Gaussian Multiaccess Channels with ISI: Capacity Region and Multiuser Water-Filling

Roger S. Cheng, *Member, IEEE,* and Sergio Verdú, *Fellow, IEEE*

Abstract—The capacity region of a two-user Gaussian multiaccess channel with intersymbol interference (ISI), where the inputs pass through respective linear systems and are then superimposed before being corrupted by an additive Gaussian noise process, is found. A novel geometrical method is given to obtain the optimal input power spectral densities and the capacity region. This method can be viewed as a nontrivial generalization of the single-user water-filling argument. We show that as in the traditional memoryless multiaccess channel, frequency-division multiaccess (FDMA), with optimally selected frequency bands for each user, achieves the total capacity of the K-user Gaussian multiaccess channel with ISI. However, the capacity region of the two-user channel with memory is, in general, not a pentagon unless the channel transfer functions for both users are identical.

Index Terms—Multiaccess channels, intersymbol interference, channels with memory, Gaussian channels, channel capacity.

I. INTRODUCTION

INFORMATION THEORETICAL LIMITS of memoryless channels have been studied extensively since Shannon in 1948. The capacities and the capacity regions of single-user and multiuser memoryless channels were found by Shannon [1] and Ahlswede [2] (see also [3], [4]), respectively. However, channels with memory did not receive as much attention as their memoryless counterparts. This is partly due to the fact that single-letter characterizations [5] for the capacities of channels with memory do not exist. Only a limiting expression [6] is known for the capacities of single-user channels with memory and a not uncommon misconception is to dismiss limiting expressions for capacity as uncomputable. The classical example where limiting expressions can be not only readily computed but lead to a celebrated solution is the single-user Gaussian channel with intersymbol interference (ISI) ([7], [8], [9], see also [10]).

In a recent paper [11], a limiting expression for the capacity regions of multiaccess channels with memory was obtained. Such a limiting expression was explicitly evaluated for some channels with memory in [11] and [12]. In particular, the

limiting expression is applied to the asynchronous code-division multiaccess (CDMA) channel (which can be viewed as a multiaccess channel with memory) [12] to obtain its capacity region. In this special case, the capacity region can be evaluated by solving an optimization problem with low computational complexity. In this paper, we show that the limiting expression of [11] can be used to obtain a computable capacity region formula for Gaussian linear vector multiple-access channels with finite ISI.

A natural extension of the single-user channel with ISI to the two-user case is shown in Fig. 1. However, we consider a general linear multiaccess channel model in Fig. 2 where the inputs and the output are vectors with possibly different dimensions. By linearity, we can obtain an equivalent model in Fig. 3 where there are two channels, one for each user, and their outputs are superimposed before being corrupted by noise. Throughout the paper, we concentrate on this channel model with the only assumptions being that the channel is linear and time-invariant with finite-length impulse response, and the noise is stationary Gaussian and m-dependent (i.e., the autocorrelation function has finite support, or in other words, any two noise samples that lie more than m positions apart are independent.) This model is a generalization of the classical memoryless multiaccess channel to channels with memory, as well as a generalization of the single-user channel with ISI to the two-user vector channel with ISI.

The capacity of a single-user Gaussian channel with ISI is obtained using the Karhunen–Loève expansion. This expansion decomposes the channel into independent parallel memoryless Gaussian channels whose capacities are well known; thereby reducing the problem to one of optimal power allocation into various channels. It is crucial to note that the kernel used in the Karhunen–Loève expansion depends on the ISI coefficients. In the two-user Gaussian channel with ISI as shown in Fig. 3, there are two sets of ISI coefficients, one for each user. If both linear systems are identical, the traditional procedures can be applied and the capacity region has been obtained in [13], [14]. However, this is a very restrictive assumption because it rules out any cases where both users do not see the same channel. If the sets of ISI coefficients are not the same, a similar decomposition into independent memoryless channels cannot be applied since no kernel can simultaneously decompose the signals from both users.

Therefore, in order to obtain the result in the multiuser case, a new approach has to be used. It turns out that the circular
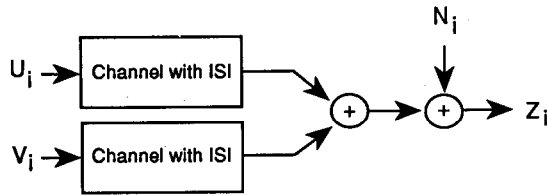
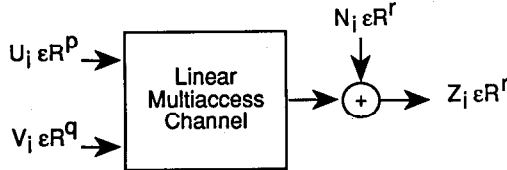Fig. 1.   Gaussian multiaccess channel with ISI.

Fig. 2.   Gaussian linear vector multiaccess channel.
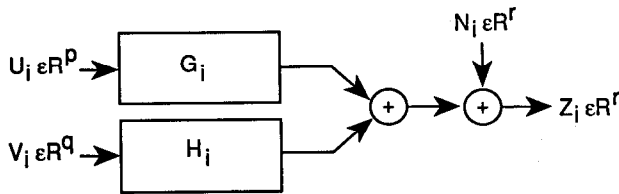

Fig. 3.   Equivalent model for Gaussian linear vector multiaccess channel.

channel methods of [12] and [15] can be employed here. This approach enables an orthogonal decomposition of the channel using the discrete Fourier transform (DFT) that is independent of the ISI coefficients. In this paper, we employ these ideas and the limiting expression for the capacity region of multiaccess channels with memory in [11] to obtain the capacity regions of the Gaussian multiaccess channels with ISI.

What makes the single-user channel with ISI interesting is not just its capacity formula, but also the well-known geometrical water-filling argument used to obtain the optimal input power spectral density (PSD). We extend the single-user water-filling argument to the two-user case when the inputs and the output are scalars. We derive a geometrical method to obtain the optimal input PSD's. It turns out that this geometrical argument can be explained via two main ideas: the equivalent channel idea and the successive decoding idea (decode one user's information while treating the other user's information as noise first and then decode the remaining user's information). The equivalent channel idea bears some resemblance to the single-user water-filling argument in the sense that it obtains graphically the optimal input power distribution over the frequency domain. It can be applied directly to the single-user channel to obtain the optimal input PSD. Roughly speaking, in the two-user case, the equivalent channel idea determines graphically the optimal distribution of the *total* power over the frequency domain, while the successive decoding idea determines, again graphically, the optimal split of the total power among the users for each frequency.

In Section II, the main theorem giving the capacity region of a Gaussian vector multiaccess channel with ISI is presented. Then, we specialize the result to the scalar case (inputs and output are scalars) in Section III where we come up with

a generalization of the water-filling argument to the two-user case. Examples of some simple channels are given to demonstrate how the optimal power spectral densities are obtained by the equivalent channel and successive decoding ideas. As we will see, the generalization of the water-filling argument to the two-user case is not straightforward except when the channels seen by both users are identical.

II. CAPACITY REGION OF GAUSSIAN MAC WITH ISI

In this section, we consider a general multiaccess linear channel with additive Gaussian noise shown in Fig. 2. It is easy to see, by linearity, that the models in Figs. 2 and 3 are equivalent and they include, as special cases, the scalar Gaussian multiaccess channel (MAC) with ISI in Fig. 1. From now on, we shall concentrate on the equivalent model in Fig. 3:

$$Z_i = \sum_{j=0}^{n} G_j I_{i-j} + H_j V_{i-j} + N_i, \tag{1}$$

where Z_i is the output of the channel in $\mathbb{R}^r$, U_i and V_i are symbols sent by user 1 and user 2 in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, and N_i is a zero-mean stationary Gaussian noise vector process with autocorrelation function $R_{i-j} \triangleq E(N_i N_j^T)$. The power constraint requires each codeword of the kth user, $(c_{k0}, \cdots, c_{k(N-1)})$, of each (N, M_1, M_2, ϵ) code (see, e.g., [5] for definition) to satisfy

$$\frac{1}{N} \sum_{i=0}^{N-1} \|c_{ki}\|^2 \leq W_k, \tag{2}$$

where $\| \cdot \|$ denotes the Euclidean norm on $\mathbb{R}^p$ (resp. $\mathbb{R}^q$). We assume that both channels have finite-length impulse responses with length less than or equal to n, and the noise process is an m-dependent stationary process (i.e., N_i and N_j are independent for all $|i - j| > m$.) These are crucial assumptions in the proof of our result. However, the causality assumption is introduced for convenience and ease of notation only. The same result follows if the channel is noncausal, provided that the impulse responses have finite lengths.

