau)\, d\tau = \frac{1}{2\hat{T}} \int_{-\hat{T}}^{\hat{T}} \Xi(t-\tau, s-\tau)\, d\tau.$$

Thus if $\bar{z}^k$ denotes the zero-mean Gaussian process with covariance function $Z^k(t, s)$ we see from Eq. 30 that

$$I(\bar{z}^k, \bar{w}^k + n^k) \geq (k-1)\,[(1-\epsilon)\log G - \log 2] \tag{31}$$

Here $\bar{w}^k + n^k$ is the output of the channel h due to the input $\bar{z}^k$. Finally from Eq. (31) we obtain

$$\lim_{k\to\infty} \frac{1}{k}\, I(\bar{z}^k, \bar{w}^k + n^k) \geq (1-\epsilon)\log G - \log 2 \tag{32}$$

Now let $u(t)$ be a zero-mean Gaussian process defined on the line and with covariance function $U(t, s)$ given by

$$U(t, s) = \frac{1}{2\hat{T}} \int_{-\hat{T}}^{\hat{T}} \fallingdotseq(t-\tau, s-\tau)\, d\tau .$$

We note that since $\fallingdotseq$ is doubly periodic with period $2\hat{T}$, $U(t+\tau, s+\tau) = U(t, s)$ so that U is stationary. Furthermore $Z^k(t, s) = U(t, s)$ for $-(k-1)\hat{T} \leq s, t \leq (k-1)\hat{T}$. Thus if $v+n$ is the output process of the channel h corresponding to the input v then

$$\lim_{k\to\infty} \frac{1}{k}\, I(u^k; v^k + n^k) = \lim_{k\to\infty} \frac{1}{k}\, I(\bar{z}^k; \bar{w}^k + n^k) \tag{33}$$

where u^k, v^k and n^k are the restrictions of u, v and n to the interval $[-k\hat{T}, k\hat{T}]$. But by the Appendix,

$$\lim_{k\to\infty} \frac{1}{k\hat{T}} I(u^k; v^k + n^k) = \int_{-\infty}^{\infty} \log\left(1 + \frac{|\tilde{h}(\nu)|^2 \tilde{s}_u(\nu)}{N}\right) dv \tag{34}$$

where $\tilde{s}_u(v)$ is the Fourier transform of $s_u(\tau) = U(0, \tau)$. Combining this equality with Eq. (32) and (33) gives us (35).

$$\int_{-\infty}^{\infty} \log\left(1 + \frac{|\tilde{h}(\nu)|^2 \tilde{s}_u(\nu)}{N}\right) dv \geq \frac{1}{\hat{T}} [(1-\epsilon)\log G - \log 2] \tag{35}$$

Taking the minimum of both sides over $h \in \mathcal{B}_f$ we obtain

$$\inf_{h\in\mathcal{B}_f} \int_{-\infty}^{\infty} \log\left(1+ \frac{|\tilde{h}(\nu)|^2 \tilde{s}_u(\nu)}{N}\right) d\nu \geq \frac{1}{\hat{T}} [(1-\epsilon)\log G - \log 2]$$

The left-hand side can be dominated by taking the supremum over all $\tilde{s}_u \in \mathcal{A}$ so that we get

$$C(\mathcal{B}_f) \geq \frac{1}{\hat{T}} [(1-\epsilon)\log G - \log 2] = \frac{1}{T+\delta} [(1-\epsilon)\log G - \log 2]$$

We have therefore proved Lemma 8:

Lemma 8.

If there exists a $(G, 1/2\epsilon, T)$ code for $\mathcal{B}$ then for every finite subset $\mathcal{B}_f$ of $\mathcal{B}$,

$$\frac{1}{T+\delta}\left[(1-\epsilon)\log G - \log 2\right] \leq C(\mathcal{C}_f)$$

Theorem 3.

$$\hat{C}(\mathcal{C}) \leq C(\mathcal{C}).$$

Proof. Let $R < \hat{C}(\mathcal{C})$ so that there is a sequence of $(e^{RT_n}, \epsilon_n, T_n)$ codes for $\mathcal{C}$ with $T_n \to \infty$ and $\epsilon_n \to 0$. For each finite subset $\mathcal{C}_f$ of $\mathcal{C}$ we get

$$C(\mathcal{C}_f) \geq \frac{1}{T_n+\delta}\left[(1-2\epsilon_n)RT_n - \log 2\right]$$

Taking limits as $n \to \infty$ this yields

$$C(\mathcal{C}_f) \geq R.$$

Taking the infimum over all finite subsets $\mathcal{C}_f$ of $\mathcal{C}$ this gives us

$$\inf_{\mathcal{C}_f \subset \mathcal{C}} C(\mathcal{C}_f) \geq R$$

It remains to show that

$$\inf_{\mathcal{C}_f \subset \mathcal{C}} C(\mathcal{C}_f) = C(\mathcal{C}). \tag{36}$$

