The Capacity of the Band-Limited Gaussian Channel

By A. D. WYNER

(Manuscript received December 27, 1965)

Shannon's celebrated formula $W \ln(1 + P_o/N_oW)$ for the capacity of a time-continuous communication channel with bandwidth W cps, average signal power P_o, and additive Gaussian noise with flat spectral density N_o has never been justified by a coding theorem (and "converse"). Such a theorem is necessary to establish $W \ln(1 + P_o/N_oW)$ as the supremum of those transmission rates at which one may communicate over this channel with arbitrarily high reliability as the coding and decoding delay becomes large.

In this paper, a number of physically consistent models for this time-continuous channel are proposed. For each model the capacity is established as $W \ln(1 + P_o/N_oW)$ by means of a coding theorem and converse.

I. INTRODUCTION

As an idealized model for the time-continuous Gaussian channel (with bandwidth W cycles per second, two-sided noise spectral density $N_o/2$, and average power P_o), Shannon[1,2] employed the mathematical time-discrete channel which passes $2W$ real numbers x per second, with the average of x^2 restricted to be P_o. Each input x is perturbed by an independent "noise" random variable which is Gaussian with mean zero and variance N_oW. If by "channel capacity" we mean the maximum rate at which a channel is capable of transmitting information with arbitrarily small error probability as the coding and decoding delay becomes large, then the capacity of this time-discrete channel is given by the celebrated formula $W \log_2 (1 + P_o/N_oW)$ bits per second (or $W \ln (1 + P_o/N_oW)$ nats per second).

In order to show that the capacity is given by this formula, it is necessary to prove a coding theorem (showing the possibility of achieving "error-free" communication at any rate less than $W \log_2 (1 + P_o/N_oW))$, and a "converse" (showing the impossibility of achieving "error-free" coding at a rate exceeding this quantity). For this — purely mathematical — channel these theorems have been proved, and there is no question as to the meaning and validity of the capacity formula.

The way in which Shannon arrived at this time-discrete model for a "physical" time-continuous channel is described in detail in Section II. It will suffice to remark here that there remain questions as to the relation of this time-discrete model (and the resulting capacity formula) to a physically meaningful time-continuous channel. These difficulties center on the fact that the inputs and outputs of the time-continuous channel are band-limited signals which are not physically realizable. As we shall see in Section II, such assumptions lead to a number of anomalies and absurdities.

Our purpose in this paper is to find physically consistant mathematical models for the time-continuous band-limited Gaussian channel, and to establish their capacity by means of a coding theorem and converse. Schematically our results are of the following form:

Let $a(T,W,P_o)$ be a class of functions which are "approximately band-limited to W cycles per second and approximately time-limited to T seconds", and which have "average power" P_o. The channel inputs must be members of a. The noise is additive, stationary, and Gaussian with flat two-sided spectral density $N_o/2$ in the band $0 - W$ cycles per second (or "approximately" given as above). Then the channel capacity, defined as the maximum rate for which arbitrarily high reliability is possible (using signals from a) as T becomes large, is given "approximately" by $W \log_2 (1 + P_o/N_oW)$. The term "approximately" used here will, of course, be given a precise meaning below.

In Section II, Shannon's model and results are discussed, and in Section III our models and results are stated completely and discussed. Our proofs follow in Sections IV and V. A glossary is included at the end of the paper.

II. THE SHANNON MODEL

2.1 *The Time-Discrete Channel*

In order to fix ideas as well as to review some results which will be required subsequently, let us consider the following class of (time-discrete) channels: Every T seconds the input to the channel is a sequence of $n = [\alpha T]$ real numbers $\mathbf{x} = (x_1, x_2, \cdots, x_n)$, where $\alpha (0 < \alpha \leqq \infty)$ is a fixed parameter. Further, the input sequence must satisfy the "energy" constraint

$$E(\mathbf{x}) = \sum_{k=1}^{n} x_k^2 \leqq PT, \tag{1}$$

where $P > 0$ is another fixed parameter, and where $E(\mathbf{x})$ is, as indicated, the sum of the squares of the components of $\mathbf{x}$.

The channel output is also a real n-sequence $\mathbf{y} = (y_1, y_2, \cdots, y_n)$, where

$$y_k = x_k + z_k, \qquad k = 1,2, \cdots, n, \tag{2}$$

and the noise digits z_k ($k = 1,2, \cdots, n$) are independent, normally distributed random variables with mean zero and variance N.

Let us assume that this channel is to be used in the communication system of Fig. 1. The output of the message source is a sequence of independent and equally likely binary digits which appear at the input of the coder at the rate of R_b digits (bits) per second. Every T seconds the coder input is one of $M = 2^{R_b T}$ binary sequences, each sequence being equally likely. Let us number the possible messages as $1,2, \cdots, M$. The coder contains a mapping of the message set $\{1,2, \cdots, M\}$ to a set (called a *code*) of M real n-sequences $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M\}$ (called *code words*) satisfying (1). If message $i (i = 1,2, \cdots, M)$ is the coder input, then the coder output (and hence channel input) is the code word $\mathbf{x}_i$. Since it takes T seconds to transmit a code word, the system can process information continuously without a "backup" at the coder input. The transmission rate is R_b bits per second or $R = (\ln 2)R_b$ nats per second.

It is the task of the receiver (or decoder) to examine the received sequence $\mathbf{y}$, and determine which of the M code words was actually transmitted. Thus, we may think of the decoder as a rule which assigns to each possible received sequence $\mathbf{y}$, a code word $\mathbf{x}_i$. Let us denote by P_{ei} the probability that the decoder chooses the wrong code word given that $\mathbf{x}_i$ was transmitted. The over-all error probability is then

$$P_e = \frac{1}{M} \sum_{i=1}^{M} P_{ei}. \tag{3}$$

A transmission rate R (nats per second) is said to be *permissible* if for every $\lambda > 0$ one can find a T sufficiently large and a code with parameter T with $M = [e^{RT}]$ code words and $P_e \leqq \lambda$. With such a code, the system could process $R_b = R/\ln 2$ bits per second. We define the *channel capacity* C as the supremum of permissible rates. For the channel under discussion the channel capacity is given by the celebrated formula

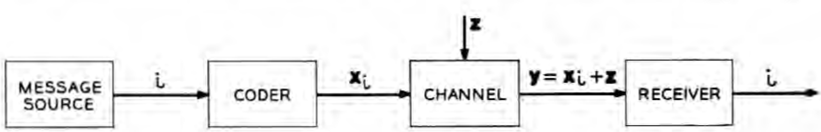

Fig. 1. — Time-discrete channel.

$$C = C_\alpha = \frac{\alpha}{2} \ln \left(1 + \frac{P}{\alpha N}\right). \tag{4}$$

In order to establish C as the capacity, one must prove two theorems. The first ("direct half") states that any $R < C$ is a permissible rate; that is, there exist codes with vanishingly small P_e as $T \to \infty$. The second theorem ("weak converse") states that no $R > C$ is a permissible rate; that is, for any sequence of codes with rate $R > C$, P_e is bounded away from zero. This has been done for the present channel for the case of a finite α by Shannon.[1,2,3] Let us observe that if we let $\alpha \to \infty$ in (4), we have $C_\alpha \xrightarrow{\alpha} P/2N$. The fact that $C_\infty = P/2N$ has been established by Ash.[4] The reader is referred to Ash [Ref. 5, Chapter 8] for a complete discussion of the above. The significance of the channel capacity then, is that it is the maximum rate for which arbitrarily high reliability is possible using signals in a certain class (i.e., those which satisfy (1)) with sufficiently long delay T.

2.2 *Application to the Band-Limited Gaussian Channel*

Shannon[1,2] has applied the above results to the communication system of Fig. 2. As above, the message source emits binary digits at the rate of R_b per second, and after T seconds, one of $M = 2^{R_b T}$ possible messages appears at the coder input. Corresponding to the ith message ($i = 1,2, \cdots, M$) the coder output is the function

$$x_i(t) = \sum_{k=1}^{n} x_{ik}\delta(t - k/2W), \tag{5a}$$

where $\delta(t)$ is the unit impulse, $n = [2WT]$, and the $\{x_{ik}\}_{k=1}^{n}$ satisfy

$$\sum_{k=1}^{n} x_{ik}^2 \leqq 2WP_oT, \qquad i = 1,2,\cdots, M. \tag{5b}$$

As for the time-discrete channel, the coder must contain a set of M real n-sequences. The channel input $s_i(t)$ is the result of passing $x_i(t)$ through an ideal low-pass filter with transfer function

$$H(\omega) = \begin{cases} \dfrac{1}{2W} & |\omega| \leqq 2\pi W, \\ 0 & |\omega| > 2\pi W, \end{cases} \tag{6}$$

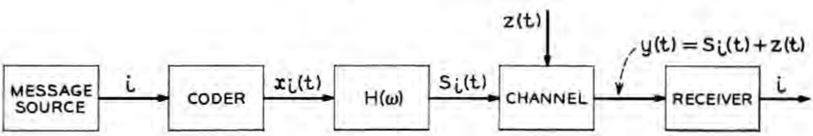

Fig. 2 — Shannon's time-continuous band-limited channel.

so that

$$s_i(t) = \sum_{k=1}^{n} x_{ik} \left[\frac{\sin 2\pi W(t - k/2W)}{2\pi W(t - k/2W)} \right]. \tag{7}$$

Thus, it takes T seconds to generate the filter input, and the system can process information at a rate of $R = (\ln 2) R_b$ nats per second without a "backup" at the coder input. Let us also remark that although the signal $s_i(t)$ is generated in T seconds, due to the physical unrealizability of $H(\omega)$, $s_i(t)$ is nonzero almost everywhere on $(-\infty, \infty)$. This leads to a fundamental difficulty which we shall discuss later.

Let $s(t)$ be the input to the channel due to a repeated application of the coding process (every T seconds). Then $s(t)$ is bandlimited to W cycles per second, and

$$\operatorname*{limit}_{T_o \to \infty} \frac{1}{T_o} \int_{-T_o/2}^{T_o/2} s^2(t)\,dt \leqq P_o . \tag{8}$$

Inequality (8) follows from (5b) and the orthogonality of

$$\frac{\sin 2\pi W(t - k/2W)}{2\pi W(t - k/2W)} \quad \text{and} \quad \frac{\sin 2\pi W(t - k'/2W)}{2\pi W(t - k'/2W)}$$

$(-\infty < k < k' < \infty)$ on the infinite interval $(-\infty, \infty)$. Thus, the channel input is a bandlimited signal with "average power" not exceeding P_o.

Again turning our attention to Fig. 2, the channel output is a function $y(t) = s(t) + z(t)$, where $z(t)$ is a sample from a Gaussian random process with spectral density

$$N(\omega) = \begin{cases} N_o/2 & |\omega| \leqq 2\pi W, \\ 0 & |\omega| > 2\pi W. \end{cases} \tag{9a}$$

The corresponding autocorrelation function of the noise is

$$R(\tau) = \mathcal{E}[z(t)z(t + \tau)] = N_o W \frac{\sin 2\pi W\tau}{2\pi W\tau}, \tag{9b}$$

where $\mathcal{E}$ denotes expectation.