Since this channel is a special case of a general (not necessarily linear) multiaccess channel with finite memory, we can make use of a result in [11] which gives the capacity region as a limit of the capacity regions, C_N, of a series of N-block channels. Applying that result here and denoting by C the capacity region of the Gaussian linear multiaccess channel, we have

$$C = \text{Closure}\left(\liminf_{N \to \infty} C_N\right), \tag{3}$$

where

$$C_N = \bigcup_{\substack{U_0^{N-1}, V_0^{N-1} \\ \sum_{i=0}^{N-1} \text{tr}(E(U_i U_i^T)) \leq NW_1 \\ \sum_{i=0}^{N-1} \text{tr}(E(V_i V_i^T)) \leq NW_2}} \left\{ (R_1, R_2) : \begin{array}{l} 0 \leq R_1 \leq \frac{1}{N} I(U_0^{N-1}; Y_0^{N-1} | V_0^{N-1}) \\ 0 \leq R_2 \leq \frac{1}{N} I(V_0^{N-1}; Y_0^{N-1} | U_0^{N-1}) \\ R_1 + R_2 \leq \frac{1}{N} I(U_0^{N-1}, V_0^{N-1}; Y_0^{N-1}) \end{array} \right\}. \tag{4}$$

In (4), U_0^{N-1}, V_0^{N-1} and Y_0^{N-1} are the inputs and the output of the N-block multiaccess channels. The inputs and the output of the N-block channel are the N-vectors of the inputs and the output of the original channel, respectively. The conditional probability of the outputs given the inputs in the N-block channel is induced by the original channel except for the first t output symbols with t being the maximum of n, the length of the impulse response, and m, the length of the noise auto-correlation function. The conditional output distribution is

$$P_{Y_0^{N-1}|U_0^{N-1},V_0^{N-1}}\left(y_0^{N-1}|u_0^{N-1},v_0^{N-1}\right)$$
$$= P\left(y_0^{t-1}|u_0^{N-1},v_0^{N-1}\right)$$
$$\cdot P_{Z_t^{N-1}|U_0^{N-1},V_0^{N-1}}\left(y_t^{N-1}|u_0^{N-1},v_0^{N-1}\right). \quad (5)$$

As pointed out in [11], important aspect of this result is that the conditional probability of the first t output symbols given the inputs can be arbitrarily assigned. This flexibility allows us to define P appropriately so that the N-block channel becomes

$$Y_0^{N-1} = G_0^{N-1} \otimes U_0^{N-1} + H_0^{N-1}$$
$$\oplus V_0^{N-1} + M_0^{N-1}, \quad (6)$$

where $\otimes$ denotes the circular convolution and M_0^{N-1} are Gaussian with mean $\mathbf{0}$ and $E(M_i M_j) = R(i-j)_N$ where $(\cdot)_N$ denotes the modulo N operation. (A similar formulation for the circular N-block channel is used in obtaining the single-user Gaussian channel with ISI [15].) Then, the discrete Fourier transform can be used to decompose the N-block channels into independent channels whose capacities can be found easily. Using these ideas, we find the capacity region of a two-user Gaussian multiaccess channel in the following theorem.

Theorem 1: The capacity region of the two-user Gaussian multiaccess channel is

$$C = \bigcup_{\substack{\Sigma_1(w) \in \mathbb{R}^{p \times p}, \Sigma_2(w) \in \mathbb{R}^{q \times q} \\ \Sigma_i(w) \geq 0 \quad \forall\, w \in [0,\pi] \\ \frac{1}{\pi}\int_0^\pi \mathrm{tr}\Sigma_i(w)\, dw \leq W_i, i=1,2}}$$
$$\cdot \left\{ (R_1, R_2): \begin{array}{l} 0 \leq R_1 \leq F(\Sigma_1, \mathbf{0}) \\ 0 \leq R_2 \leq F(\mathbf{0}, \Sigma_2) \\ R_1 + R_2 \leq F(\Sigma_1, \Sigma_2) \end{array} \right\}, \quad (7)$$

where

$$F(A, B) = \frac{1}{2\pi}\int_0^\pi \log\det$$
$$[I + G(w)A(w)G^*(w)N^{-1}(w)$$
$$+ H(w)B(w)H^*(w)N^{-1}(w)]\, dw, \quad (8)$$

$A^*(w)$ denotes the conjugate transpose of $A(w)$, and $G(w)$, $H(w)$, and $N(w)$ are the Fourier transforms of G_i, H_i, and R_i, respectively.

Proof: Appendix A. $\qquad\qquad\qquad\square$

Remark 1: In the single-user case, it is easy to check that the above capacity region degenerates in the multivariate (or vector) case to that obtained in [16] and in the unvariate (or scalar) case to the well known water-filling result (e.g., [8]).

Remark 2: Theorem 1 can be easily extended to include compound and K-user channels. Since the limiting expression in (3) and (4) can accommodate compound channels [11], it is straightforward to generalize the result to the case where there are several linear systems and the transmitters do not know which one will be used. In such case, the capacity region of the N-block channel, C_N, will be the union of intersections of pentagons instead of the union of pentagons as in (4). The intersection will be taken over all possible linear systems. Using a similar characterization for K-user capacity regions as in [17], one can easily generalize the result to the K-user case.

In order to get a better understanding of Theorem 1, we consider two special cases in the following corollaries which correspond to cases where the inputs and the outputs are scalars.

Corollary 1: If both the inputs and the output are scalars (i.e., $p = q = r = 1$), then the capacity region of the two-user Gaussian multiaccess channel is shown in (9) at the page where $T_1(w) = |G(w)|^2/N(w)$ and $T_2(w) = |H(w)|^2/N(w)$.

Corollary 2: Suppose that both the inputs and the output are scalars (i.e., $p = q = r = 1$), and the channels for both users are identical (i.e., $G(w) = H(w)$ for all w). Let $T(w) \triangleq T_1(w) = T_2(w)$. Then, the capacity region of the two-user Gaussian multiaccess channel becomes the pentagon

$$C = \left\{ (R_1, R_2): \begin{array}{l} 0 \leq R_1 \leq \frac{1}{2\pi}\int_0^\pi \log[1 + S_1(w)T(w)]\, dw \\ 0 \leq R_2 \leq \frac{1}{2\pi}\int_0^\pi \log[1 + S_2(w)T(w)]\, dw \\ R_1 + R_2 \leq \frac{1}{2\pi}\int_0^\pi \log[1 + S_{12}(w)T(w)]\, dw \end{array} \right\}, \quad (10)$$

where $S_1(w) = \left[c_1 - T^{-1}(w)\right]^+$, $S_2(w) = \left[c_2 - T^{-1}(w)\right]^+$, $S_{12}(w) = \left[c_{12} - T^{-1}(w)\right]^+$, and c_1, c_2, c_{12} are chosen so that

$$\frac{1}{\pi}\int_0^\pi S_k(w)\, dw = W_k, \qquad k = 1, 2, \quad (11)$$

$$\frac{1}{\pi}\int_0^\pi S_{12}(w)\, dw = W_1 + W_2. \quad (12)$$

$$C = \bigcup_{\substack{S_1(w), S_2(w) \in \mathbb{R} \\ S_i(w) \geq 0 \quad \forall\, w \in [0,\pi] \\ \frac{1}{\pi}\int_0^\pi S_i(w)\, dw \leq W_i, i=1,2}} \left\{ (R_1, R_2): \begin{array}{l} 0 \leq R_1 \leq \frac{1}{2\pi}\int_0^\pi \log[1 + S_1(w)T_1(w)]\, dw \\ 0 \leq R_2 \leq \frac{1}{2\pi}\int_0^\pi \log[1 + S_2(w)T_2(w)]\, dw \\ R_1 + R_2 \leq \frac{1}{2\pi}\int_0^\pi \log[1 + S_1(w)T_1(w) + S_2(w)T_2(w)]\, dw \end{array} \right\}, \quad (9)$$

The proofs of these corollaries are in Appendix B. A novel two-user water-filling argument to obtain the optimal input PSD's is presented in Section III. Corollary 2 corresponds to the simplest scalar channel where the users' channels are identical and the result agrees with those reported in [13] and [14]. In such case, the capacity region is a pentagon and the corresponding optimal input PSD's can be obtained through the classical single-user water-filling argument.

III. TWO-USER WATER-FILLING

In this section, we consider the Gaussian multiaccess channel with ISI in the scalar case (i.e., $p = q = r = 1$) and give a geometrical characterization of the optimal input power spectral densities (PSD's). It is important to note that unlike the single-user channel, the two-user capacity region is a two-dimensional region and different PSD pairs are needed to achieve different boundary points of that region. Hence, instead of finding the optimal PSD that achieves the capacity as in the single-user case, the task in the two-user case is to find, for each boundary point of the capacity region, the corresponding optimal PSD pair.

We start by reviewing the classical single-user water-filling argument in Section III-A. Then, a special case of the two-user channel in which the channels seen by the users are identical (i.e., $G(w) = H(w)$) is studied in Section III-B, where we show how the single-user water-filling argument can be applied repeatedly to obtain the optimal PSD pairs. Unfortunately, this is no longer true in the general case when $G(w) \neq H(w)$. We discuss the difficulties in extending the water-filling argument to the two-user case in Section III-C. After briefly outlining the new ideas needed in the two-user water-filling argument, we present in detail the procedure and the interpretation for obtaining the PSD pairs that achieve the total capacity and other boundary points in Section III-C-1 and III-C-2, respectively. Finally, some examples are given in Section III-D.

A. Classical Water-Filling

The classical water-filling argument is a geometrical method to obtain the optimal input PSD that achieves the capacity. The optimal PSD is the unique solution to the following equations:

$$S(w) = \left[c - T^{-1}(w)\right]^{+}, \qquad (13)$$

$$W = \frac{1}{\pi} \int_{0}^{\pi} S(w) \, dw, \qquad (14)$$

where W is the total power and $T(w) = |H(w)|^{2}/N(w)$, the magnitude square of the channel transfer function over the noise power spectral density. The geometrical interpretation is illustrated in the water-filling diagram as shown in Fig. 4. We can imagine that $T^{-1}(w)$ is the bottom of a container and a fixed amount, W, of water (power) is poured into the container. The water (power) will distribute itself to maintain a water level, c. Then, the shaded area indicated in Fig. 4 gives the optimal PSD.