Clearly, the left-hand side is not less than the right-hand side. Now since the set $\{\tilde{h}(\nu) \,|\, h \in \mathcal{B}\}$ is a conditionally compact subset of L_2, and since the Fourier transform is an isometry, the set $\{h(t) \,|\, h \in \mathcal{B}\}$ is a conditionally compact subset of L_2. Also if $h \in \mathcal{B}$ then $h(t) = 0$, $|t| \geq \delta$, so that $\{h(t) \,|\, h \in \mathcal{B}\}$ is a conditionally compact subset of L_1 and therefore $\{\tilde{h}(\nu) \,|\, h \in \mathcal{B}\}$ is a conditionally compact subset of L_∞. Hence given $\epsilon > 0$ there is a finite subset $\mathcal{B}_f^\epsilon$ of $\mathcal{B}$ such that for each $h \in \mathcal{B}$, there is an $h_f \in \mathcal{B}_f^\epsilon$ such that

$$||\tilde{h} - \tilde{h}_f||_\infty \leq \epsilon .$$

Therefore for all $\tilde{s} \in \mathcal{D}$,

$$\left| \int_{-\infty}^{\infty} \log \left(1 + \frac{|\tilde{h}(\nu)|^2 \tilde{s}(\nu)}{N} \right) d\nu - \int_{-\infty}^{\infty} \log \frac{(1 + |\tilde{h}_f|^2 \tilde{s}(\nu))}{N} d\nu \right|$$

$$\leq \frac{1}{N} \int_{-\infty}^{\infty} \left| |\tilde{h}(\nu)|^2 - |\tilde{h}_f(\nu)|^2 \right| \tilde{s}(\nu) \quad d\nu$$

$$\leq \frac{1}{N} \epsilon^2 \int_{-\infty}^{\infty} \tilde{s}(\nu)\, d\nu \leq \frac{1}{N} \epsilon^2 .$$

It follows that $C(\mathcal{B}_f^\epsilon) \leq C(\mathcal{B}) + \frac{1}{N} \epsilon^2$ which proves Eq. 36.

APPENDIX

In this appendix we prove Eq. (34). Recall that u(t) is a stationary Gaussian process with zero-mean and covariance function $s_u(t)$. $v(t)+n(t)$ is the output of the channel h corresponding to input u(t). u^k, v^k, n^k are the restrictions of the processes u, v, and n to the interval $[-k\hat{T}, k\hat{T}]$.

Lemma.

$$I(u^k; v^k + n^k) = \frac{1}{2} \sum_{i=1}^{\infty} \log\left(1 + \frac{\lambda_i^k}{N}\right)$$

where $\lambda_1^k \geq \lambda_2^k \geq \ldots$ are the eigenvalues of the operator

$$\Delta_T = P_T H P_T S_u P_T H^* P_T$$

where

(i) $T = k\hat{T}$

and

(ii) H and S_u are the operators defined by the kernels h and s_u respectively.

Proof. $\Delta_T = H_T S_T H_T^*$ where $H_T = (P_T H_T P_T)$ and $S_T = P_T S_u P_T$. Furthermore, S_T is a positive semi-definite self-adjoint compact operator and hence can be expressed as $S_T = SS$ where S is positive semi-definite, self-adjoint and compact. Thus $\Delta_T = AA^*$ where $A = H_T S$. The operator $\hat{\Delta}_T = A^* A$ is also positive semi-definite, self-adjoint and compact, and has the same eigenvalues as the operator Δ_T. Therefore there is a complete orthonormal basis $\{\delta_i\}_{i=1}^{\infty}$ of $L_2(T)$ such that $\hat{\Delta}_T = A^* A$ is defined by the kernal

$$\hat{\delta}(t, \tau) = \sum_{i=1}^{\infty} \lambda_i^k \delta_i(t) \delta_i(\tau)$$

Let

$$\eta_i(t) = \frac{1}{\sqrt{\lambda_i^k}} (A \delta_i)(t).$$

Then

$$\left(\eta_i, \eta_j\right) = \frac{1}{\sqrt{\lambda_i^k}\sqrt{\lambda_j^k}} \left(A\delta_i, A\delta_j\right) = \frac{1}{\sqrt{\lambda_i^k \lambda_j^k}} \left(\delta_i, A^* A \delta_j\right)$$

$= \delta_i^j$ - the Kronecker delta; so that $\{\eta_i\}_{i=1}^{\infty}$ is also a complete orthonormal basis of $L_2(T)$. Finally, relative to these basis functions we have the representations.

$$u^k(t) = \sum (u^k)_i \delta_i(t)$$

$$v^k(t) = \sum (v^k)_i \eta_i(t)$$

$$n^k(t) = \sum (n^k)_i \eta_i(t)$$

where $(u^k)_i$, $(v^k)_i$ and $(n^k)_i$ are appropriate Gaussian random variables. Furthermore the variance of $(v^k)_i$ is λ_i^k and variance of $(n^k)_i$ is N. It is easy to check that with the various independance relations

$$I((u^k)_1, \ldots, (u^k)_\ell \; ; \; (y^k)_1 + (n^k)_1, \ldots, (y^k)_\ell + (n^k)_\ell)$$

$$= \frac{1}{2} \sum_{i=1}^{\ell} \log \left(1 + \frac{\lambda_i^k}{N} \right) ,$$

and the lemma is proved.