Again it is the function of the receiver (or decoder) to examine $y(t)$ and determine what the input information was. Let us consider the signal $s_i(t)$ (7), which was generated during the interval $[0,T]$. The coefficients $\{x_{ik}\}_{k=1}^{n}$ are the values of $s_i(t)$ at the "sampling instants" $t = k/2W$, $k = 1,2, \cdots, n$. Since the noise is also bandlimited, the received signal $y(t)$ is bandlimited and may be completely characterized by its values at the sampling instants $y_k = y(k/2W)$, $k = 0, \pm 1, \pm 2, \cdots$. Clearly

$$y_k = x_{ik} + z_k, \qquad k = 1,2, \cdots, n, \tag{10}$$

where $z_k = z(k/2W)$ is the value of the noise $z(t)$ at the sampling instant $t = k/2W$. Since $s_i(k/2W) = 0$, for $k < 1$ and $k > n$, the only useful samples of y are $\{y_k\}_{k=1}^n$. Further it follows directly from (9b) that the z_k are independent, normally distributed random variables with mean zero and variance N_oW. Thus, it suffices to consider the input and output as n-sequences $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{in})$ and $\mathbf{y} = (y_1, \cdots, y_n)$ $(n = 2WT)$ related by (10). Let us remark here, that the code words corresponding to previous and successive intervals will not cause any interference with the code word corresponding to the interval [0,T], since these other code words are zero at the sampling instants.

Inequality (5b) and (10) permit us to apply the results for the time-discrete Gaussian channel discussed above with parameters $\alpha = 2W$, $P = 2WP_o$, and $N = N_oW$. We conclude that this communication system (in Fig. 2) is capable of processing information at any rate R less than

$$C = W \ln\left(1 + \frac{P_o}{N_oW}\right), \tag{11}$$

with vanishingly small error probability as T becomes large. Since the channel inputs are bandlimited to W cycles per second, and by (8) have average power not exceeding P_o, it is generally believed that the *capacity* (taken as the maximum "error-free rate") of a channel which admits only bandlimited signals with average power P_o is given by (11). In fact, it has only been shown that it is possible to do at least as well as C (using the system of Fig. 2), and no converse has been proven. This is the first difficulty with the Shannon model which we shall attempt to remedy.

Further, there are other difficulties inherent in the use of this model. We are taking "capacity" to be a (maximum) transmission *rate*, but what is the rate for the system of Fig. 2? We have said merely that the *coder* can process information at a rate of R nats per second. However, because of the physical unrealizability of $H(\omega)$, we must discard all temporal notions about the channel input $s_i(t)$ as well as the output $y(t)$. The notion of *rate*, therefore, has only a limited meaning. In fact, since the received signal $y(t)$ is an entire function, it is perfectly predictable for all time from observations over a finite interval. Thus the receiver, by observing $y(t)$ in a tiny interval, could extrapolate $y(t)$ for all time and obtain sample values at an arbitrarily high rate. This anomaly is the second difficulty with the Shannon model.

It is the purpose of this paper to present a model for the time-continuous band-limited Gaussian channel for which the capacity (defined as the maximum "error-free rate") is given by (11). This will necessitate proving a "direct half" and "converse" to a coding theorem. Further, the model should avoid the second difficulty mentioned above. We shall obtain results of the following form:

Let $a(T,W,P_o)$ be a class of functions which are "approximately bandlimited to W cycles per second and approximately time-limited to T seconds", and which have total "energy" not exceeding P_oT. The noise is taken to be stationary and Gaussian with spectral density given (or "approximately" given) by (9a). Then the channel capacity, defined as the maximum rate for which arbitrarily high reliability is possible (using signals from a) as T becomes large, is given "approximately" by $W \ln (1 + P_o/N_oW)$. The term "approximately" used here will, of course, be given a precise meaning below.

III. SUMMARY OF RESULTS

We shall propose four models for the channel and find the capacity of each. Each model is of the following form:

(*i*) Definition of a suitable class of allowable signal functions, $a(T,W,P_o)$, which are "approximately bandlimited to W cycles per second, approximately time-limited to T seconds", and with total energy not exceeding P_oT.

(*ii*) Definition of the noise — taken to be stationary additive Gaussian noise with spectral density $N(\omega)$, which is "approximately" given by (9a).

We shall take W and P_o to be fixed parameters. A *code* with parameter T is a set of M functions (called *code words*) in $a(T,W,P_o)$. The transmission rate R is defined by $R = (1/T) \ln M$, so that $M = e^{RT}$. A decoding scheme is a mapping of the space of possible received signals (code word plus a noise sample) onto the code. If code word i ($i = 1, 2, \cdots, M$) is transmitted, we take P_{ei} to be the conditional probability that the decoder chooses a code word other than i, and hence makes an error. Since all code words are equally likely to be transmitted, the over-all error probability P_e is given by (3), i.e.,

$$P_e = \frac{1}{M} \sum_{i=1}^{M} P_{ei} .$$

A transmission rate R is said to be *permissible*, if for every $\lambda > 0$ one can find a T sufficiently large and a code with $M = [e^{RT}]$ code words for which $P_e \leqq \lambda$. The *channel capacity* C is defined as the supremum of permissible rates. We shall find the capacity corresponding to a number of different $a(T,W,P_o)$ and $N(\omega)$. This will, as for the time-discrete channel, necessitate proving two coding theorems — a "direct half" and a "weak converse".

Before beginning the summary we shall need the following definitions. Let $s(t)$, $-\infty < t < \infty$, be a real-valued square-integrable function and $S(\omega)$ be its Fourier transform. Let the norm of $s(t)$ be

$$\| s \| = \left[\int_{-\infty}^{\infty} s^2(t)dt \right]^{\frac{1}{2}}. \tag{12}$$

The frequency and time "concentration" of s are

$$K_B(s,2\pi W) = \frac{1}{2\pi} \int_{-2\pi W}^{2\pi W} | S(\omega) |^2 d\omega / \| s \|^2, \tag{13a}$$

and

$$K_D(s,T) = \int_{-T/2}^{T/2} s^2(t)dt / \| s \|^2, \tag{13b}$$

respectively. Further, let D_T be the "time-truncation" operator defined by

$$D_T s = \begin{cases} s(t) & |t| \leqq T/2, \\ 0 & |t| > T/2. \end{cases} \tag{14}$$

With these definitions in hand, we are able to state our results. In each case we shall define the channel model and then give the channel capacity. Although there are some difficulties inherent in these models, each model leads to a *mathematical theorem* which justifies Shannon's capacity formula.

Model 1: To begin with, let us take for the set a of "allowable" inputs. $a_1(T,W,P_o)$, the set of functions $s(t)$ satisfying

$$s(t) = 0, \quad | t | > T/2, \tag{15a}$$

$$\| s \|^2 \leqq P_o T, \tag{15b}$$

$$K_B(s,2\pi W) \geqq 1 - \eta \quad (0 < \eta < 1). \tag{15c}$$

Hence, our allowable signals are functions which are strictly time-limited and approximately band-limited. As $\eta \rightarrow 0$, the allowable signals become more perfectly bandlimited. The noise spectrum is taken to be

$$N(\omega) = \begin{cases} N_o/2 & |\omega| \leqq 2\pi W, \\ \nu N_o/2 & |\omega| > 2\pi W, \end{cases} \tag{16}$$

where $0 < \nu \leqq 1$. As $\nu \to 0$, (16) is in some sense "approximately" the same as (9a). The average noise power outside the band ($|\omega| > 2\pi W$), however, is infinite. In this case, Theorem 3 establishes

$$C = C_\eta = W \ln \left(1 + (1 - \eta) \frac{P_o}{N_o W}\right) + \eta \frac{P_o}{\nu N_o} \tag{17}$$

as the channel capacity. As $\eta \to 0$, the capacity approaches the classical formula $W \ln (1 + P_o/N_o W)$.

The principal difficulty with this model is the assumption of infinite average noise power, which is hardly a physically acceptable notion. Further, there are mathematical difficulties inherent in a spectral density given by (16) which implies a covariance containing an impulse function. Often the assumption of a spectrum in (16) can be justified by the fact that it can be approximated as closely as desired in the frequency range of interest by a spectrum with finite power. However, the following theorem, the proof of which is Appendix B, renders this justification meaningless in this case.

Theorem 5: Let $a(T,W,P_o)$ be as in (15) and let the noise be additive and Gaussian with spectral density $N(\omega)$, where

$$\int_{-\infty}^{\infty} N(\omega) d\omega < \infty.$$

Then the capacity $C_\eta = \infty$ regardless of how small η may be.

Intuitively, we may see that this is true by observing that, since the above integral exists, $N(\omega)$ must be arbitrarily small in some frequency range. Hence, by placing some signal energy into this frequency range, we can make the "signal-to-noise" ratio arbitrarily large, and therefore, the permissible rate of transmission arbitrarily high.

Accordingly, we shall assume for the remaining models that the noise is additive, Gaussian, with spectral density

$$N(\omega) = \begin{cases} N_o/2 & |\omega| \leqq 2\pi W, \\ 0 & |\omega| > 2\pi W. \end{cases} \tag{18}$$

This corresponds more closely with the usual formulation of a band-limited channel. It remains to find a suitable class of input signals, $a(T,W,P_o)$. We consider some possibilities.

Model 2: This model defines $a = a_2(T,W,P_o)$ as the set of functions $s(t)$

satisfying

$$S(\omega) = 0, \qquad |\omega| > 2\pi W, \tag{19a}$$

$$\| s \|^2 \leqq P_o T, \tag{19b}$$

$$K_D(s,T) \geqq 1 - \eta \quad (0 < \eta < 1). \tag{19c}$$

Thus, a_2 is a set of strictly band-limited, approximately time-limited functions. As $\eta \to 0$, the allowable signals become more perfectly time-limited. With the noise as defined in (18), Theorem 2 establishes

$$C = C_\eta = W \ln \left(1 + (1 - \eta) \frac{P_o}{N_o}\right) + \eta \frac{P_o}{N_o} \tag{20}$$

as the channel capacity. Again, as $\eta \to 0$, C_η approaches the classical formula $W \ln [1 + (P_o/N_oW)]$.

Model 2 is an intuitively plausible model for the band-limited channel, and Theorem 2 which establishes its capacity is a mathematically rigorous result which, in the limit, yields the desired capacity formula. There are, however, two difficulties inherent in this formulation. The first is that since the allowable signals $s(t)$ are band-limited, it is not possible to generate them in finite time. Thus the central idea of a transmission *rate* has, at best, a limited meaning. The Shannon model (Fig. 2) also suffers from this difficulty (see Section II). The other problem with this formulation is that if code words are transmitted sequentially, we will have an interference problem (i.e., the tails of successive signals will overlap), the resolution of which is not known at present. The following two models contain neither of these difficulties.

Model 3: This model avoids the difficulties of Model 2 by letting the code words be strictly time-limited and approximately band-limited. However, as we have seen in Theorem 5, the definition of approximately band-limited functions employed above (15) yields an infinite capacity. Thus we seek an alternate way of characterizing "approximately" band-limited or "slowly changing" functions. We proceed as follows. Let $x(t)$ be a function satisfying $x(t) = 0$, $|t| > T/2$, and $\| x \|^2 < \infty$. If $x = D_T\hat{x}$, where $\hat{x}$ is a strictly bandlimited function and D_T is defined by (14), we may define a "frequency concentration" of x by

$$K_B'(x,2\pi W) = \frac{\| x \|^2}{\| \hat{x} \|^2}. \tag{21}$$

If we cannot express x as $D_T\hat{x}$, we take $K_B' = 0$. For example, if $x(t)$ or any of its derivatives has even a small discontinuity then we cannot write $x = D_T\hat{x}$, so that $K_B'(x,2\pi W) = 0$ and x is not approximately bandlimited in this sense. This is so no matter how large $K_B(x,2\pi W)$ may be. Conversely, it is shown in Appendix C that for any function x

$$K_B(x,2\pi W) \geqq 1 - 2\sqrt{\frac{1 - K_B'(x,2\pi W)}{K_B'(x,2\pi W)}}, \tag{22}$$

so that a K_B' close to unity implies a K_B close to unity. Thus, saying that a function x has a K_B' close to unity implies that x is "slowly changing" and that K_B is also close to unity.