The water-filling argument gives a nice interpretation for obtaining the optimal PSD. As we will see in the sequel, in order to extend the method to the general multiuser channel,
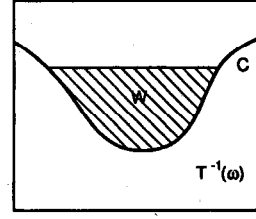

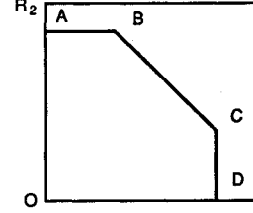Fig. 4. Classical water-filling argument for single-user channel.

Fig. 5. Capacity region when $G(w) = H(w)$, $\forall w \in [0, \pi]$.

we need to modify the water-filling diagram to an equivalent, but less appealing form. However, in the case where both channels seen by the users are identical (i.e., $G(w) = H(w)$), the problem is greatly simplified. We show in the next section how the classical water-filling method is used to obtain the optimal PSD pairs in that particular case.

B. Two-User Channel $(G(w) = H(w))$

We now consider the two-user channel with $G(w) = H(w)$. It turns out that this special case avoids most of the difficulties we will encounter in the general multiaccess channel, and only the classical water-filling argument is needed to obtain the optimal PSD pairs.

In Corollary 2, we found that the capacity region of this channel is a pentagon as shown in Fig. 5 and the optimal PSD pair achieving point C at the corner of the capacity region is the solution of the following equations:

$$S_1(w) + S_2(w) = \left[c_{12} - T^{-1}(w)\right]^{+}, \qquad (15)$$

$$\frac{1}{\pi} \int_{0}^{\pi} S_1(w) + S_2(w) \, dw = W_1 + W_2, \qquad (16)$$

$$S_1(w) = \left[c_1 - T^{-1}(w)\right]^{+}, \qquad (17)$$

$$\frac{1}{\pi} \int_{0}^{\pi} S_1(w) \, dw = W_1. \qquad (18)$$

The geometrical interpretation is similar to that in the single-user case, and is illustrated in Fig. 6. The water-filling argument is applied twice: first with the sum of the users' power to obtain the sum of the PSD's, and second with the power of user 1 to obtain the PSD of user 1.

By symmetry, the optimal PSD pair achieving point B can be obtained using the same procedure with the roles of the users interchanged. The points on line BC can be achieved by time-sharing between the coding strategies for points B and C; however, if stationary inputs are desired, those points can also be achieved by PSD pairs equal to convex combinations of the optimal PSD pairs for points B and C. In fact, any nonnegative PSD pair satisfying (15) and the power constraints
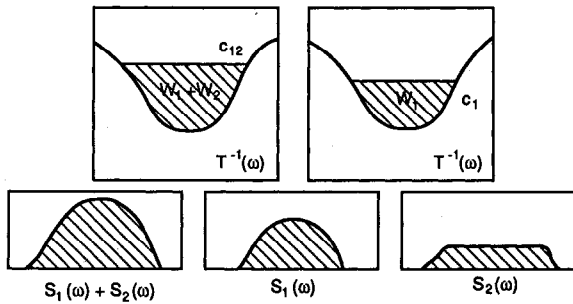
Fig. 6. Optimal PSD's for point C in the capacity region when $G(w) = H(w), \forall w \in [0, \pi]$.

achieves a point on line BC. In particular, frequency-division multiaccess (FDMA) with appropriately selected bands always achieves the total capacity. As we will see in Section III-C, this remains true even when the channels seen by the users are not identical.

The relative simplicity for obtaining the optimal PSD in this special case is mainly due to the fact that the capacity region is a pentagon. Once we determine the optimal PSD pairs for the corners, the optimal PSD pairs for all points along the side of the pentagon are just convex combinations of the PSD pairs for the corners. Since each corner represents a rate pair achieving both the total capacity and the single-user capacity of one user, the PSD of that user and the sum of the PSD's can be obtained by applying the classical water-filling argument twice. Notice that in the general case, when other points are under consideration, no user achieves its single-user capacity and hence, none of their optimal PSD's can be obtained via the classical water-filing argument.

This procedure can be generalized straightforwardly to the K-user channel in which all users see the same channel. In that case, the capacity region is a polytope. For example, in the three-user case, $S_1(w) + S_2(w) + S_3(w)$, $S_1(w) + S_2(w)$, and $S_1(w)$ can be obtained via three water-filling diagrams with total power $W_1 + W_2 + W_3$, $W_1 + W_2$, and W_1, respectively. Then, the resulting PSD triple achieves a vertex of the capacity region. With different combinations of users, we can obtain six different PSD triples, each achieving a vertex in the capacity region for the three-user channel. Similarly, convex combinations of the PSD triples for three vertices achieve all points on the plane connecting those vertices.

C. Two-User Channel $(G(w) \neq H(w))$

Now, we proceed to the general two-user case where the channels seen by the users are not necessarily identical.

The capacity region of the general two-user channel is a convex region, but not necessarily a pentagon. One way to describe the capacity region (a convex set in the positive quadrant) is to bound it by straight lines (hyperplanes), $\alpha R_1 + (1 - \alpha)R_2 = C(\alpha)$, for $\alpha \in [0, 1]$. The capacity region can then be written as

$$C = \{(R_1, R_2) \in \mathbb{R}_+^2 : \alpha R_1 + (1 - \alpha)R_2$$
$$\leq C(\alpha), \forall \alpha \in [0, 1]\} \qquad (19)$$

The points achieving $\alpha R_1 + (1 - \alpha)R_2 = C(\alpha)$ with equality

are the boundary points of the capacity region. The objective of the two-user water-filling argument is to find, for each $\alpha \in [0, 1]$, the optimal PSD pair maximizing $C(\alpha)$ and thus the maximum achievable $\alpha R_1 + (1 - \alpha)R_2$. It is also useful to view α as a parameter determining the relative priorities of the users. The closer α is to 1, the higher the priority given to user 1.

In contrast to the special case considered in the previous section, this general channel poses two main difficulties in obtaining a generalized water-filling argument.

1) The capacity region is no longer a pentagon and the optimal PSD pair for every boundary point, not just the corners, have to be found using a generalized water-filling method. Unlike those corners in the special case where the highest priority is given to one user to achieve its single-user capacity, a typical boundary point in the general case maximizes a weighted average of the users' rates. In that case, no user can have an optimal PSD equal to its single-user optimal PSD, and neither of the PSD's can be obtained via the classical single-user water-filling argument.

2) Since the two channels seen by the users are different, we have two curves, $T_1^{-1}(w) \triangleq R(w)/|G(w)|^2$ and $T_2^{-1}(w) \triangleq R(w)/|H(w)|^2$, and two water-filling diagrams. In general, the PSD's of the users cannot be found separately in two water-filling diagrams since their signals interfere with each other. At first glance, it is not clear how these two curves should be combined in one water-filling diagram with a single water level in the multiuser case.

Despite these differences, the study of the special case in the previous section does help to develop a general approach. As in the special case, we find the sum of the PSD's first, and then split the sum into two optimal individual PSD's. In the process, we need two ideas: the equivalent channel idea that leads to the proper combination of the two curves, $T_1^{-1}(w)$ and $T_2^{-1}(w)$, to find the sum of the PSD's, and the successive decoding idea that determines the optimal split of the PSD sum into individual PSD's. We explore the equivalent channel idea in Section III-C-1 where it is used to obtain the optimal PSD pair for the boundary point achieving the total capacity (i.e., $\alpha = 1/2$). It turns out that when the two users are given the same priority, the split of the PSD sum is trivial and only the equivalent channel idea is needed. In particular, the PSD's of the users should never overlap and FDMA with appropriately selected bands always achieves the total capacity for any number of users. Then, in Section III-C-2, the successive decoding idea is introduced and combined with the equivalent channel idea to find the optimal PSD pairs for all the boundary points of the capacity region.

1) PSD Pair for Achieving Total Capacity (Equivalent Channel Idea): In this section, we consider the optimal PSD pair for the boundary point achieving the total capacity. In other words, we find the optimal PSD pair maximizing the rate sum, $R_1 + R_2$, over the capacity region. It turns out that the total capacity can always be achieved by FDMA, even for channels with more than two users.
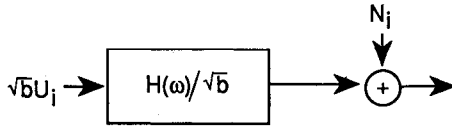
Fig. 7. Equivalent channel idea.

Fig. 8. Water-filling diagram for the equivalent channel.

Fig. 9. Equivalent channel idea applied to the two-user channel.

We shall motivate the equivalent channel idea by considering a two-user channel where $G(w)$ and $H(w)$ are two bandpass filters with nonoverlapping passbands. In that case, the multiaccess channel decouples into two single-user channels and the optimal input PSD's can be determined separately via two single-user water-filling diagrams. Depending on the gains of the filters and the input power constraints, the water levels may not be the same. This difference in the water levels causes no problem in this special case because the PSDs optimal for the single-user channels will never overlap. However, in the general case, the PSD's have to be determined in one water-filling diagram with a single water level and this difference in the water levels prevents us from combining the single-user water-filling diagrams directly. The equivalent channel idea introduces channels equivalent to the original channels, but with scaled versions of the respective water filling diagrams. Then, by proper selection of the equivalent channels, the two water-filling diagrams can be properly scaled, and then combined into one diagram with a single water level where the sum of the PSD's can be determined.