Finally, combining this lemma with Theorem 1 establishes Eq. (34).

REFERENCES

1. Root, W. L. and Varaiya, P. P., "Capacity of Classes of Gaussian Channels. Part I: Discrete-Time," Tech. Memo M-208. Electronics Res. Lab., Univ. of California, Berkeley, (1967).

2. Blackwell, D., Breiman, L. and Thomasian, A. J., "The Capacity of a Class of Channels," Ann. Math. Stat. 30 (1959) p. 1229.

3. Gallager, R. G., "Continuous Time Channels," Chapter of a book in preparation.

4. Kac, M., Murdock, W. L. and Szego, G., "On the Eigenvalues of Certain Hermitian Forms," Jour. of Rat. Mech. and Anal. 2 (1953) p. 767.

5. Dunford, N., and Schwartz, J. T., Linear Operators Part 1, Interscience, New York (1958).

6. Bochner, S. and Chandrasekharan, K., Fourier Transforms, Ann. of Math. Studies No. 19, Princeton Univ. Press (1949).

7. Riesz, F. and Sz-Nagy, B., Functional Analysis, (translated from the French) Ungar, New York (1955).

8. Root, W. L., "Estimation of ϵ-Capacity for Certain Linear Communication Channels," Willow Run Laboratories Report, Univ. of Michigan, (1967).

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

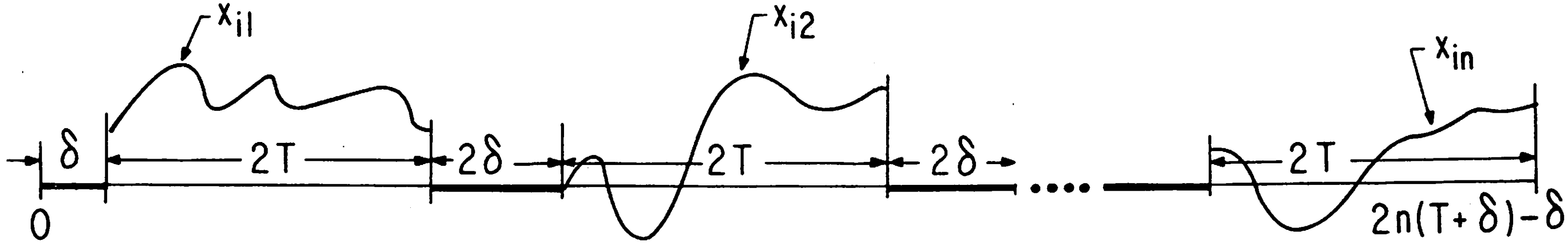

Fig. 1. $\hat{u}_i(t)$.

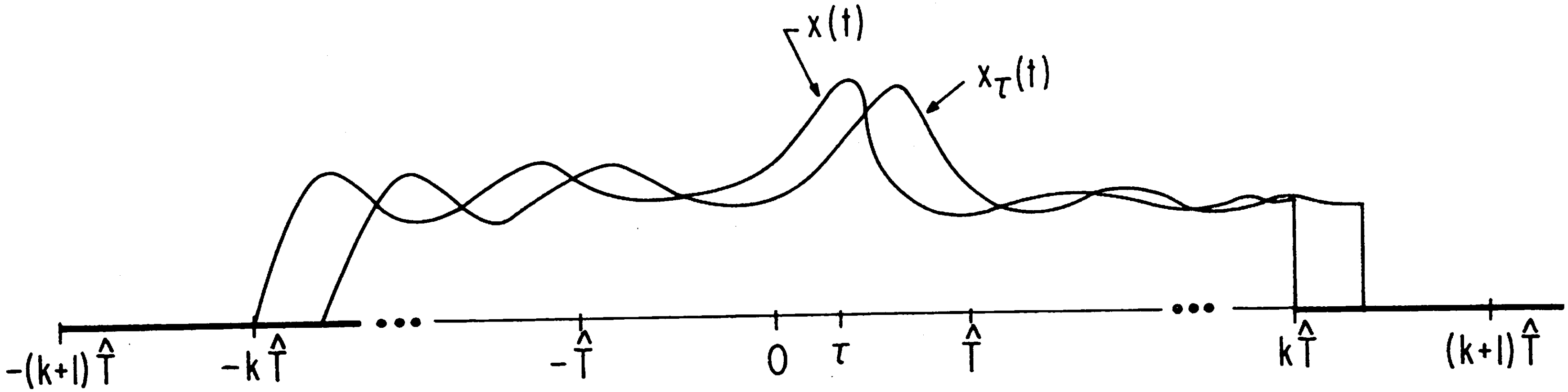

Fig. 2. $x(t)$, $x_\tau(t)$.