We now choose that set $a = a_3(T,W,P_o)$ of allowable inputs as the set of functions $s(t)$ for which

$$s(t) = 0, \qquad |t| > T/2, \tag{23a}$$

$$\|s\|^2 \leqq P_oT, \tag{23b}$$

$$K_B'(s,2\pi W) \geqq 1 - \eta \quad (0 < \eta < 1). \tag{23c}$$

Thus a_3 is a set of strictly time-limited, and approximately band-limited functions. In this case, Theorem 4 establishes

$$C = C_\eta = W \ln\left(1 + \frac{P_o}{N_oW}\right) + \frac{\eta}{1-\eta}\frac{P_o}{N_o} \tag{24}$$

as the channel capacity. Again $C_\eta \to W \ln [1 + (P_o/N_oW)]$ as $\eta \to 0$.

The significance of constraint (23c) is that it makes it impossible for the communicator to make any *use* of the high-frequency components which must of necessity be included in the signal (since it is time-limited). Model 3, therefore, provides a mathematically rigorous theorem which does not involve any complications concerning physical realizability, and yields the desired capacity.

Our final formulation is as follows:

Model 4: Let $a = a_4(T,W,P_o)$ be the set of strictly time-limited, approximately band-limited functions $s(t)$ which satisfy

$$s(t) = 0, \qquad |t| \geqq T/2, \tag{25a}$$

$$\|s\|^2 \leqq P_oT, \tag{25b}$$

$$K_B(s,2\pi W) \geqq 1 - \eta. \tag{25c}$$

Now Theorem 5 (stated above) tells us that if the noise were as in (18), then the capacity is infinite. In actuality one could not be sure that the noise was absolutely band-limited. In fact, whether or not the noise is strictly band-limited is not verifiable in the laboratory. It is reasonable, therefore, to assume that the noise is given by $z(t) = z_1(t) + z_2(t)$, where $z_1(t)$ is a sample from a Gaussian random process with spectral density (18). For $z_2(t)$ we require only that

$$\int_{-T/2}^{T/2} z_2^2(t)\,dt \leqq \nu N_o W T, \tag{26}$$

where $\nu > 0$ is small. We place no other restrictions on the spectrum of z_2 or on its probability structure. Since the expected value of the energy of $z_1(t)$ in $[-T/2,T/2]$ is $N_o WT$, (26) implies that the energy of $z(t)$ is nearly all in $z_1(t)$ $(\nu \ll 1)$. We shall assume that $z_2(t)$ may depend on the code and decoding rule used, on the code word transmitted, and the sample $z_1(t)$. We require our communication system to perform well no matter what $z_2(t)$ may be.

Let us say that a code (satisfying (25)) and a decoding rule have been chosen. Let us also assume that the rule for selecting $z_2(t)$ has been chosen. Let $P_e(z_2)$ be the resulting error probability. Then define

$$P_e = \max_{z_2} P_e(z_2), \tag{27}$$

where the maximization in (27) is over all rules for choosing $z_2(t)$ — with the code and decoding rule fixed. The channel capacity is the supremum of those rates for which P_e may be made to vanish as $T \to \infty$.

It can be shown (see Appendix D) that the capacity C is given by

$$C = C_{\eta,\nu} = W \ln\left(1 + \frac{P_o}{N_o W}\right) + \varepsilon(\eta,\nu), \tag{28}$$

where $\varepsilon(\eta,\nu) \to 0$ as $\eta,\nu \to 0$ provided $\nu/\eta > P_o/N_oW$, the signal-to-noise ratio. Since we may consider η and ν to be limits on the accuracy of our measuring equipment, the former on measuring the signal* and the latter on measuring the noise, it is reasonable to assume, as we did in (28), that η and ν go to zero at the same rate.

An alternate and mathematically equivalent formulation of Model 4 is as follows: Let the signals $s(t)$ be as in (25) and the noise $z(t)$ be as in (18). Now in reality one could not expect the decoder to be capable of infinitesimally accurate measurements. It is reasonable, therefore, to assume that there is an inherent uncertainty in all measurements made by the decoder, and to require that the communication system perform well despite this uncertainty. Specifically, we require that the decoding regions satisfy the following condition: If $y_1(t)$ is decoded as s_i, and $y_2(t)$ is decoded as $s_j (i \neq j)$, then

* *I.e.*, η represents a limit on the measurement of the frequency component of the signal outside the band.

$$\int_{-T/2}^{T/2} (y_1(t) - y_2(t))^2 \, dt \geqq 2\nu N_o WT. \tag{29}$$

In other words, if a received signal $y(t)$ is close to the "border" between decoding regions, we cannot, because of the uncertainty in the accuracy of our measurements, be sure to which region $y(t)$ belongs. Condition (29) forces the decoder to give up on such a $y(t)$ and to announce an error. The capacity for this alternate model is also given by (28). Let us remark that here ν is again a measure of the accuracy of our measuring instruments, this time at the decoder, so that again it is reasonable to expect η and ν to tend to zero at the same rate.

IV. PRELIMINARIES TO PROOFS

4.1 *The Product of Time-Discrete Channels*

The *product* or parallel combination of r time-discrete Gaussian channels is defined as follows. Every T seconds the input to the channel is an r-tuple $(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(r)})$, where

$$\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \cdots, x_{n_i}^{(i)}) \qquad (i = 1, 2, \cdots, r)$$

is a real n_i-vector ($n_i = \alpha_i T$, α_i a fixed parameter). Each vector $\mathbf{x}^{(i)}$ satisfies the energy constraint

$$E[\mathbf{x}^{(i)}] = \sum_{k=1}^{n_i} [x_k^{(i)}]^2 \leqq P_i T, \qquad i = 1,2,\cdots, r, \tag{30}$$

where the $P_i > 0$ are fixed parameters. The channel output is also an r-tuple $(\mathbf{y}^{(1)}, \cdots, \mathbf{y}^{(r)})$, where the $\mathbf{y}^{(i)}$ are n_i-vectors given by

$$\mathbf{y}^{(i)} = \mathbf{x}^{(i)} + \mathbf{z}^{(i)}, \tag{31}$$

where the $\mathbf{z}^{(i)}$ are n_i-vectors whose coordinates are independent Gaussian random variables with mean zero and variance $N_i (i = 1, 2, \cdots, r)$. Further, the $\{\mathbf{z}^{(i)}\}_{i=1}^{r}$ are statistically independent. Codes, permissible rates of transmission, and channel capacity are defined as in Section I. The following is proved in Ref. 6.

Lemma A: The capacity C of the product of r time-discrete Gaussian channels, with parameters (α_i, P_i, N_i), $i = 1, 2, \cdots, r$, is given by the sum of the capacities of the component channels:

$$C = \sum_{i=1}^{r} \frac{\alpha_i}{2} \ln\left(1 + \frac{P_i}{\alpha_i N_i}\right). \tag{32}$$

Equation (32) also holds when one or more of the $\alpha_i = \infty$. In this case we read $x \ln [1 + (c/x)] \mid_{x=\infty}$ as c.

4.2 *The Jointly-Constrained Product Channel*

We define the jointly-constrained product of time-discrete channels exactly as the ordinary product with constraint (30) replaced by constraints of the following form:

Type 1: Let $r = 2$ and $N_1 = N_2 = N$ and instead of (30) we have

$$E(\mathbf{x}) = E(\mathbf{x}^{(1)}) + E(\mathbf{x}^{(2)}) \leqq PT. \tag{33a}$$

If $\alpha_1 \leqq \alpha_2$ we introduce an additional constraint on $\mathbf{x}^{(2)}$

$$E(\mathbf{x}^{(2)}) \leqq \hat{\eta} E(\mathbf{x}) \tag{33b}$$

where $\hat{\eta}(0 \leqq \hat{\eta} \leqq 1)$ is another fixed parameter. In other words, we have constrained the total energy of the two input vectors (33a), and introduced another constraint on the second input vector $\mathbf{x}^{(2)}$ requiring it to have no more than $\hat{\eta}$ of the total energy (33b). If $\alpha_2 \leqq \alpha_1$, we replace (33b) by a similar constraint on $\mathbf{x}^{(1)}$.

Type 2: Let $r = 3$, $N_1 = N_2$, and $N_1 \geqq N_3$. Further, let $\alpha_3 = \infty$. Instead of (30) we require that $\mathbf{x}$ satisfy

$$E(\mathbf{x}) = E(\mathbf{x}^{(1)}) + E(\mathbf{x}^{(2)}) + E(\mathbf{x}^{(3)}) \leqq PT, \tag{34a}$$

$$E(\mathbf{x}^{(3)}) \leqq \hat{\eta} E(\mathbf{x}). \tag{34b}$$

This is a special case of type 1 when $\alpha_2 = 0$, $N_1 = N_3$.

Type 3: Let $r = 2$, $N_1 = N_2 = N$, and $\alpha_2 = \infty$. Instead of (30) require $\mathbf{x}$ to satisfy

$$E(\mathbf{x}^{(1)}) \leqq PT, \tag{35a}$$

$$E(\mathbf{x}^{(2)}) \leqq \hat{\eta} E(\mathbf{x}). \tag{35b}$$

We now ask what is the capacity C of these channels? The answer is the following theorem which is proven in Appendix A.

Theorem 1: The capacity C of the jointly-constrained product channel as defined above is

Type 1 $(r = 2, N_1 = N_2 = N)$:

$$C = C_1((1 - \beta)P) + C_2(\beta P), \tag{36}$$

where

$$\beta = \min\left(\hat{\eta}, \frac{\alpha_2}{\alpha_1 + \alpha_2}\right), \tag{37a}$$

and

$$C_i(x) = \frac{\alpha_i}{2} \ln\left(1 + \frac{x}{\alpha_i N}\right), \qquad i = 1,2. \tag{37b}$$

Again when $\alpha_i = \infty$, we interpret $x \ln [1 + (c/x)] \mid_{x=\infty} = c$. In particular, when $\alpha_2 = \infty$ $(\alpha_1 < \infty)$, $\beta = \hat{\eta}$, so that (36) implies that we can do no better than putting as much energy into Channel 2 as (33b) will permit.

Type 2 $(r = 3, N_1 = N_2 \geqq N_3, \alpha_3 = \infty)$:

$$C = \frac{\alpha_1}{2} \ln\left(1 + \frac{(1-\hat{\eta})P}{(\alpha_1+\alpha_2)N_1}\right) + \frac{\alpha_2}{2} \ln\left(1 + \frac{(1-\hat{\eta})P}{(\alpha_1+\alpha_2)N_1}\right) + \hat{\eta}\,\frac{P}{2N_3} \tag{38}$$

Type 3 $(r = 2, N_1 = N_2 = N, \alpha_2 = \infty)$:

$$C = \frac{\alpha_1}{2} \ln\left(1 + \frac{P}{\alpha_1 N}\right) + \frac{\hat{\eta}P}{(1-\hat{\eta})2N}. \tag{39}$$

4.3 *Prolate Spheroidal Wave Functions*

The following material can be found in Ref. 7. Given any $W,T > 0$ we can find a countably infinite set of real functions $\{\psi_i(t)\}_{i=1}^{\infty}$, called *prolate spheroidal wave functions* (PSWF), and a set of real positive numbers

$$1 > \lambda_1 > \lambda_2 > \cdots \tag{40}$$

with the following properties:*

(*i*) The $\psi_i(t)$ are bandlimited to W cycles per second, orthonormal on the real line, and complete in the space of bandlimited functions of bandwidth W cycles per second.

(*ii*) The restrictions of the $\psi_i(t)$ to the interval $[-T/2, T/2]$ are orthogonal:

$$\int_{-T/2}^{T/2} \psi_i(t)\psi_j(t)dt = \begin{cases} \lambda_i & i = j, \\ 0 & i \neq j. \end{cases} \tag{41}$$

* Note that the first PSWF is $\psi_1(t)$. In Ref. 7, on the other hand, the first PSWF is $\psi_0(t)$.