We now proceed to define the equivalent channels. For any linear additive Gaussian noise channel with transfer function, $H(w)$, and power constraint W, we define a family of equivalent channels, parameterized by $b > 0$, each of which has transfer function $H(w)/\sqrt{b}$ and power constraint bW as shown in Fig. 7. Then, the optimal PSD for the equivalent channel is the unique solution to the following equations:

$$\hat{S}(w) = \left[a - bT^{-1}(w)\right]^{+}, \qquad (20)$$

$$\frac{1}{\pi}\int_{0}^{\pi}\hat{S}(w)\,dw = bW. \qquad (21)$$

The capacities of the equivalent channel and the original channel are naturally the same while the optimal PSD of the equivalent channel is a scaled version of the optimal PSD of the original channel.

In order to extend the water-filling argument to the two-user case, we need to scale two water-filling diagrams such that they can be combined appropriately to maintain a single water level for all users. Hence, using the equivalent channel idea, we define a modified version of the water-filling diagram, as shown in Fig. 8, in which the water level is fixed to 1, and the bottom of the container is scaled by adjusting the parameter b such that the amount of water (power) is equal to bW, the power constraint of the equivalent channel. As in the classical water-filling argument, the height of the shaded region gives the PSD of the equivalent channel or the PSD of the original channel scaled by b.

The equivalent channel idea is applied to the two-user channel in Fig. 9 where the equivalent two-user channels are parameterized by b_1 and b_2. If we find the optimal PSD pair
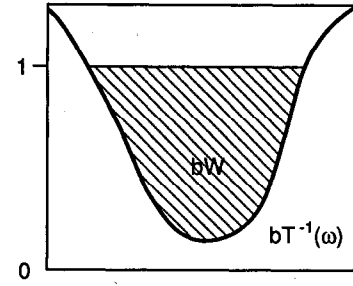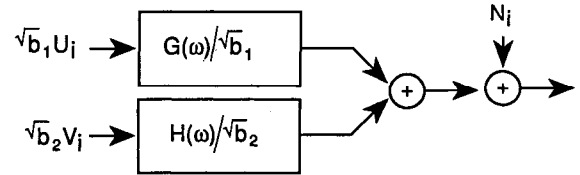
for one of these equivalent channels, we can find the optimal PSD pair for the original channel by scaling.

The optimal PSD pair, (S_1, S_2), achieving the total capacity is the solution to the following equations:

$$S_k(w) = \frac{\hat{S}_k(w)}{b_k}, \qquad k = 1, 2, \qquad (22)$$

$$\hat{S}_1(w) + \hat{S}_2(w) = \left[1 - \min\{b_1 T_1^{-1}(w), b_2 T_2^{-1}(w)\}\right]^{+}, \quad (23)$$

$$\frac{1}{\pi}\int_{0}^{\pi}\hat{S}_1(w) + \hat{S}_2(w)\,dw = b_1 W_1 + b_2 W_2, \qquad (24)$$

$$\hat{S}_1(w) =$$
$$\begin{cases} \left[1 - b_1 T_1^{-1}(w)\right]^{+}, & \text{if } b_1 T_1^{-1}(w) \le b_2 T_2^{-1}(w), \\ 0, & \text{otherwise,} \end{cases} \quad (25)$$

$$\frac{1}{\pi}\int_{0}^{\pi}\hat{S}_1(w)\,dw = b_1 W_1. \qquad (26)$$

The geometrical procedure is shown in Fig. 10. We first fix the water level to be 1 and plot the two curves, $b_1 T_1^{-1}(w)$ and $b_2 T_2^{-1}(w)$, on the same diagram. Treating the minimum of the two curves as the bottom of the container, we adjust the parameters b_1 and b_2 such that a) the total amount of water is equal to $b_1 W_1 + b_2 W_2$ and b) the amount of water in the region where $b_1 T_1^{-1}(w) \le b_2 T_2^{-1}(w)$ is equal to $b_1 W_1$. Then, the two shaded regions in Fig. 10 give the PSD's for the two users.

Using the equivalent channel idea, we can see that the parameters b_1 and b_2 select two equivalent channels such that their water-filling diagrams can be combined into one diagram with a common water level equal to 1. The selection of these parameters depends on the shapes of $T_1^{-1}(w)$ and $T_2^{-1}(w)$, and the power constraints, W_1 and W_2, in a complicated fashion. However, we can obtain b_1 and b_2 graphically as described above (or numerically).
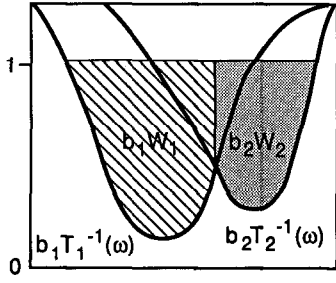
Fig. 10. Power spectral densities achieving total capacity.

Regarding this procedure, we want to justify a) why do we treat the minimum of the two curves as the bottom of the container in determining the sum of the PSD's? b) why do we split the sum of the PSD's into two individual PSDs as in (25)? The rationale behind these can be seen by examining the equivalent channel model in Fig. 9 again. Using the same idea as in the single-user ISI channel, we consider a two-user ISI channel as a series of parallel independent channels, each of which is a two-user channel, $Y = (|H(w_i)|/\sqrt{b_1})X_1 + (|G(w_i)|/\sqrt{b_2})X_2 + N(w_i)$. Let us imagine that there is a virtual user who has total power equal to $b_1W_1 + b_2W_2$. We assume that this virtual user can transmit over all the parallel channels and in each parallel two-user channel, he can transmit over either or both subchannels simultaneously. Then, the virtual user has to determine the distribution of the total power over these parallel channels as well as the split of the power allocated for each two-user channel between its two subchannels. For each parallel channel, since the output of the two subchannels are superimposed before being corrupted by noise, the optimal split between the two subchannels is to send all the power over the subchannel with a larger gain as this strategy maximizes the power of the superimposed signal $(|H(w_i)|/\sqrt{b_1})X_1 + (|G(w_i)|/\sqrt{b_2})X_2$ and hence the signal-to-noise ratio. Therefore, each two-user channel appears to be a single-user channel with gain equal to $\max\{(|H(w_i)|/\sqrt{b_1}), (|G(w_i)|/\sqrt{b_2})\}$ and the distribution of power over the parallel channels can be obtained using the single-user water-filling argument with $\min\{b_1T_1^{-1}(w), b_2T_2^{-1}(w)\}$ as the bottom of the container. Moreover, the water (power) over the region where $b_1T_1^{-1}(w) \leq b_2T_2^{-1}(w)$ is the PSD transmitted over the channel seen by user 1 and the remaining part is the PSD transmitted over the channel seen by user 2. Hence, as with the multiaccess channel without ISI, FDMA maximizes total capacity even in the presence of ISI.

Although we have only discussed the two-user case, the argument holds for arbitrary number of users. The following theorem shows that FDMA achieves the total capacity for any K-user Gaussian multiaccess channel with ISI and its proof

illustrates the recursive procedure to obtain the optimal PSD's (or the frequency bands) for the users.

Theorem 2: For any K-user m-dependent Gaussian multiaccess scalar channel with finite intersymbol interference and power constraints $W_1, \cdots, W_K$:

$$Z_i = \sum_{k=1}^{K} \sum_{j=0}^{n} H_{k,j} X_{k,i-j} + N_i, \qquad (27)$$

the total capacity can be achieved by FDMA with optimal input PSD K-tuple, $(S_1(w), \cdots, S_K(w))$, where

$$S_k(w) = \frac{\hat{S}_k(w)}{b_k}, \qquad (28)$$

(28) (shown at the bottom of the page), and $b_1, \cdots, b_K$ are chosen such that

$$\frac{1}{\pi} \int_0^\pi \hat{S}_k(w)\, dw = b_k W_k, \qquad (29)$$

for $k = 1, \cdots, K$.

Proof: Generalizing the result in Theorem 1 and applying it to the scalar case where both the inputs and the output are scalars, we find that the total capacity, C, is

$$C = \max_{\substack{S_k(w) \in \mathbb{R}_+\ \forall w \in [0,\pi] \\ \frac{1}{\pi}\int_0^\pi S_k(w)\,dw \leq W_k,\ k=1,\cdots,K}} \frac{1}{2\pi}$$
$$\cdot \int_0^\pi \log\left[1 + \sum_{k=1}^{K} S_k(w)T_k(w)\right] dw, \qquad (31)$$

where $T_k(w) = |H_k(w)|^2/N(w)$, the magnitude squared of the kth user channel transfer function over the noise spectral density. Using the equivalent channel idea, we have that for any $(b_1, \cdots, b_K) \in \mathbb{R}^K_+$,

$$C = \max_{\substack{\hat{S}_k(w) \in \mathbb{R}_+\ \forall w \in [0,\pi] \\ \frac{1}{\pi}\int_0^\pi \hat{S}_k(w)\,dw \leq b_k W_k,\ k=1,\cdots,K}} \frac{1}{2\pi}$$
$$\cdot \int_0^\pi \log\left[1 + \sum_{k=1}^{K} \hat{S}_k(w)\frac{T_k(w)}{b_k}\right] dw \qquad (32)$$