The restrictions of the $\psi_i(t)$ are also complete in $\mathfrak{L}_2$ $[-T/2,T/2]$, the space of square integrable functions on $[-T/2,T/2]$.

(*iii*) For all t, the $\psi_i(t)$ satisfy the integral equation

$$\lambda_i\psi_i(t) = \int_{-T/2}^{T/2} \psi_i(s) \frac{\sin 2\pi W(t\text{-}s)}{\pi(t\text{-}s)} ds. \tag{42}$$

Thus the λ_i are the eigenvalues, and the ψ_i the eigenfunctions of the integral equation (42). It follows immediately from (42) that the time-limited functions $D_T\psi_i$ (see (14)) have frequency concentration (see (13a))

$$K_B(D_T\psi_i, 2\pi W) = \lambda_i, \qquad i = 1, 2, \cdots . \tag{43}$$

It can be shown that the λ_i and ψ_i depend upon W and T only through the product WT. Further,

(*iv*) For a fixed $\delta > 0$:

$$\lambda_{2WT(1-\delta)} \to 1 \quad \text{as} \quad WT \to \infty \tag{44a}$$

and

$$\lambda_{2WT(1+\delta)} \to 0 \quad \text{as} \quad WT \to \infty. \tag{44b}$$

Thus roughly speaking, for large WT, approximately $2WT$ of the λ_i are approximately unity, and the remainder are approximately zero.

4.4 *Karhunen-Loeve Expansion*

Let $z(t)$ be a Gaussian random process with spectral density $N(\omega)$ given by (18). Then, using the Karhunen-Loeve Theorem[8], we may write $z(t)$ as

$$z(t) = \sum_{k=1}^{\infty} z_k\psi_k(t), \qquad -\frac{T}{2} \leqq t \leqq \frac{T}{2}, \tag{45}$$

where the $\psi_k(t)$ are PSWF's, and the z_k are independent random variables which are normally distributed with mean zero and variance $N_o/2$. The sum in (45) converges to $z(t)$ with probability 1 for every t.

If $N(\omega)$ is given by (16), then we may formally represent $z(t)$ by

$$z(t) = \sum_{k=1}^{\infty} z_k \frac{\psi_k(t)}{\sqrt{\lambda_k}}, \qquad -\frac{T}{2} \leqq t \leqq \frac{T}{2}, \tag{46}$$

where the λ_k are the eigenvalues of the PSWF's (40), and the z_k are independent normally distributed random variables with mean zero and variance

$$\mathcal{E}(z_k^2) = \frac{N_o}{2}[\lambda_k(1-\nu)+\nu].$$

Thus from (44) roughly speaking, for large WT, approximately $2WT$ of the z_k have variance $N_o/2$, and the remainder variance $\nu N_o/2$.

V. PROOFS OF THE THEOREMS

The general ideal of the proofs in this section is as follows. All the time continuous input signals (i.e., members of $a(T,W,P_o)$ can be written in a Fourier series in PSWF's in which, roughly speaking, the first $2WT$ terms correspond to the part of the signal which is simultaneously approximately confined to the frequency band $|\omega| \leqq 2\pi W$ and to the time interval $|t| \leqq T/2$. The noise sample $z(t)$ may also be written in a Karhunen-Loeve expansion in PSWF's. The result is to reduce the time-continuous channel into a jointly-constrained product of time-discrete channel (discussed in Section 4.2). Channel 1 corresponds to the first $2WT$ PSWF's so that the parameter $\alpha_1 = 2W$. Channel 2 corresponds to the remaining PSWF's so that $\alpha_2 = \infty$. The energy requirement on the time continuous signal $\| s \|^2 \leqq PT$ yields a joint energy constraint for the product channels (as in (33a) for example), and the requirement that the energy outside the frequency band (or time-interval) be small yields a second energy constraint on the input to Channel 2 (as in (33b) for example). Application of Theorem 1 then yields the desired theorems. In the remainder of this section we shall make these ideas precise.

We begin by establishing the capacity of the channel defined by Model 2.

Theorem 2: Let the allowable signal set be $a_2(T,W,P_o)$, the set of functions $s(t)$ satisfying

$$S(\omega) = 0, \qquad |\omega| > 2\pi W, \tag{47a}$$

$$\| s \|^2 \leqq P_o T, \tag{47b}$$

$$K_D(s,T) \geqq 1 - \eta \; (0 < \eta < 1). \tag{47c}$$

The noise is a sample from a Gaussian random process with spectral density

$$N(\omega) = \begin{cases} N_o/2 & |\omega| \leqq 2\pi W, \\ 0 & |\omega| > 2\pi W. \end{cases} \tag{48}$$

Then the channel capacity is

$$C = C_\eta = W \ln\left(1 + (1-\eta)\frac{P_o}{N_o W}\right) + \eta \frac{P_o}{N_o}. \tag{49}$$

Proof:

(i) Direct Half: Let R be given satisfying

$$R < W \ln\left[1 + (1-\eta)\frac{P_o}{N_oW}\right] + \eta\frac{P_o}{N_o}. \tag{50}$$

Since the right member of (50) is continuous in η and W, we may find a $\delta > 0$ and $\sigma > 0$ sufficiently small so that

$$R < W(1-\delta)\ln\left[1 + (1-\eta+\sigma)\frac{P_o}{N_oW(1-\delta)}\right] + \frac{(\eta-\sigma)P_o}{N_o} \triangleq C^*.$$

We see from (36) that C^* is the capacity of a type 1 jointly constrained product channel with parameters

$$P = P_o, \quad N = N_o/2, \quad \hat{\eta} = \eta - \sigma, \quad \alpha_1 = 2W(1-\delta), \quad \alpha_2 = \infty. \tag{51}$$

We now show how to construct codes for the time-continuous "channel' with rate R and with vanishingly small error probability (as $T \to \infty$). Let $\mathbf{x} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ be an allowable input vector for the type 1 time-discrete product channel with parameters given by (51). Then the corresponding input for the time-continuous channel is

$$s(t) = \sum_{k=1}^{2W(1-\delta)T=\alpha_1T} x_k^{(1)}\psi_k(t) + \sum_{k=1}^{\infty=\alpha_2T} x_k^{(2)}\psi_{k+2W(1-\delta)T}(t) \tag{52}$$

where the $\{\psi_i(t)\}_{i=1}^{\infty}$ are the PSWF's (Section 4.3) with parameters W and T. We first verify that signals of the form of (52) are allowable inputs, i.e., belong to $a_2(T,W,P_o)$ and satisfy (47). That the $s(t)$ are bandlimited and satisfy (47a), follows from the fact that the PSWF's have this property (Section 4.3). Further, the energy of $s(t)$ satisfies

$$\| s \|^2 = \sum_{k=1}^{\alpha_1T}[x_k^{(1)}]^2 + \sum_{k=1}^{\alpha_2T}[x_k^{(2)}]^2 = E(\mathbf{x}) \leqq PT, \tag{53}$$

where use has been made of the orthonomality of the PSWF's on $(-\infty, -\infty)$ (Section 4.3(i)), and the joint energy constraint on $\mathbf{x}$ (33a). Thus $s(t)$ satisfies (47b). Finally, from the orthogonality of the PSWF's on $[-T/2, T/2]$ (41), and the monotonicity of the λ_k (40) we have

$$\begin{aligned}1 - K_D(s,T) &= \frac{\|(1-D_T)s\|^2}{\|s\|^2} \\ &= \sum_{k=1}^{2W(1-\delta)T}\frac{(1-\lambda_k)[x_k^{(1)}]^2}{E(\mathbf{x})} + \sum_{k=1}^{\infty}\frac{(1-\lambda_{2WT(1-\delta)+k})}{E(\mathbf{x})}[x_k^{(2)}]^2 \\ &\leqq [1-\lambda_{2WT(1-\delta)}]\frac{E(\mathbf{x}^{(1)})}{E(\mathbf{x})} + \frac{E(\mathbf{x}^{(2)})}{E(\mathbf{x})}.\end{aligned} \tag{54}$$

Now since $\lambda_{2WT(1-\delta)} \to 1$ as $T \to \infty$ (44a), and $E(\mathbf{x}^{(1)})/E(\mathbf{x}) \leqq 1$, with T sufficiently large we have

$$[1 - \lambda_{2WT(1-\delta)}] \frac{E(\mathbf{x}^{(1)})}{E(\mathbf{x})} \leqq \sigma.$$

Since $E(\mathbf{x}^{(2)})$ must satisfy (33b) (with $\hat{\eta} = \eta - \sigma$), we have (with T sufficiently large)

$$1 - K_D(s,T) \leqq \sigma + \eta - \sigma = \eta, \tag{55}$$

so that $s(t)$ satisfies (47c). Thus $s(t)$ belongs to $a_2(T,W,P_o)$

Now we may express the noise in a Karhunen-Loeve expansion as

$$z(t) = \sum_{k=1}^{\alpha_1 T} z_k^{(1)} \psi_k(t) + \sum_{k=1}^{\infty} z_k^{(2)} \psi_{\alpha_1 T+k}(t), \tag{56}$$

where again the ψ_k are PSWF's and the $\{z_k^{(i)}\}_{1 \leqq k < \infty}^{i=1,2}$ are independent normally distributed random variables with mean zero and variance $N = N_o/2$. The output signal $y(t) = s(t) + z(t)$ is

$$y(t) = \sum_{k=1}^{\alpha_1 T} y_i^{(1)} \psi_k(t) + \sum_{k=1}^{\infty} y_k^{(2)} \psi_{\alpha_1 T+k}(t), \tag{57}$$

where the $y_k^{(i)}$ are obtainable by integration from the signal $y(t)$. Further,

$$y_k^{(i)} = x_k^{(i)} + z_k^{(i)}, \tag{58}$$

so that we conclude that our time-continuous channel with signals constructed in this way is equivalent to the type 1 jointly-constrained product channel with parameters (51) and capacity C^* (see Appendix E). Since $R < C^*$, we may therefore construct codes with rate R for either channel with error probability $P_e \to 0$ as $T \to \infty$. This is the direct half of Theorem 2.

(*ii*) Weak Converse: Say we are given a sequence of codes for our time-continuous channel with parameters $\{T_i\}_{i=1}^{\infty}$, with code words belonging to $a_2(T_i, W, P_o)$ (as defined in (47)), with error probability $P_e^{(i)}$, and rate

$$R > W \ln\left(1 + (1 - \eta) \frac{P_o}{N_o W}\right) + \eta \frac{P_o}{N_o}. \tag{59}$$

We shall show that $P_e^{(i)}$ must be bounded away from zero so that the capacity C (the maximum permissible rate) cannot exceed the right member of (59).

Now as in the proof of the direct half we may (by (59)) find a $\delta > 0$ and $\sigma > 0$ sufficiently small so that

$$R > W(1 + \delta) \ln \left[1 + \left(1 - \frac{\eta}{1 - \sigma} \right) \frac{P_o}{N_o W(1 + \delta)} \right] + \frac{\eta}{1 - \sigma} \frac{P_o}{N_o} \triangleq C^*. \tag{60}$$

Again, as in the direct half, C^* is the capacity of the type 1 jointly-constrained product channel with parameters

$$P = P_o, \qquad N = N_o/2, \qquad \hat{\eta} = \frac{\eta}{1 - \sigma}, \qquad \alpha_1 = 2W(1 + \delta), \qquad \alpha_2 = \infty. \tag{61}$$

Now if $s(t)$ is a code word from the code with parameter T_i, (so that $s \,\varepsilon\, a_2(T_i, W, P_o))$, we may write $s(t)$ as a Fourier series in PSWF's (due to the completeness of the PSWF's on the space of band-limited functions) (Section 4.3),

$$s(t) = \sum_{k=1}^{2WT_i(1+\delta)} x_k^{(1)} \psi_k(t) + \sum_{k=1}^{\infty} x_k^{(2)} \psi_{k+2WT_i(1+\delta)}(t), \qquad -\infty < t < \infty. \tag{62}$$

Hence, to each code word $s(t)$ for the time-continuous channel, there corresponds a vector $\mathbf{x} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ whose coordinates are the coefficients in the above Fourier series. We now show that $\mathbf{x}$ is an allowable input to the type 1 jointly-constrained product channel with parameters given by (61). From the orthonormality of the PSWF's on $(-\infty, \infty)$ we have from (62), $\| s \|^2 = E(\mathbf{x})$. Since $s(t) \,\varepsilon\, a_2(T_i, W, P_o)$, we have $E(\mathbf{x}) \leqq PT_i$, so that $\mathbf{x}$ satisfies (33a). Further, from the orthogonality of the PSWF's on $[-T/2, T/2]$ and the monotonicity of the λ_k we have

$$\begin{aligned} 1 - K_D(s,T_i) &= \frac{\| (1 - D_{T_i}) s \|^2}{\| s \|^2} \\ &= \sum_{k=1}^{2WT_i(1+\delta)} \frac{[x_k^{(1)}]^2 (1 - \lambda_k)}{E(\mathbf{x})} + \sum_{k=1}^{\infty} \frac{[x_k^{(2)}]^2 (1 - \lambda_{2WT_i(1+\delta)+k})}{E(\mathbf{x})} \\ &\geqq [1 - \lambda_{2WT_i(1+\delta)}] \frac{E(\mathbf{x}^{(2)})}{E(\mathbf{x})}. \end{aligned} \tag{64}$$

With T_i sufficiently large (from 44b) we may put $\lambda_{2WT_i(1+\delta)} \leqq \sigma$, and since $1 - K_D(s,T_i) \leqq \eta$,

$$E(\mathbf{x}^{(2)}) \leqq \frac{\eta}{1 - \sigma} E(\mathbf{x}) = \hat{\eta} E(\mathbf{x}), \tag{65}$$

so that $\mathbf{x}^{(2)}$ satisfies (33b).

Finally, if we proceed as in the proof of the direct half of this theorem and express the noise in a Karhunen-Loeve expansion in PSWF's, we can conclude that for each code for this time-continuous channel we can obtain a code for the time-discrete jointly-constrained product channel with the same rate and error probability (see Appendix E). Since the rate R exceeds the capacity of the latter channel we conclude from the weak converse to Theorem 1 that the error probability is bounded away from zero. This completes the proof.

The following theorems establish the capacity of the channels defined by Models 1 and 3.

Theorem 3: (Model 1) Let the allowable signal set be $a_1(T,W,P_o)$ *the set of functions* $s(t)$ *satisfying*

$$s(t) = 0, \qquad |t| \geqq T/2, \tag{66a}$$

$$\| s \|^2 \leqq P_o T, \tag{66b}$$

$$K_B(s,2\pi W) \geqq 1 - \eta \quad (0 < \eta < 1). \tag{66c}$$

The noise is a sample from a Gaussian random process with spectral density

$$N(\omega) = \begin{cases} N_o/2 & & |\omega| \leqq 2\pi W, \\ \nu N_o/2 & (\nu \leqq 1) & |\omega| > 2\pi W. \end{cases} \tag{67}$$

Then the channel capacity is

$$C = C_{\eta,\nu} = W \ln \left(1 + (1 - \eta) \frac{P_o}{N_o W} \right) + \frac{\eta P_o}{\nu N_o}. \tag{68}$$

Theorem 4: (Model 3) Let the allowable signal set be $\alpha_3(T,W,P_o)$ *the set of functions* $s(t)$ *satisfying*

$$s(t) = 0, \qquad |t| \geqq T/2, \tag{69a}$$

$$\| s \|^2 \leqq P_o T, \tag{69b}$$

$$K_B'(s,2\pi W) \geqq 1 - \eta \quad (0 < \eta < 1), \tag{69c}$$

where K_B' *is the frequency concentration defined by (21). The noise is as in Theorem 2 (48). Then the channel capacity is*

$$C = C_\eta = W \ln \left(1 + \frac{P_o}{N_o W} \right) + \frac{\eta}{1 - \eta} \frac{P_o}{N_o}. \tag{70}$$

Proofs of Theorems 3, 4: Since the proofs of Theorems 3 and 4 parallel that of Theorem 2 (which was given in detail above) we shall confine ourselves to a few remarks which will enable the interested reader to fill in the details on his own.

Theorem 3: In the direct half we consider, as in the proof of Theorem 2, a jointly-constrained product channel. In this case it is a type-2 channel with parameters

$$\alpha_1 = 2W(1-\delta), \quad \alpha_2 = 0, \quad \alpha_3 = \infty, \quad P = P_o,$$
$$N_1 = \frac{N_o}{2}(1-\xi), \quad N_2 = \frac{\nu N_o}{2}, \quad \hat{\eta} = \eta - \sigma, \tag{71}$$

where $\xi,\delta,\sigma > 0$ are "small". In the present proof, this channel plays the role that the type-1 channel played in the proof of the direct half of Theorem 2. Since $\alpha_2 = 0$, we may write a channel input as $\mathbf{x} = (\mathbf{x}^{(1)},\mathbf{x}^{(3)})$. Corresponding to $\mathbf{x}$ we construct an input signal for our time-continuous channel as

$$s(t) = D_T\left[\sum_{k=1}^{2W(1-\delta)=\alpha_1 T} x_k^{(1)} \frac{\psi_k(t)}{\sqrt{\lambda_k}} + \sum_{k=1}^{\infty} x_k^{(3)} \frac{\psi_{k+W(1-\delta)T}}{\sqrt{\lambda_{k+2W(1-\delta)T}}}\right]. \tag{72}$$

where the ψ_k are PSWF's, the λ_k the associated eigenvalues (40), and D_T the time-truncation operator (14). Equation (72) replaces (52) in the proof of Theorem 2. It is easily verified that signals of the form (72) belong to $a_1(T, W, P_o)$ as defined by (66). If we write the noise in the expansion of (46) we can, as in Theorem 2, establish the equivalence of the time-discrete and time-continuous channels, and establish the direct-half of Theorem 3. The weak converse is proved in a similar manner, the jointly-constrained product channel employed here being of type-2 with parameters

$$\alpha_1 = 2W(1-\delta), \quad \alpha_2 = 4W\delta, \quad \alpha_3 = \infty, \quad P = P_o,$$
$$N_1 = \frac{N_o}{2}, \qquad N_2 = \frac{N_o}{2}, \tag{73}$$
$$N_3 = (\nu - \xi)\frac{N_o}{2}, \qquad \eta = \frac{\eta}{1-\sigma},$$

where again $\delta,\xi,\sigma > 0$ are "small".

Theorem 4: For the direct-half we consider a type-3 jointly-constrained product channel with parameters

$$\alpha_1 = 2W(1-\delta), \quad \alpha_2 = \infty, \quad P = P_o, \quad N = \frac{N_o}{2}, \quad \hat{\eta} = \eta - \sigma. \tag{74}$$

The signals are constructed from vectors $\mathbf{x}$ as in (72). For the converse we use a type-3 channel with parameters

$$\alpha_1 = 2W(1 + \delta), \quad \alpha_2 = \infty, \quad P = P_o, \quad N = \frac{N_o}{2}, \quad \hat{\eta} = \frac{\eta}{1 - \sigma}. \tag{75}$$

APPENDIX A

Proof of Theorem 1

We shall give a proof of Theorem 1 for the type-1 jointly-constrained product channel only. The proofs for types 2 and 3 are similar.

The proof as usual is in two parts.

A.1 *Direct Half*

We set $P_1 = (1 - \beta)P$, $P_2 = \beta P$ and consider codes for the ordinary product channel (Section 4.1). If (30) is satisfied for all code words with these values of P_1 and P_2, then the joint constraint (33a) is also satisfied. Further since $\beta \leqq \hat{\eta}$, (33b) is also satisfied. Hence the direct half of Lemma A for the ordinary product channel implies that any rate less than $C_1(P_1) + C_2(P_2) = C_1((1 - \beta)P) + C_2(\beta P)$ is permissible, and the direct-half of Theorem 1 follows.

A.2 *Converse*

Let us define $C^* = C_1((1 - \beta)P) + C_2(\beta P)$. We must show that any rate $R > C^*$ is not permissible. Let us assume the contrary, i.e.; for some $R = C^* + \varepsilon(\varepsilon > 0)$, there exists a sequence of numbers $\{T_i\}_{i=1}^{\infty}$ where $T_i \to \infty$ as $i \to \infty$, and a corresponding sequence of codes for the jointly constrained product channel (satisfying (33a) and (33b), with parameters $\hat{\eta}$ and P); with the ith code ($i = 1,2, \cdots$) having parameter $T = T_i$ and e^{RT_i} code words, and error probability $P_e = P_e^{(i)}$ where $P_e^{(i)} \to 0$ as $i \to \infty$.

Since $C_1(x)$ is uniformly continuous on the closed interval $[O,P]$, let us choose an integer J_o (sufficiently large) so that

$$\left| C_1(x) - C_1\left(x - \frac{\hat{\eta}P}{J_o}\right)\right| < \frac{\varepsilon}{2}, \qquad 0 \leqq x \leqq \hat{\eta}P. \tag{76}$$

We now partition the ith code ($i = 1,2, \cdots$) into J_o classes $S_i(j)$ ($j = 1,2, \cdots, J_o$). A code word $(\mathbf{x}^{(1)},\mathbf{x}^{(2)})$ in the ith code will belong to the jth class $S_i(j)$, according as the energy of its second component satisfies

$$(j - 1)\frac{\hat{\eta}PT}{J_o} < \sum_{k=1}^{n_2} [x_k^{(2)}]^2 \leqq \frac{j\hat{\eta}PT_i}{J_o}, \qquad j = 1,2, \cdots, J_o. \tag{77}$$

Since $\mathbf{x}^{(2)}$ satisfies (33b), each code word belongs to exactly one class.

(To be precise, we assign code words for which the energy in $\mathbf{x}^{(2)}$ is zero to class $S_i(1)$.)

For each $i(i = 1,2, \cdots)$, let S_i^* be the subcode of the ith code (with parameter $T = T_i$) consisting of the class $S_i(j)\,(j = 1,2, \cdots, J_o)$ containing the most members. Since S_i^* is the largest class in a partition of a code with e^{RT_i} code words into J_o classes, the number of code words in $S_i^* \geqq e^{RT_i}/J_o$, so that the corresponding transmission rate for S_i^* is

$$R^* \geqq R - \frac{1}{T_i} \ln J_o . \tag{78}$$

Further, since S_1^* is a subcode of the ith code (which has error probability $P_e^{(i)}$), the error probability of S_i^* is not more than $P_e^{(i)}$.