$$\leq \max_{\substack{\hat{S}_k(w) \in \mathbb{R}_+\ \forall w \in [0,\pi] \\ \frac{1}{\pi}\int_0^\pi \hat{S}_k(w)\,dw \leq b_k W_k,\ k=1,\cdots,K}} \frac{1}{2\pi}$$
$$\cdot \int_0^\pi \log\left[1 + \sum_{k=1}^{K} \hat{S}_k(w) \max_{j=1,\cdots,K} \frac{T_j(w)}{b_j}\right] dw \qquad (33)$$

$$\hat{S}_k(w) = \begin{cases} \left[1 - b_k T_k^{-1}(w)\right]^+, & \text{if } b_k T_k^{-1}(w) \leq b_l T_l^{-1}(w), \quad \text{for all } l \neq k, \\ 0, & \text{otherwise}. \end{cases} \qquad (30)$$

$$= \max_{\substack{\hat{S}(w) \in \mathbb{R}_+ \ \forall w \in [0,\pi] \\ \frac{1}{\pi} \int_0^\pi \hat{S}(w)\,dw \le \sum_{k=1}^K b_k W_k}} \frac{1}{2\pi}$$
$$\cdot \int_0^\pi \log\left[1 + \hat{S}(w) \max_{j=1,\cdots,K} \frac{T_j(w)}{b_j}\right] dw, \qquad (34)$$

where the optimal $\hat{S}(w) = \sum_{k=1}^K \hat{S}_k(w)$ can be obtained via the single-user water-filling argument:

$$\sum_{k=1}^K \hat{S}_k(w) = \left[1 - \min_{k=1,\cdots,K} b_k T_k^{-1}(w)\right]^+, \qquad (35)$$

$$\frac{1}{\pi} \int_0^\pi \hat{S}(w)\,dw = \sum_{k=1}^K b_k W_k. \qquad (36)$$

Now, we have a set of upper bounds on the total capacity parameterized by $(b_1, \cdots, b_K)$. If there exists $(b_1, \cdots, b_K)$ satisfying (29) and (30), it is easy to check that $(\hat{S}_1(w), \cdots, \hat{S}_K(w))$ obtained in (29) satisfies (35) and (33) with equality. Hence, it achieves the upper bound and is the optimal PSD K-tuple.

In order to complete the proof, we need to show that for arbitrary power constraints, $W_1, \cdots, W_K$, there exists $b_1, \cdots, b_K$ satisfying (29) and (30) for $k = 1, \cdots, K$. Note that (29) is equivalent to

$$\sum_{l=1}^k \hat{S}_l(w) =$$
$$\begin{cases} \left[1 - \min_{l=1,\cdots,k} b_l T_l^{-1}(w)\right]^+, & \text{if } \min_{l=1,\cdots,k} b_l T_l^{-1}(w) \\ & \le \min_{l=k+1,\cdots,K} b_l T_l^{-1}(w), \\ 0, & \text{otherwise}, \end{cases}$$
$$\hspace{11cm} (37)$$

for $k = 1, \cdots, K$. (We have adopted the convention that minimizing over an empty set is equal to infinity.) Now, we show the result by induction on the following proposition.

P_L : For any $b_{L+1}, \cdots, b_K$, there exists $b_1, \cdots, b_L$ satisfying (37) and (30) for $k = 1, \cdots, L$. Moreover, each b_i for $i = 1, \cdots, L$ is increasing in b_j for $j = L+1, \cdots, K$.

The assertion is true for $L = 1$. Let us define

$$\Delta_1(b_1, \cdots, b_K) = \frac{1}{\pi} \int_0^\pi \hat{S}_1(w)\,dw - b_1 W_1, \qquad (38)$$

where $\hat{S}_1(w)$ is the solution to (37) for $k = 1$. Since $\Delta_1(b_1, \cdots, b_K) \ge 0$ when $b_1 = 0$ and decreases to $-\infty$ as b_1 increases to ∞, there must exist a b_1 satisfying (37) and (30) for $k = 1$. Moreover, as b_j for $j > 1$ increases, it follows from (37) that $\hat{S}_1(w)$, and thus $\Delta_1(b_1, \cdots, b_K)$, increases. Since $\Delta_1(b_1, \cdots, b_K)$ is a decreasing function in b_1, b_1 satisfying (30) for $k = 1$ must increase. This shows the second part of the assertion.

Let us assume that P_L is true. Then, for any $b_{L+1}, \cdots, b_K$, there exists $b_1, \cdots, b_L$ satisfying (37) and (30) for $k =$

$1, \cdots, L$. Using these $b_1, \cdots, b_K$, we find $\hat{S}_k(w)$ from (37) and define

$$\Delta_k(b_{L+1}, \cdots, b_K) = \frac{1}{\pi} \int_0^\pi \hat{S}_1(w)\,dw - b_k W_k, \qquad (39)$$

for $k = 1, \cdots, L+1$. Note that in this definition, for any $b_{L+1}, \cdots, b_K$, $\Delta_k(b_{L+1}, \cdots, b_K) = 0$ for $k = 1, \cdots, L$ because $b_1, \cdots, b_L$ are chosen to satisfy (30) for $k = 1, \cdots, L$. For b_{L+1} small enough, it is easy to see that $\Delta_{L+1}(b_{L+1}, \cdots, b_K) > 0$. Let $b'_{L+1} > b_{L+1}$, and b'_k and $\hat{S}'_k(w)$ be the corresponding solutions to (37) and (30) for $k = 1, \cdots, L$, Then, we have

$$\Delta_{L+1}(b'_{L+1}, b_{L+2}, \cdots, b_K)$$
$$= \sum_{l=1}^{L+1} \Delta_l(b'_{L+1}, b_{L+2}, \cdots, b_K) \qquad (40)$$

$$= \frac{1}{\pi} \int_0^\pi \sum_{l=1}^{L+1} \hat{S}'_l(w)\,dw - \sum_{l=1}^{L+1} b'_l W_l \qquad (41)$$

$$\le \frac{1}{\pi} \int_0^\pi \sum_{l=1}^{L+1} \hat{S}_l(w)\,dw - \sum_{l=1}^{L+1} b_l W_l \qquad (42)$$

$$= \sum_{l=1}^{L+1} \Delta_k(b_{L+1}, b_{L+2}, \cdots, b_K) \qquad (43)$$

$$= \Delta_{L+1}(b_{L+1}, b_{L+2}, \cdots, b_K), \qquad (44)$$

where (40) and (44) follow from $\Delta_k(b'_{L+1}, b_{L+2}, \cdots, b_K) = 0$ and $\Delta_k(b_{L+1}, \cdots, b_K) = 0$, respectively, for $k = 1, \cdots, L$. Equation (42) follows from the facts that $b'_k \ge b_k$ for $k = 1, \cdots, L$, (assured by the second part of P_L) and its implication in (37), $\sum_{l=1}^{L+1} \hat{S}'_l(w) \le \sum_{l=1}^{L+1} \hat{S}_l(w)$. Therefore, $\Delta_{L+1}(b_{L+1}, \cdots, b_K)$ is also a decreasing function in b_{L+1} and tends to $-\infty$ as $b_{L+1} \to \infty$. Hence, there must exist b_{L+1} and $b_1, \cdots, b_L$, whose existence is guaranteed by P_L, satisfying (37) and (30) for $k = 1, \cdots, L+1$. Now, we show the second part of P_{L+1}. If b_j for $j = L+2, \cdots, K$ increases to b'_j, it follows from P_L that each b_i for $i = 1, \cdots, L$ increases. Then, we observe from (29) that $\hat{S}_{L+1}(w)$ increases and $\Delta_{L+1}(b_{L+1}, \cdots, b'_j, \cdots, b_K) > 0$. Therefore, b_{L+1} satisfying (30) must increase, and by P_L again, so are $b_1, \cdots, b_L$ and the proof is completed.

By induction, we have shown that the proposition P_L is true for $L = 1, \cdots, K$ and hence, the desired result which is P_K.
□

2) PSD Pairs for all Boundary Points (Successive Decoding Idea): In this section, we derive the two-user water-filling procedure to obtain the optimal PSD pairs achieving the boundary points of the capacity region. Specifically, we seek to find the optimal PSD pair that maximizes a weighted rate sum, $\alpha R_1 + (1 - \alpha)R_2$.

The main differences between the procedures for obtaining the optimal PSD pair for the total capacity and other points are the way the two water-filling diagrams are combined and the
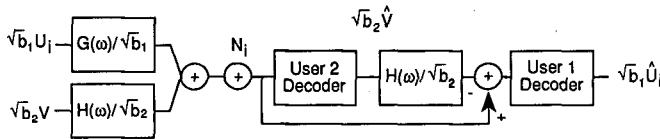
Fig. 11. Successive decoding idea in two-user channel.

way the sum of the PSD's is split into the individual PSD's. When $\alpha \neq 1/2$, the relative priorities given to the users are not the same and it turns out that when combining the two water-filling diagrams, offsets equal to the relative priorities, α and $1 - \alpha$, have to be added. Instead of $b_1 T_1^{-1}(w)$ and $b_2 T_2^{-1}(w)$, we have to draw $b_1 T_1^{-1}(w) + (1 - \alpha)$ and $b_2 T_2^{-1}(w) + \alpha$ on a diagram with a fixed water level equal to 1. Moreover, splitting of the PSD sum into individual PSD's is no longer trivial and the successive decoding idea is needed to explain it.