Since there are a finite number (J_o) of classes in the partition of the ith code $(i = 1,2, \cdots)$, there must be at least one j_o $(1 \leqq j_o \leqq J_o)$ such that for an infinite number of i, the largest partition S_i^* is the j_oth partition $S_i(j_o)$. Let $(i_1, i_2, \cdots)$ be the subsequence of i's for which $S_i^* = S_i(j_o)$. Thus the $\{S_{i_t}^*\}_{t=1}^{\infty}$ are a sequence of codes with rate R^* satisfying (78), and error probability not more than $P_e^{(i_t)}$, where $P_e^{(i_t)} \to 0$ as $t \to \infty$. Further, if a code word $(\mathbf{x}^{(1)},\mathbf{x}^{(2)}) \,\varepsilon\, S_{i_t}^*$, it belongs to the class $S_{i_t}(j_o)$, so that from (77) the energy of the second component satisfies

$$E(\mathbf{x}^{(2)}) = \sum_{k=1}^{n_2} [x_k^{(2)}]^2 \leqq \frac{j_o \hat{\eta} P T_{i_t}}{J_o}, \tag{79}$$

and from (77) and (33a), the energy of the first component satisfies

$$E(\mathbf{x}^{(1)}) = \sum_{k=1}^{n_1} [x_k^{(1)}]^2 \leqq \left[1 - \frac{(j_o - 1)\hat{\eta}}{J_o}\right] P T_{i_t} . \tag{80}$$

We conclude that $\{S_{i_t}^*\}_{t=1}^{\infty}$ is a sequence of codes which satisfy the constraints for the ordinary product channel (30) with parameters

$$P_1 = [1 - \{(j_o - 1)/J_o\}\hat{\eta}]\, P \quad \text{and} \quad P_2 = (j_o\hat{\eta}/J_o)\, P.$$

Since the error probability for $S_{i_t}^*$, $P_e^{(i_t)} \to 0$ as $t \to \infty$, we conclude that the rate R^* is a permissible rate for the ordinary product channel. By the converse half of Lemma A we have that R^* does not exceed the capacity of this product channel, i.e.,

$$R^* \leqq C_1\left(\left(1 - \frac{(j_o - 1)}{J_o}\,\hat{\eta}\right)P\right) + C_2\left(\frac{j_o\hat{\eta}}{J_o}\,P\right), \tag{81}$$

where $C_i(x)\,(i = 1,2)$ is defined by (37b). Applying (76) to (81) we obtain

$$R^* \leqq C_1((1 - \delta)P) + C_2(\delta P) + \frac{\varepsilon}{2}, \tag{82}$$

where $\delta = j_o\hat{\hat{\eta}}/J_o$. Now it follows immediately (by differentiation) from the definition of $C_1(x)$ and $C_2(x)$ that if $\alpha_2 \geqq \alpha_1$, $f(\delta) \triangleq C_1((1 - \delta)x) + C_2(\delta x)$ is an increasing function for δ for $\delta < \alpha_2/(\alpha_1 + \alpha_2)$, and $f(\delta)$ is a decreasing function of δ for $\delta > \alpha_2(\alpha_1 + \alpha_2)$. We conclude that since $\delta = (j_o/J_o)\hat{\hat{\eta}} \leqq \hat{\hat{\eta}}$,

$$C_1((1 - \delta)P) + C_2(\delta P) \leqq C_1((1 - \beta)P) + C_2(\beta P) = C^*, \tag{83}$$

where $\beta = \min(\hat{\hat{\eta}}, \alpha_2/(\alpha_1 + \alpha_2))$. Combining (78), (82) and (83), we obtain

$$R \leqq C^* + \frac{\varepsilon}{2} + \frac{1}{T_{i_t}} \ln J_o. \tag{84}$$

If we let $t \to \infty$, then $T_{i_t} \to \infty$ and have from (84)

$$R \leqq C^* + \frac{\varepsilon}{2}.$$

But $R = C^* + \varepsilon$, and the contradiction establishes the weak converse to Theorem 1.

APPENDIX B

Proof of Theorem 5

Theorem 5: Let $\mathfrak{a}(T,W,P_o)$ *be the set of all* $s(t)$ *satisfying*

$$(i) \qquad s(t) = 0, \qquad |t| > T/2, \tag{85a}$$

$$(ii) \qquad \|s\|^2 \leqq P_oT, \tag{85b}$$

$$(iii) \qquad K_B(s,2\pi W) \geqq 1 - \eta \quad (0 < \eta < 1). \tag{85c}$$

Let the Gaussian noise be additive with spectral density $N(\omega)$ *where*

$$\int_{-\infty}^{\infty} N(\omega)d\omega = \bar{N} < \infty \tag{86}$$

Then $C_\eta = \infty$ *(all* η*).*

Proof: Let $R > 0$ and $\varepsilon > 0$, and $\eta(0 < \eta \leqq 1)$ be specified and fixed. We shall construct a code satisfying (85) with $M = e^{RT}$ code words with error probability $P_e \leqq \varepsilon$.

To begin with let us choose T sufficiently large so that

$$\frac{1}{\sqrt{4\pi RT}} \leqq \varepsilon, \tag{87a}$$

$$\lambda_1 \geqq 1 - \frac{\eta}{2}, \tag{87b}$$

where λ_1 is the first PSWF eigenvalue (40). With T fixed we now construct the code.

Let us expand the noise in a series of PSWF's

$$z(t) = \sum_{k=1}^{\infty} z_k \frac{\psi_k(t)}{\sqrt{\lambda_k}}, \qquad -\frac{T}{2} \leqq t \leqq \frac{T}{2}, \tag{88}$$

where

$$z_k = \int_{-T/2}^{T/2} z(t) \frac{\psi_k(t)}{\sqrt{\lambda_k}} dt, \tag{89}$$

and the $\{z_k\}_{k=1}^{\infty}$ are Gaussian random variables with mean zero, but not necessarily independent.

Now from (86) we have

$$\mathcal{E} \int_{-T/2}^{T/2} z^2(t)dt = \bar{N}T, \tag{90}$$

where "$\mathcal{E}$" denotes expectation. From the orthogonality of the PSWF's (41) we have from (88)

$$\bar{N}T = \mathcal{E} \int_{-T/2}^{T/2} z^2(t)dt = \sum_{k=1}^{\infty} \mathcal{E}(z_k^2). \tag{91}$$

Thus we can find an integer K sufficiently large so that

$$\mathcal{E}(z_{K+i}^2) \leqq \frac{\eta P_o}{16R}, \qquad i = 1,2, \cdots, M. \tag{92}$$

With K so chosen, let the M code words be

$$s_i(t) = D_T\left[\sqrt{P_oT\left(1 - \frac{\eta}{2}\right)} \frac{\psi_1(t)}{\sqrt{\lambda_1}} + \sqrt{\frac{\eta}{2} P_oT} \frac{\psi_{K+i}(t)}{\sqrt{\lambda_{K+i}}}\right], \tag{93}$$

$$i = 1,2, \cdots, M$$

Let us first verify that $s_i(t)$, as given by (93), satisfies (85). Equation (85a) follows from the definition of D_T (14). From the orthogonality of PSWF's (41) we have

$$\| s_i \|^2 = \left(1 - \frac{\eta}{2}\right) P_oT + \frac{\eta}{2} P_oT = P_oT \tag{94}$$

so that (85b) is satisfied. Finally,

$$K_B(s_i, 2\pi W) = \frac{P_oT\left(1 - \frac{\eta}{2}\right)\lambda_1 + \frac{\eta}{2} P_oT\lambda_{K+i}}{\| s_i \|^2}$$
$$\geqq \frac{P_oT}{P_oT}\left(1 - \frac{\eta}{2}\right)\lambda_1 \geqq \left(1 - \frac{\eta}{2}\right)\left(1 - \frac{\eta}{2}\right) \tag{95}$$
$$\geqq 1 - \eta,$$

where the next to last inequality follows from (87b). Thus $s_i(t) \varepsilon \mathfrak{a}(T,W,P_o)$. It remains to show that $P_e \leqq \varepsilon$.

We can write the received signal $y(t)$ in a Fourier series in PSWF's

$$y(t) = S_i(t) + z(t) = \sum_{k=1}^{\infty} y_k \frac{\psi_k(t)}{\sqrt{\lambda_k}}, \tag{96}$$

where the y_k are recoverable from $y(t)$ by integration. Say that the receiver disregards all the y_k except $y_{K+1}, y_{K+2}, \cdots, y_{K+M}$. We may write

$$y_{K+j} = \begin{cases} z_{K+j} + \sqrt{\frac{\eta}{2} P_oT}, & j = i, \\ z_{K+j}, & j \neq i, \end{cases} \tag{97}$$
$$(j = 1,2, \cdots, M).$$

If y_{K+i} is the maximum of the $\{y_{K+j}\}_{j=1}^{M}$, the receiver decodes $y(t)$ as $s_i(t)$. Thus if code word i is transmitted, the error probability is

$$P_{ei} = P_r \bigcup_{j \neq i} \left[z_{K+j} > z_{K+i} + \sqrt{\frac{\eta}{2} P_oT} \right]$$
$$\leqq MP_r \left[z_{K+j} - z_{K+i} > \sqrt{\frac{\eta}{2} P_oT} \right]. \tag{98}$$

Now $z_{K+j} - z_{K+i}$ is Gaussian, with mean zero and variance

$$\mathcal{E}((z_{K+j} - z_{K+i})^2) \leqq [E(z_{K+j}^{\,2})^{\frac{1}{2}} + E(z_{K+i}^{\,2})^{\frac{1}{2}}]^2$$
$$\leqq \frac{\eta P_o}{4R}, \tag{99}$$

where the last inequality follows from (92). Thus (98) becomes

$$P_e \leqq M \operatorname{erf}(-\sqrt{2RT}), \tag{100}$$

where

$$\text{erf }(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

is the cumulative error function. Since erf $(-x) \leqq e^{-x^2/2}/(\sqrt{2\pi}x)$, (100) yields with the help of (87a)

$$P_e \leqq \frac{e^{RT}e^{-RT}}{\sqrt{4\pi RT}} = \frac{1}{\sqrt{4\pi RT}} \leqq \varepsilon. \tag{101}$$

Thus the theorem is proven.

APPENDIX C

In this appendix we verify inequality (22)

$$K_B(x,2\pi W) \geqq 1 - 2\sqrt{\frac{1 - K_B'(x,2\pi W)}{K_B'(x,2\pi W)}}, \tag{102}$$

where K_B is defined by (13a) and K_B' by (21). Let $f(t)$ be a function with Fourier transform $F(\omega)$, and define the operator B by

$$g = Bf, \tag{103a}$$

where

$$g(t) = \frac{1}{2\pi} \int_{-2\pi W}^{2\pi W} F(\omega) e^{i\omega t} d\omega. \tag{103b}$$

Thus Bf is the result of passing f through an ideal low-pass filter with bandpass W cycles per second. Then

$$K_B(f,2\pi W) = \frac{\| Bf \|^2}{\| f \|^2}. \tag{104}$$

Say that $x(t) = 0$, $|t| \leqq T/2$ and $\| x \|^2 < \infty$. We assume that we may write $x = D_T \hat{x}$, where $\hat{x}$ is bandlimited to W cycles per second. (If we cannot then $K_B'(x,2\pi W) = 0$, and (102) follows immediately.) Let us write

$$\hat{x}(t) = x(t) + y(t), \tag{105}$$

where $y(t) = 0$, $|t| \leqq T/2$. Then

$$\| \hat{x} \|^2 = \| x \|^2 + \| y \|^2, \tag{106}$$

and from the definition of K_B',

$$K_B'(x,2\pi W) = \frac{\| x \|^2}{\| \hat{x} \|^2}. \tag{107}$$

Hence, from (107) and (106),

$$\frac{\| y \|^2}{\| x \|^2} = \frac{1 - K_B'(x,2\pi W)}{K_B'(x,2\pi W)} . \tag{108}$$

Now, since $\hat{x}$ is bandlimited, $B\hat{x} = \hat{x}$ and we have

$$\begin{aligned} \| \hat{x} \|^2 = \| B\hat{x} \|^2 = \| Bx + By \|^2 &\leqq [\, \| Bx \| + \| By \| \,]^2 \\ &\leqq [\, \| Bx \| + \| y \| \,]^2 = \| Bx \|^2 + \| y \|^2 + 2 \| Bx \| \| y \| . \end{aligned} \tag{109}$$

Combining (106) and (109) we have

$$\| x \|^2 + \| y \|^2 = \| \hat{x} \|^2 \leqq \| Bx \|^2 + \| y \|^2 + 2 \| Bx \| \| y \| , \tag{110}$$

so that (from (104))

$$K_B(x,2\pi W) = \frac{\| Bx \|^2}{\| x \|^2} \geqq 1 - 2 \frac{\| Bx \| \| y \|}{\| x \|^2} \geqq 1 - 2 \frac{\| y \|}{\| x \|} . \tag{111}$$

Finally, from (108) and (111) we have

$$K_B(x,2\pi W) \geqq 1 - 2 \sqrt{\frac{1 - K_B'(x,2\pi W)}{K_B'(x,2\pi W)}} . \tag{112}$$

This is inequality (102).