Let us consider the successive decoding idea first put forward by Cover in [18] to explain the capacity region (Cover–Wyner pentagon) for the memoryless Gaussian multiaccess channel. The idea is illustrated in Fig. 11 for the case when user 1 has a higher priority than user 2. The receiver decodes the signal sent by user 2, the user with lower priority, first, treating the signal component from user 1 as noise. Then, the receiver regenerates the signal from user 2 and subtracts it from the received signal. The difference is then used to decode the signal sent by user 1. Notice that user 2 sees a new channel whose noise is the sum of the original channel noise and the signal sent by user 1. If user 2 communicates at rate less than the capacity of this new channel, the component due to user 2 can be eliminated almost perfectly from the received signal and the presence of user 2 is virtually transparent to user 1. Since the shape of the PSD of user 2 has no effect on the rate of user 1, the PSD for user 2 should be set to maximize the rate of user 2. The optimal PSD of user 2 then follows from the single-user water-filling argument with the bottom of the container equal to

$$\hat{T}_2^{-1}(w) = \frac{N(w) + \hat{S}_1(w)\frac{|G(w)|^2}{b_1}}{\frac{|H(w)|^2}{b_2}} \qquad (45)$$

$$= b_2 T_2^{-1}(w) + \hat{S}_1(w)\frac{b_2 T_1(w)}{b_1 T_2(w)} \qquad (46)$$

because the noise seen by user 2 is the sum of the original channel noise and the signal sent by user 1. Note that the PSD of user 1 cannot be obtained via the classical water-filling argument since its signal will become noise to user 2 and the effect on the rate of user 2 cannot be ignored when one tries to maximize a weighted sum of the users' rates.

Now, we are ready to give our result on the optimal PSD pairs.

Theorem 3: If both the inputs and the output are scalars (i.e., $p = q = r = 1$), the capacity region of the two-user Gaussian multiaccess channel is

$$C = \left\{ (R_1, R_2) \in \mathbb{R}^2_+ : \alpha R_1 + (1 - \alpha)R_2 \leq C(\alpha) \right\}, \quad (47)$$

where (see (48) at the bottom of the page)

$$F(A, B) = \frac{1}{2\pi} \cdot \int_0^\pi \log[1 + A(w)T_1(w) + B(w)T_2(w)] \, dw, \qquad (49)$$

$T_1(w) = |G(w)|^2/N(w)$, $T_2(w) = |H(w)|^2/N(w)$, and $(S_{1,\alpha}(w), S_{2,\alpha}(w))$ is the corresponding optimal PSD pair. For each α,

$$S_{k,\alpha}(w) = \frac{\hat{S}_k(w)}{b_k}, \qquad k = 1, 2, \qquad (50)$$

where $b_1 > 0$, $b_2 > 0$, $\hat{S}_1(w)$, and $\hat{S}_2(w)$ are the solutions to

$$\frac{1}{\pi} \int_0^\pi \hat{S}_k(w) \, dw = b_k W_k, \qquad k = 1, 2, \qquad (51)$$

$$\hat{S}_1(w) + \hat{S}_2(w) = [1 - \min\{b_1 T_1^{-1}(w) + (1 - \alpha), \\ b_2 T_2^{-1}(w) + \alpha\}]^+, \qquad (52)$$

and if $\alpha \in [0, 1/2]$,

$$\hat{S}_1(w) = \left[1 - \left(b_1 T_1^{-1}(w) + (1 - \alpha) \right. \right. \\ \left. \left. + \hat{S}_2(w)\frac{b_1 T_1^{-1}(w)}{b_2 T_2^{-1}(w)}\right)\right]^+, \qquad (53)$$

$$= \left[\alpha - b_1 T_1^{-1}(w) \right. \\ \left. \cdot \left(1 + \frac{[b_1 T_1^{-1}(w) - b_2 T_2^{-1}(w) + (1 - 2\alpha)]^+}{[b_2 T_2^{-1}(w) - b_1 T_1^{-1}(w)]^+}\right)\right]^+, \qquad (54)$$

and if $\alpha \in [1/2, 1]$,

$$\hat{S}_2(w) = \left[1 - \left(b_2 T_2^{-1}(w) + \alpha \right. \right. \\ \left. \left. + \hat{S}_1(w)\frac{b_2 T_2^{-1}(w)}{b_1 T_1^{-1}(w)}\right)\right]^+, \qquad (55)$$

$$C(\alpha) = \begin{cases} \alpha F(S_{1,\alpha}, S_{2,\alpha}) + (1 - 2\alpha)F(0, S_{2,\alpha}), & \text{if } \alpha \in [1, 1/2] \\ (1 - \alpha)F(S_{1,\alpha}, S_{2,\alpha}) + (2\alpha - 1)F(S_{1,\alpha}, 0), & \text{if } \alpha \in [1/2, 1] \end{cases} \qquad (48)$$
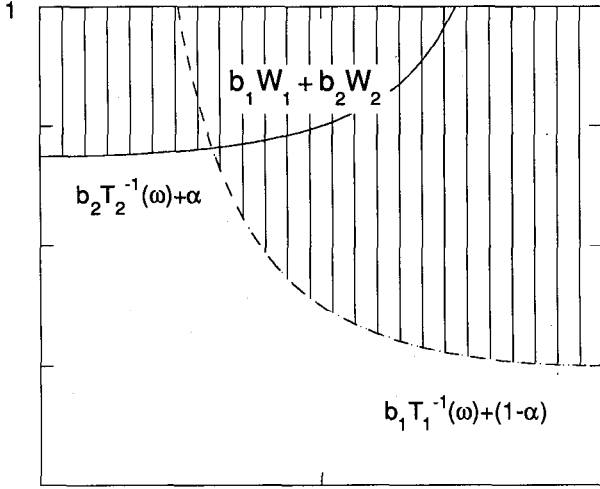
Fig. 12. Sum of the user's optimal power spectral densities in example 4.

Fig. 13. Optimal power spectral density of user 2 in example 4.

$$= \left[1 - \alpha - b_2 T_2^{-1}(w) \right.$$

$$\left. \cdot \left(1 + \frac{\left[b_2 T_2^{-1}(w) - b_1 T_1^{-1}(w) + (2\alpha - 1) \right]^+}{\left[b_1 T_1^{-1}(w) - b_2 T_2^{-1}(w) \right]^+} \right) \right]^+ . \tag{56}$$

Proof: Though tedious, the proof basically involves applying the Kuhn–Tucker theorem to the capacity region in Corollary 1. □

First of all, note that when $\alpha = 1$ or 0, the result reduces to the single-user water-filling argument. For example, let us consider the case when $\alpha = 1$. Since $b_1 T_1^{-1}(w) \leq b_2 T_2^{-1}(w) + 1$ if $b_1 T_1^{-1}(w) < 1$ and the right-hand side of (54) must be 0 if $b_1 T_1^{-1}(w) \geq 1$, (54) becomes

$$\hat{S}_1(w) = \left[1 - b_1 T_1^{-1}(w) \right]^+ . \tag{57}$$

Hence, the result reduces to the single-user water-filling argument.

When $\alpha = 1/2$, the result reduces to that in Section III-C-1 (except that the common water level, which can be arbitrarily fixed, is set to $1/2$ instead of 1.)

For $0 < \alpha < 1/2$ or $1/2 < \alpha < 1$, both the equivalent channel and the successive decoding ideas are needed to obtain the optimal PSD pairs. By symmetry, we will consider only the case when $1/2 < \alpha < 1$ and the corresponding water-filling diagram is illustrated in Figs. 12 and 13. Since the priorities given to the users are not the same, as pointed out before, different offsets, corresponding to the relative priorities, have to be added to $b_1 T_1^{-1}(w)$ and $b_2 T_2^{-1}(w)$ before they can be combined in a single water-filling diagram. It turns out that this is the only modification we need to obtain the sum of the PSD's. Similarly to the total capacity case, in (51) and (52) and in Fig. 12, the minimum of the offset curves, $b_1 T_1^{-1}(w) + (1 - \alpha)$ and $b_2 T_2^{-1}(w) + \alpha$, is treated as the bottom of the container and the water-filling method is used with the sum of the power, $b_1 W_1 + b_2 W_2$, to determine the sum of the PSD's, $\hat{S}_1(w) + \hat{S}_2(w)$.
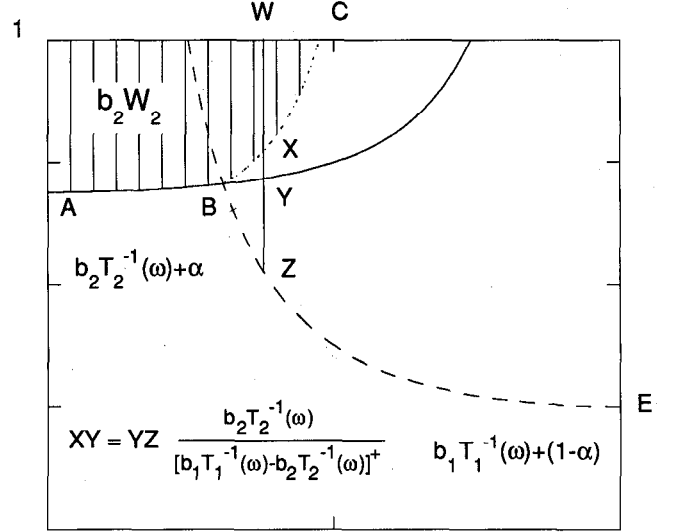
The split of the sum of the PSD's into individual PSD's is more complicated. In the case of total capacity, since the rates of the two users are weighted equally, all the signal power assigned to a frequency is sent over the channel with a higher gain at that frequency (i.e., FDMA is optimal). However, this is not optimal if the users' rates are weighted unequally. For example, if the rate of user 1 is weighted heavily, a portion of the signal power should still be transmitted over channel 1 even at frequencies where its gain is lower. Hence, at some frequencies, the signal power should be split between both users rather than completely assigned to one of the users as in the case of achieving the total capacity. As a result, the PSD's overlap.

When the PSD's of the two users overlap, the successive decoding idea suggests that the user with lower priority should be decoded first. Moreover, once the PSD of the user with higher priority is fixed, the PSD of the user with lower priority can be obtained via the classical water-filling argument with the noise equal to the sum of the original channel noise and the other user's signal. When $1/2 < \alpha < 1$, user 2 has a lower priority. Hence, from the successive decoding idea, the optimal PSD of user 2 is obtained, in (55) or in Fig. 13, via the water-filling argument with the bottom of the container equal to $\hat{T}_2^{-1}(w)$ in (45) or curve ABXC. In Fig. 13, $XZ = \hat{S}_1(w)$ and hence,

$$XY = XZ \frac{b_2 T_2^{-1}(w)}{b_1 T_1^{-1}(w)} . \tag{58}$$

After some manipulations, we can obtain the relationship

$$XY = YZ \frac{b_2 T_2^{-1}(w)}{\left[b_1 T_1^{-1}(w) - b_2 T_2^{-1}(w) \right]^+} , \tag{59}$$

which can be used to construct the curve BXC in Fig. 13.

Now, we concentrate on the numerical procedure of obtaining the PSD pair for each $1/2 < \alpha < 1$. As in the single-user channel where the PSD can be obtained by an algorithm iterating on b until the area of $[1 - bT^{-1}(w)]^+$ is equal to bW, a similar algorithm which has two levels of iterations: one on
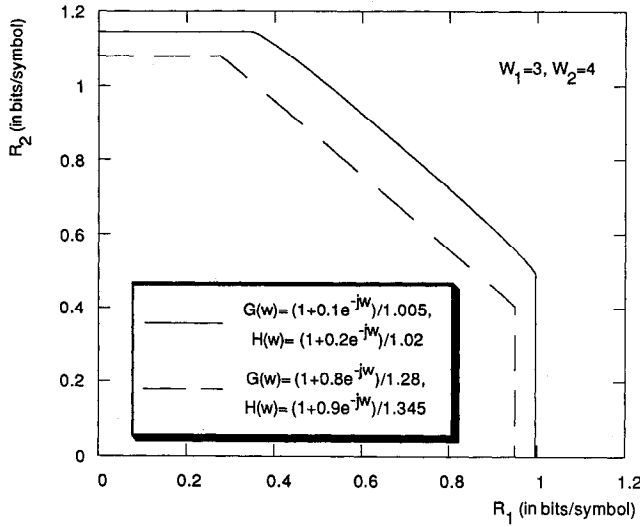
Fig. 14. Capacity regions of the channels in examples 1 and 2.

Fig. 15. Capacity regions of the channels in examples 3 and 4.

b_1 and the other on b_2, is given for the two-user channel. First, we arbitrarily fix b_1 and plot $b_1 T_1^{-1}(w) + (1 - \alpha)$ as in Fig. 12. Then, we plot $b_2 T_2^{-1}(w) + \alpha$ and adjust b_2 such that the total shaded area is equal to $b_1 W_1 + b_2 W_2$. On the same graph, we construct curve BXC, as shown in Fig. 13, using (59). If the shaded are in Fig. 13 is larger than $b_2 W_2$, we decrease the value of b_1 (thereby lowering the curve BZE); otherwise, we increase the value of b_1. The same procedure is repeated until the area above the curve ABXC is equal to $b_2 W_2$. Then, the shaded regions in Figs. 12 and 13 give the sum of the PSD's and the PSD of user 2, respectively and their difference gives the PSD of user 1.

D. Examples

To illustrate the application of the two-user water-filling argument, we give four examples representing different levels of intersymbol interference:

1) $g_1 : \left[1 + 0.1e^{-jw}\right]$, $h_1 : \left[1 + 0.2e^{-jw}\right]$
2) $g_2 : \left[1 + 0.8e^{-jw}\right]$, $h_2 : \left[1 + 0.9e^{-jw}\right]$
3) $g_3 : \left[1 + 0.1e^{-jw}\right]$, $h_3 : \left[1 - 0.2e^{-jw}\right]$
4) $g_2 : \left[1 + 0.8e^{-jw}\right]$, $h_4 : \left[1 - 0.9e^{-jw}\right]$.

In all four examples, we assume that $W_1 = 3$, $W_2 = 4$, the white Gaussian noise has variance 1 and the impulse responses of all channels are normalized to have unit energy. (i.e., $G_1(w) = (1 + 0.1e^{-jw})/\sqrt{1.01}$, $H_1(w) = (1 + 0.2e^{-jw})/\sqrt{1.04}$, etc.) In Examples 1 and 3, intersymbol interference is mild while in Examples 2 and 4, intersymbol interference is much stronger. In Examples 1 and 2, the channels for both users are low-pass while in Examples 3 and 4, one user transmits via a low-pass channel and the other user transmits via a high-pass channel. The capacity regions of the channels in Example 1 and 2 are shown in Fig. 14 while those in Example 3 and 4 are shown in Fig. 15. It is clear, from all four examples, that with the energy of the transfer functions fixed, intersymbol interference decreases the single-user capacity. However, the same is not necessarily true for the total capacity. When both channels are low-pass, the total
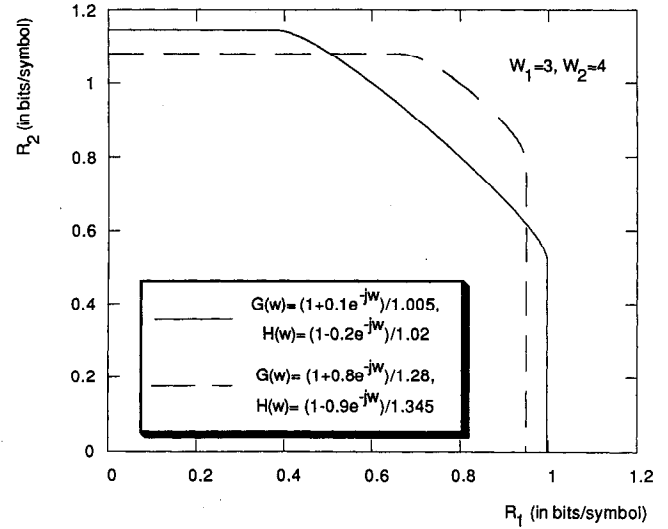
capacity decreases as intersymbol interference increases as shown in Figs. 14 and 15. When one channel is low-pass and the other is high-pass, the total capacity actually *increases* as intersymbol interference becomes stronger. This is because in that case, the increase in single-user ISI is more than offset by the decrease in multiuser interference. However, it is important to note that the single-user rate is degraded. Therefore, when the system designer has some latitude shaping the channel spectra, there is a trade-off between the single-user capacities and the total capacity.

APPENDIX A

Proof of Theorem 1: Following from the discussion in Section II, in order to find the capacity region of the two-user Gaussian linear vector channel, we need to solve the limiting expression in (3) and (4) with U_0^{N-1}, V_0^{N-1}, and Y_0^{N-1} related by (6).

Taking the DFT on both sides of (6), we have

$$\tilde{Y}_i = \tilde{G}_i \tilde{U}_i + \tilde{H}_i \tilde{V}_i + \tilde{M}_i, \qquad (A.1)$$

where $\tilde{X}_0^{N-1}$ is the vector DFT of X_0^{N-1} (i.e., the jth entries of $\tilde{X}_i$, $i = 0, \cdots, N - 1$ is the DFT of the jth entries of X_i, $i = 0, \cdots, N - 1$.) It is well known [19, p. 122] that if X_0^{N-1} is a real vector sequence, $\tilde{X}_i$ and $\tilde{X}_{N-i}$ are complex conjugate pair for $i = 1, \cdots, N - 1$. Therefore, it is sufficient to consider (A.1) for $i = 0, 1, \cdots, L$ where $L = \lfloor N/2 \rfloor$. Moreover, if N is even, $\tilde{X}_0$ is a real vector while if N is odd, both $\tilde{X}_0$ and $\tilde{X}_L$ are real vectors. For simplicity, we shall concentrate on the case when N is even, analogous argument follows for the case when N is odd.

It can be shown that since M_0^{N-1} is a sequence of zero-mean, jointly Gaussian vectors with $EM_i M_j^T = R_{(i-j)_N}$, $\tilde{M}_0^L$ is a sequence of zero-mean, independent complex Gaussian vectors. Moreover, $\tilde{M}_0$ is real with $E\tilde{M}_0 \tilde{M}_0^T = N\tilde{R}_0$ and for $i = 1, \cdots, L$, $\tilde{M}_i$ is complex with $E\tilde{M}_i \tilde{M}_i^* = N\tilde{R}_i$ where $\tilde{R}_0^{N-1}$ is the DFT of R_0^{N-1} and * denotes the complex conjugate transpose.