APPENDIX D

The Capacity of Model 4

To establish the capacity of the channel defined by Model 4 we must, as always, prove a direct-half and converse. In this appendix we give an outline of the proof of the direct-half, and a remark about the proof of the converse.

D.1 *Direct-Half*

Let $R < W \ln [1 + (P_o/N_oW)]$ be given. We show here that for ν sufficiently small we may construct codes for Model 4 with rate R and with vanishing error probability (as $T \to \infty$). By the continuity of the "ln" function we may find a $\delta > 0$, $a > 0$ sufficiently small so that

$$R < W(1 - \delta) \ln \left[1 + \frac{P_o(1 - a)}{N_oW(1 - \delta)} \right] = C^* . \tag{113}$$

We observe that C^* is the capacity of a single time-discrete channel

(Section 2.1) with parameters

$$P = P_o(1 - a), \qquad N = N_o/2, \qquad \alpha = 2W(1 - \delta). \quad (114)$$

Since $R < C^*$, we can find a code $\mathcal{C} = \{\mathbf{x}_i\}_{i=1}^M$ for this time-discrete channel (so that $E(\mathbf{x}_i) \leqq P_o(1 - a)T$) with $M = e^{RT}$ code words, and with error probability given that $\mathbf{x}_i$ is transmitted ($i = 1,2, \cdots, M$) (using the minimum distance decoder)

$$\begin{aligned} P_{ei} &= \Pr \bigcup_{j \neq i} [d_E(\mathbf{x}_i, \mathbf{y}) > d_E(\mathbf{x}_j, \mathbf{y})] \\ &= \Pr \bigcup_{j \neq i} \left[\| \mathbf{z}^{ij} \| > \frac{d_{ij}}{2} \right] = e^{-\beta T + o(T)}, \end{aligned} \quad (115)$$

where $\mathbf{y}$ is the received vector, $d_E(\mathbf{u},\mathbf{v})$ is the Euclidean distance between n-vectors $\mathbf{u}$ and $\mathbf{v}$, $d_{ij} = d_E(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{z}^{ij}$ is the projection of the noise vector $\mathbf{z}$ on the line passing through code words $\mathbf{x}_i$ and $\mathbf{x}_j$, and $\| \mathbf{u} \| = [E(\mathbf{u})]^{\frac{1}{2}}$ is the square root of the sum of the squares of the components of $\mathbf{u}$. The exponent β has been estimated by Shannon.[3] Since $\| \mathbf{z}^{ij} \|$ is a Gaussian random variable with mean zero and variance $N_o/2$ we may lower bound P_{ei} by

$$P_{ei} \geqq \Pr \left[\| \mathbf{z}^{ij} \| > \frac{d_{ij}}{2} \right] = \operatorname{erf} \left(- \frac{d_{ij}}{\sqrt{2N_o}} \right), \quad (j = 1,2, \cdots, M \quad j \neq i) \quad (116)$$

where

$$\operatorname{erf} x = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

is the cumulative error function. Since for large x,

$$\operatorname{erf}(-x) \approx \left(\frac{1}{\sqrt{2\pi x}} \right) e^{-x^2/2},$$

(115) and (116) yield for large T

$$d_{ij}^2 \geqq 4\beta N_o T, \qquad i,j = 1,2, \cdots, M \qquad i \neq j. \quad (117)$$

From the code $\mathcal{C}$, let us construct a new code $\hat{\mathcal{C}} = \{\mathbf{x}_i\}_{i=1}^M$, where

$$\hat{\mathbf{x}}_i = \frac{1}{1 - a} \mathbf{x}_i, \qquad i = 1,2, \cdots, M. \quad (118)$$

Thus the members $\hat{\mathbf{x}}_i$ of $\hat{\mathcal{C}}$ satisfy

$$E(\hat{\mathbf{x}}_i) \leqq P_o T. \quad (119)$$

Let us now assume that there are two noises in the channel, i.e., the noise vector $\mathbf{z} = \mathbf{z}_1 + \mathbf{z}_2$. The first noise $\mathbf{z}_1$ is the usual spherical Gaussian noise (with variance $N_o/2$), and the second $\mathbf{z}_2$ is an unknown n-vector ($n = \alpha T = 2W(1 - \delta)T$) for which we require only

$$E(\mathbf{z}_2) \leqq \nu N_o W(1 - \delta)T = \nu \frac{N_o}{2} n. \tag{120}$$

We place no other restrictions on the probability structure of $\mathbf{z}_2$. The vector $\mathbf{z}_2$ may depend on the code $\hat{\mathfrak{C}}$, the code word transmitted and the value of $\mathbf{z}_1$. The noise vector $\mathbf{z}_1$ corresponds to the noise function $z_1(t)$ in Model 4, and the noise vector $\mathbf{z}_2$ corresponds to $z_2(t)$ in Model 4. If we use $\hat{\mathfrak{C}}$ on the time-discrete channel with this noise and use the minimum distance decoder, we have an error probability given that $\hat{\mathbf{x}}_i$ is transmitted

$$\begin{aligned}\hat{P}_{ei} &= \Pr \bigcup_{j \neq i} \left[d_E^2(\hat{\mathbf{x}}_i, \hat{\mathbf{y}}) > d_E(\hat{\mathbf{x}}_j, \hat{\mathbf{y}}) \right] \\ &= \Pr \bigcup_{j \neq i} \left[\| (\mathbf{z}_1 + \mathbf{z}_2)^{ij} \| > \frac{\hat{d}_{ij}}{2} \right]\end{aligned} \tag{121}$$

where

$$\hat{d}_{ij} = d_E(\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j) = \frac{d_{ij}}{(1 - a)}.$$

Now since " $\| \ \|$ " is a norm

$$\| (\mathbf{z}_1 + \mathbf{z}_2)^{ij} \| \leqq \| \mathbf{z}_1^{ij} \| + \| \mathbf{z}_2^{ij} \| \leqq \| \mathbf{z}_1^{ij} \| + \sqrt{\nu N_o W(1 - \delta)T}. \tag{122}$$

Thus the event

$$\begin{aligned}&\left[\| (\mathbf{z}_1 + \mathbf{z}_2)^{ij} \| > \frac{\hat{d}_{ij}}{2} \right] \\ &\qquad \subseteq \left[\| \mathbf{z}_1^{ij} \| > \frac{d_{ij}}{2(1 - a)} - \sqrt{\nu N_o W(1 - \delta)T} \right],\end{aligned} \tag{123}$$

where "$\subseteq$" denotes set inclusion. Now we would like to say that the right member of (123)

$$\left[\| \mathbf{z}_1^{ij} \| > \frac{d_{ij}}{(1 - a)2} - \sqrt{\nu N_o W(1 - \delta)T} \right] \subseteq \left[\| \mathbf{z}_1^{ij} \| > \frac{d_{ij}}{2} \right]. \tag{124}$$

If this is so, then $\hat{P}_{ei} \leqq P_{ei} \to 0$ as $T \to \infty$. In fact (124) is satisfied if

$$\frac{d_{ij}}{2} \leqq \frac{d_{ij}}{2(1 - a)} - \sqrt{\nu N_o W(1 - \delta)T}, \tag{125}$$

or

$$\nu \leqq \frac{d_{ij}}{\sqrt{4N_oW(1-\delta)T}} \left(\frac{a}{1-a}\right). \tag{126}$$

Now from (117), $d_{ij} \geqq \sqrt{4\beta N_o T}$ so that if

$$\nu \leqq \sqrt{\frac{\beta}{w(1-\delta)}} \left(\frac{a}{1-a}\right), \tag{127}$$

(126) is satisfied. Hence $\hat{P}_{ei} \xrightarrow{T} 0$.

If we now make the same correspondence between the time-continuous channel and the time-discrete channel which was made in the proof of Theorem 3, we deduce the existence of codes for Model 4 {with rate $R < W \ln [1 + (P_o/N_oW)]$} with $P_e \to 0$ as $T \to \infty$ (provided ν is sufficiently small — the choice of ν depending on W, P_o/N_o, and R). Note that this construction was done for any η. Thus we have shown in effect that the capacity of Model 4 is

$$C = C_{\eta,\nu} \geqq W \ln \left(1 + \frac{P_o}{N_oW}\right) + \varepsilon_1(\nu), \tag{128}$$

where $\varepsilon_1(\nu) \to 0$ as $\nu \to 0$ independent of η.

D.2 *Converse*

The proof of the converse also parallels the proofs of the converse halves of Theorems 2, 3, and 4. However, since the noise may depend on the entire code and decoding scheme used (which is not the usual assumption of information theory coding theorems), it is necessary to go back and re-prove Theorem 1 (which in turn depends on Lemma A) for this new situation. Although this task is not a terribly difficult one it is rather tedious and we shall side step this chore here. It will suffice to state the version of Lemma A which is required here and to leave the rest of the proof to the interested reader.

Lemma A′: Let us say that we are given time-discrete channel as defined in Section I (with parameters α,P) where the noise vector is $\mathbf{z} = \mathbf{z}_1 + \mathbf{z}_2$ where $\mathbf{z}_1$ is the usual spherical Gaussian noise with variance N and $\mathbf{z}_2$ is an unknown vector for which require only

$$E(\mathbf{z}_2) \leqq \xi T. \tag{129}$$

We place no other restriction on the probability structure of $\mathbf{z}_2$. The noise vector $\mathbf{z}_2$ may depend on the entire code and decoding scheme, the code word transmitted and the value of $\mathbf{z}_1$. We define the error probability P_e as we did in (26) for Model 4 and do likewise for the capacity. Let $C(\alpha,P,N,\xi)$ be the capacity of this channel.

Now consider the product of r time-discrete channels as in Section 4.1 with parameters (α_i, P_i, N_i) $i = 1, 2, \cdots, r$. *Here too, we assume a second noise vector*

$$\mathbf{z}_2 = (\mathbf{z}_2^{(1)}, \mathbf{z}_2^{(2)}, \cdots, \mathbf{z}_2^{(r)}), \tag{130}$$

which is unknown but must satisfy

$$\sum_{i=1}^{r} E(\mathbf{z}_2^{(i)}) \leqq \xi T, \tag{131}$$

and as above may depend on the entire code and decoding scheme, the code word transmitted, and the values of the spherical Gaussian noises.

Lemma A′ states that the capacity C^* *of this channel satisfies*

$$C^* \leqq \sum_{i=1}^{r} C(\alpha_i, P_i, N_i, \gamma_i \xi), \tag{132a}$$

where

$$\sum_{i=1}^{r} \gamma_i = 1. \tag{132b}$$

APPENDIX E

Equivalence of Time-Discrete and Time-Continuous Models

In this appendix, we give some details on the validity of the equivalence of the time-discrete and time-continuous channel models which is the key to the proofs of our capacity theorems.

To begin with, let us consider the direct-half of our theorems. In these proofs we deduce the existence of time-continuous coding and decoding schemes from the existence of time-discrete coding and decoding schemes. To be specific let us consider the proof of the direct half of Theorem 2. We may omit the reference to the Karhunen-Loeve expansion (5.10) and consider the received signal $y(t) = s_i(t) + z(t)$. Now it follows from Loeve (Ref. 9, p. 472, A) that

$$\mathcal{E} \int_{-T/2}^{T/2} z^2(t)\,dt = \int_{-T/2}^{T/2} R(0)\,dt = N_0 W T < \infty, \tag{133}$$

so that with probability 1, $z(t)$ and, therefore, $y(t)$ is square-integrable. It then follows that the integrals

$$y_k^{(1)} = \frac{1}{\sqrt{\lambda_k}} \int_{-T/2}^{T/2} y(t)\psi_k(t)\,dt \quad \text{and}$$

$$y_k^{(2)} = \frac{1}{\sqrt{\lambda_{\alpha_1 T+k}}} \int_{-T/2}^{T/2} y(t)\psi_{\alpha_1 T+k}(t)\,dt \tag{134}$$

(where $\psi_k(t)$ and the λ_k are the kth PSWF and eigenvalue, respectively) exist for all k with probability 1. Further, it follows directly on substituting $y(t) = s_i(t) + z(t)$ into (134) that

$$y_k^{(i)} = s_{ik} + z_k^{(i)}, \qquad i = 1,2, \tag{135}$$

where the $z_k^{(i)}$ are independent normally distributed random variables with mean zero and variance $N_0/2$. Thus, the decoder for the time-continuous code may obtain the $y_k^{(i)}$ from the $y(t)$ and make use of the decoding scheme for the time-discrete code and obtain the same error probability. Hence, the direct-half of this and the subsequent theorems is valid.

Let us now consider the converse half of our theorems. In each of these proofs we assume that for a fixed rate R exceeding capacity, we are given a sequence of codes for the time-continuous channel with error probability P_e. We must show that P_e is bounded away from zero. To do this we deduce the existence of a corresponding sequence of codes with rate R and error probability P_e for a time-discrete channel with the same capacity as the time-continuous channel. Since we can invoke a converse for this time-discrete channel (Theorem 1), we then conclude that P_e is bounded away from zero. We will now show how to make this correspondence precise. Again let us refer specifically to the proof of Theorem 2, the others following similarly.

Let $\{s_i(t)\}_{i=1}^{M}$ be the code for the time-continuous channel, and $\mathbf{x} = (\mathbf{x}^{(1)},\mathbf{x}^{(2)})$ be the corresponding input to the time-discrete (product) channel. Further, we may write the noise signal $z(t)$ and the received signal $y(t)$ in Fourier series in PSWF's where, as above, all the coordinates are finite with probability 1. We then let $\mathbf{z} = (\mathbf{z}^{(1)},\mathbf{z}^{(2)})$ and $\mathbf{y} = (\mathbf{y}^{(1)},\mathbf{y}^{(2)})$ be the vectors whose coordinates are the coefficients in these expansions. We can easily show that

$$\mathbf{y} = \mathbf{x} + \mathbf{z}, \tag{136}$$

where the coordinates of $\mathbf{z}$ are independent random variables with mean zero and variance $N_0/2$. Thus, we have established the correspondence of the time-continuous and time-discrete channels and codes. We must now show that the time-continuous and time-discrete codes have the same error probability. In other words, we mush show that there exists a decoding scheme for the $\mathbf{y}$ which has the same error probability as the decoding scheme for the continuous received signal $y(t)$. We proceed as follows:

Let $\mathcal{B}$ be the usual (Kolmogorov) σ-algebra on $\mathcal{L}_2[-\bar{T}/2,\bar{T}/2]$, i.e., $\mathcal{B}$ is the σ-algebra generated by the "intervals" of the form

$$\{y(t): y(t_1) \leqq \rho_1, y(t_2) \leqq \rho_2, \cdots, y(t_n) \leqq \rho_n\}.$$

Corresponding to the code for the time-continuous channel $\{s_i(t)\}_{i=1}^{M}$, we define M probability measures $P_1, P_2, \cdots, P_M$ on $\mathcal{B}$ as follows. If $B \in \mathcal{B}$, then

$$P_i(B) = \text{Prob}\,[(s_i(t) + z(t)) \in \mathcal{B}], \tag{137}$$

where the probability in (137) is computed for $z(t)$, a noise sample function. A decoding-coding rule for this code is a set of M disjoint $\Lambda_i \in \mathcal{B}$ ($i = 1,2, \cdots, M$), called decoding regions. The error probability given that $s_i(t)$ is transmitted is

$$P_{ei} = 1 - P_i(\Lambda_i). \tag{138}$$

Now let $\hat{\mathcal{B}} \subseteq \mathcal{B}$ be the sub-σ-algebra on $\mathfrak{L}_2[-T/2,T/2]$, consisting of those sets determined by the coefficients of a representation of a function in PSWF's. That is, if $y(t) \in \mathfrak{L}_2(-T/2,T/2)$, let

$$y_k^{(1)} = \frac{1}{\sqrt{\lambda_k}} \int_{-T/2}^{T/2} y(t)\psi_k(t)\,dt \quad \text{and} \quad y_k^{(2)} = \frac{1}{\sqrt{\lambda_k}} \int_{-T/2}^{T/2} y(t)\psi_{\alpha_1 T+k}(t)\,dt.$$

Then $\hat{\mathcal{B}}$ is the σ-algebra generated by intervals of the form

$$\{y(t): y_{k_1}^{(1)} \leqq \rho_1^{(1)}, y_{k_2}^{(1)} \leqq \rho_n^{(1)}, \cdots, y_{k_m} \leqq \rho_m^{(1)},$$
$$y_{j_1}^{(2)} \leqq \rho_1^{(2)}, y_{j_2}^{(2)} \leqq \rho_2^{(2)}, \cdots, y_{j_n} \leqq \rho_n^{(2)}\}.$$

A decoding rule for a time-discrete code with M code words is a set of M disjoint $\hat{\Lambda}_i \in \hat{\mathcal{B}}$ ($i = 1,2, \cdots, M$) (decoding regions), and the error probability given that vector $\mathbf{x}_i$ ($\mathbf{x}_i$ is the representation of $s_i(t)$ in PSWF's) is transmitted is

$$\hat{P}_{ei} = 1 - P_i(\hat{\Lambda}_i).$$

Kadota [Ref. 10, Appendix D] has shown that for each $\Lambda_i \in \mathcal{B}$, there exists a $\Lambda_i \in \hat{\mathcal{B}}$ such that

$$P(\Lambda_i \Delta \hat{\Lambda}_i) = 0,$$

where Δ denotes "symmetric difference". Thus, if $\{\Lambda_i\}_{i=1}^{M}$ are the decoding regions for a time-continuous code we can find a set $\{\hat{\Lambda}_i \in \hat{\mathcal{B}}\}_{i=1}^{M}$ of decoding regions for the corresponding time-discrete code such that the error probabilities $P_{ei} = \hat{P}_{ei}$.

We conclude that the error probability for the time-discrete code equals the error probability for the time-continuous code, and the converse is valid.

GLOSSARY

The following symbols are used throughout the paper:

M = the number of members of a code.
T = time required to transmit a code word.
R = $(1/T) \ln M$ = transmission rate in nats per second.
C = channel capacity = maximum "error free" rate.
P_{ei} = probability that the receiver makes an incorrect decoding decision when code word i is transmitted ($i = 1,2, \cdots, M$).
P_e = $(1/M) \sum_{i=1}^{M} P_{ei}$ = over-all error probability.
$\mathcal{E}(X)$ = expected value of the random variable X.
ψ_k, λ_k = kth prolate spheroidal wave function (PSWF) and eigenvalue respectively ($k = 1,2, \cdots$).

The following symbols are used in connection with time-discrete or time-continuous channels:

Time-Discrete Channels:

$\mathbf{x},\mathbf{y},\mathbf{z}$ = input, output, and noise vectors, respectively.
$n = \alpha T$ = dimension of above vectors, so that α is the rate at which the channel passes real numbers.
$E(\mathbf{x})$ = sum of the squares of the coordinates of the vector $\mathbf{x}$.
P = parameter constraining $E(\mathbf{x})$ ($\mathbf{x}$ is channel input).
N = variance of the normally distributed noise.
r = number of components in the product (or parallel combination) of channels.
$\mathbf{x}^{(i)}, \mathbf{y}^{(i)}, \mathbf{z}^{(i)}$ = input, output, and noise vectors, respectively for the ith component of a product of channels ($i = 1,2, \cdots, r$).
n_i, α_i, P_i, N_i = parameters n,α,P,N, respectively, for the ith component of a product of channels ($i = 1,2, \cdots, r$).
$\hat{\eta}$ = parameter constraining the relative values of $E(\mathbf{x}^{(i)})$ in the product of channels.

Time-Continuous Channels:

$s(t),y(t),z(t)$ = input, output, and noise signals, respectively.
$S(\omega)$ = Fourier transform of $s(t)$.

$$\| s \|^2 = \int_{-\infty}^{+\infty} s^2(t)dt = \text{"energy" of } s(t).$$

$$K_B(s,2\pi W) = \frac{1}{2\pi} \int_{-2\pi W}^{2\pi W} | S(\omega) |^2 d\omega / \| s \|^2$$

$$= \text{(energy) concentration in frequency band } O\text{-}W \text{ cps.}$$

$K_D(s, T) = \int_{-T/2}^{T/2} s^2(t)dt / \| s \|^2 =$ (energy) concentration in time interval $[-(T/2), (T/2)]$.

$K_B'(s,2\pi W) =$ an alternate measure of frequency concentration defined by (21).

$D_T =$ operator which truncates a signal outside the time interval $[-(T/2), (T/2)]$ (see (14)).

$\mathfrak{L}_2[-T/2,T/2] =$ the space of square integrable functions defined on $[-T/2,T/2]$.

$W =$ bandwidth of channel.

$P_o =$ average "power" of input signals.

$N_o =$ one-sided spectral density of noise $z(t)$.

$a = a_i(T,W,P_o) =$ set of allowable channel input signals (for Model i, $i = 1,2,3,4$). These signals are approximately time-limited to T secs, approximately band-limited to W cps, and have energy not exceeding P_oT.

$\eta =$ parameter which measures the extent to which signals in a are not strictly time or bandlimited.

$\nu =$ parameter which measures the extent to which the noise spectral density is not zero for $|\omega| > 2\pi W$.

ACKNOWLEDGMENT

I wish to thank D. Slepian for many stimulating discussions and helpful suggestions.

REFERENCES

1. Shannon, C. E., A Mathematical Theory of Communication, B.S.T.J., *27*, July and October, 1948, pp. 379–423, 623–656.
2. Shannon, C. E., Communication in the Presence of Noise, Proc. IRE, *37*, January, 1949, pp. 10–21.
3. Shannon, C. E., Probability of Error for Optimal Codes in the Gaussian Channel, B.S.T.J., *38*, May, 1959, pp. 611–656.
4. Ash, R. B., Capacity and Error Bounds for a Time-Continuous Gaussian Channel, Information and Control, *6*, March, 1963, pp. 14–27.
5. Ash, R. B., *Information Theory*, John Wiley and Sons, New York, 1965.
6. Wyner, A. D., The Capacity of the Product of Channels, submitted to Information and Control.
7. Slepian, D., Landau, H., and Pollak, H., Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty I, II, B.S.T.J., *40*, January, 1961, pp. 43–84.
8. Davenport, W. and Root, W., *Random Signals and Noise*, McGraw-Hill Book Company, Inc., New York, 1958.
9. Loeve, M., *Probability Theory*, D. Van Nostrand, Princeton, New Jersey, 1955.
10. Kadota, T. T., Optimum Reception of Binary Gaussian Signals, B.S.T.J., *43*, November, 1964, pp. 2767–2810.