By the invertibility of the DFT, the transformation from X_0^{N-1} to $\tilde{X}_0^L$ is also invertible and hence, we have

$$\frac{1}{N} I(U_0^{N-1}; Y_0^{N-1} | V_0^{N-1}) = \frac{1}{N} I\left(\tilde{U}_0^L; \tilde{Y}_0^L | \tilde{V}_0^L\right)$$

$$= \frac{1}{N} \sum_{i=0}^{L} I\left(\tilde{U}_i; \tilde{Y}_i | \tilde{V}_i\right), \quad \text{(A.2)}$$

$$\frac{1}{N} I(V_0^{N-1}; Y_0^{N-1} | U_0^{N-1}) = \frac{1}{N} I\left(\tilde{V}_0^L; \tilde{Y}_0^L | \tilde{U}_0^L\right)$$

$$= \frac{1}{N} \sum_{i=0}^{L} I\left(\tilde{V}_i; \tilde{Y}_i | \tilde{U}_i\right), \quad \text{(A.3)}$$

and

$$\frac{1}{N} I(U_0^{N-1}, V_0^{N-1}; Y_0^{N-1}) = \frac{1}{N} I\left(\tilde{U}_0^L, \tilde{V}_0^L; \tilde{Y}_0^L\right)$$

$$= \frac{1}{N} \sum_{0}^{L} I\left(\tilde{U}_i, \tilde{V}_i; \tilde{Y}_i\right). \quad \text{(A.4)}$$

On the other hand, Parseval's theorem transforms the power constraint to

$$\frac{1}{N} \sum_{i=0}^{N-1} \text{tr}\left(E(U_i U_i^T)\right) = \frac{1}{N^2} \text{tr}\left(E(\tilde{U}_0 \tilde{U}_0^T)\right)$$

$$+ \frac{2}{N^2} \sum_{i=1}^{L} \text{tr}\left(E\left(\tilde{U}_i \tilde{U}_i^*\right)\right) \leq W_1. \quad \text{(A.5)}$$

Similarly, we have

$$\frac{1}{N^2} \text{tr}\left(E\left(\tilde{V}_0 \tilde{V}_0^T\right)\right) + \frac{2}{N^2} \sum_{i=1}^{L} \text{tr}\left(E\left(\tilde{V}_i \tilde{V}_i^*\right)\right) \leq W_2.$$

$$\text{(A.6)}$$

Therefore, C_N can be expressed as

$$C_N = \bigcup_{\substack{\tilde{U}_0^L, \tilde{V}_0^L: \\ \frac{1}{N^2}\left[E\tilde{U}_0\tilde{U}_0^T + \sum_{i=1}^{L} 2\,\text{tr}(E(\tilde{U}_i\tilde{U}_i^*))\right] \leq W_1 \\ \frac{1}{N^2}\left[E\tilde{V}_0\tilde{V}_0^T + \sum_{i=1}^{L} 2\,\text{tr}(E(\tilde{V}_i\tilde{V}_i^*))\right] \leq W_2}} \left\{ (R_1, R_2): \begin{array}{c} 0 \leq R_1 \leq \frac{1}{N} \sum_0^L I\left(\tilde{U}_i; \tilde{Y}_i \mid \tilde{V}_i\right) \\ 0 \leq R_2 \leq \frac{1}{N} \sum_0^L I\left(\tilde{V}_i; \tilde{Y}_i \mid \tilde{U}_i\right) \\ R_1 + R_2 \leq \frac{1}{N} \sum_0^L I\left(\tilde{U}_i, \tilde{V}_i; \tilde{Y}_i\right) \end{array} \right\},$$

$$\text{(A.7)}$$

where the union is taken over all real random vectors $\tilde{U}_0$ and $\tilde{V}_0$, and all complex random vectors $\tilde{U}_i$ and $\tilde{V}_i$, $i = 1, \cdots, L$.

Since $\tilde{M}_i$ are complex Gaussian random vectors, each channel in (A.1) is a complex Gaussian vector channel, it can be shown, by extending the proof in the real random vector case (e.g., [20, p. 248]), that complex Gaussian distributions maximize the three mutual information quantities simultaneously. Then, we have (A.8) (shown at the bottom of the page).

Finally, using standard properties of Riemann integration, we have, in the limit as $N \rightarrow \infty$, the desired result. $\qquad \square$

APPENDIX B

Proof of Corollary 1: When the inputs are scalars, we can apply the equality $\det(I + AB) = \det(I + BA)$ and we have

$$\log \det\left[I + G(w)S_1(w)G^*(w)N^{-1}(w)\right]$$
$$= \log\left[1 + S_1(w)G^*(w)N^{-1}(w)G(w)\right], \quad \text{(B.1)}$$

$$\log \det\left[I + H(w)S_2(w)H^*(w)N^{-1}(w)\right]$$
$$= \log\left[1 + S_2(w)H^*(w)N^{-1}(w)H(w)\right], \quad \text{(B.2)}$$

and

$$\log \det[I + G(w)S_1(w)G^*(w)N^{-1}(w)$$
$$+ H(w)S_2(w)H^*(w)N^{-1}(w)]$$

$$= \log \det\left[I[G(w)H(w)] \begin{bmatrix} S_1(w) & 0 \\ 0 & S_2(w) \end{bmatrix} \right.$$
$$\left. \cdot \begin{bmatrix} G^*(w) \\ H^*(w) \end{bmatrix} N^{-1}(w)\right] \quad \text{(B.3)}$$

$$= \log \det\left[I + \begin{bmatrix} S_1(w) & 0 \\ 0 & S_2(w) \end{bmatrix} \right.$$
$$\left. \cdot \begin{bmatrix} T_1(w) & T_{12}(w) \\ T_{12}^*(w) & T_2(w) \end{bmatrix}\right] \quad \text{(B.4)}$$

$$= \log[1 + S_1(w)T_1(w) + S_2(w)T_2(w)$$
$$+ S_1(w)S_2(w)\left[T_1(w)T_2(w)\right.$$
$$\left. - |T_{12}(w)|^2\right]\right]. \quad \text{(B.5)}$$

$$C_N = \bigcup_{\substack{\sum_{10} \in \mathbb{R}^{p \times p}, \sum_{20} \in \mathbb{R}^{q \times q}, \sum_{1i} \in \mathbb{C}^{p \times p}, \sum_{2i} \in \mathbb{C}^{q \times q}: \\ \sum_{1i}, \sum_{2i} \geq 0 \quad \forall i = 0, \cdots, L \\ \frac{1}{N}\left[\text{tr}(\sum_{10}) + 2 \sum_{i=0}^{L} \text{tr}(\sum_{1i})\right] \leq W_1 \\ \frac{1}{N}\left[\text{tr}(\sum_{20}) + 2 \sum_{i=1}^{L} \text{tr}(\sum_{2i})\right] \leq W_2}} \left\{ (R_1, R_2): \begin{array}{c} 0 \leq R_1 \leq \frac{1}{N} \sum_{i=0}^{L} \log \det\left[I + \tilde{G}_i \sum_{1i} \tilde{G}_i^* \tilde{R}_i^{-1}\right] \\ 0 \leq R_2 \leq \frac{1}{N} \sum_{i=0}^{L} \log \det\left[I + \tilde{H}_i \sum_{2i} \tilde{H}_i^* \tilde{R}_i^{-1}\right] \\ R_1 + R_2 \leq \frac{1}{N} \sum_{i=0}^{L} \log \det\left[I + \tilde{G}_i \sum_{1i} \tilde{G}_i^* \tilde{R}_i^{-1} + \tilde{H}_i \sum_{2i} \tilde{H}_i^* \tilde{R}_i^{-1}\right] \end{array} \right\} \quad \text{(A.8)}$$

When the output is also scalar, it is easy to see that $T_1(w)T_2(w) = |T_{12}(w)|^2$ and we have the desired result. $\square$

Proof of Corollary 2: The result follows straightforwardly from Corollary 1, the convexity of the capacity region and traditional water-filling argument. The power spectral densities that achieve the corner points in the pentagon are discussed in Section III. $\square$

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, July & October 1948.

[2] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Inform. Theory*, Tsahkadsor, Armenia, U.S.S.R., Sept. 1971, pp. 23–52.

[3] H. Liao, "A coding theorem for multiple access communications," presented at *Int. Symp. Inform. Theory*, Asilomar, 1972.

[4] R. Ahlswede, "The capacity of a channel with two senders and two receivers," *Ann. Probab.*, vol. 2, pp. 805–814, Oct. 1974.

[5] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* London: Academic Press, 1981.

[6] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. New York: Springer, 1978.

[7] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, Jan. 1949.

[8] R. G. Gallager, *Information Theory and Reliable Communication.* New York: John Wiley, 1968.

[9] B. S. Tsybakov, "Capacity of a discrete-time Gaussian channel with a filter," *Prob. Peredach. Inform.*, vol. 6, no. 3, pp. 78–82, July–Sept. 1970.

[10] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 6, pp. 267–271, Nov. 1972.

[11] S. Verdú, "Multiple-access channels with memory with and without frame synchronism," *IEEE Trans. Inform. Theory*, vol. 35, no. 3, pp. 605–619, May 1989.

[12] ——, "The capacity region of the symbol-asynchronous Gaussian multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 35, no. 4, pp. 733–751, July 1989.

[13] C. W. Keilers, "The capacity of the spectral Gaussian multiple-access channel," Ph.D. thesis, Stanford Univ., May 1976.

[14] C. M. Zeng, N. He, and F. Kuhlmann, "Capacity region of a waveform Gaussian multiple-access channel," in *Book of Abstracts of the 1990 Int. Symp. Inform. Theory*, San Diego, CA, January 1990, p. 93.

[15] W. Hirt and J. L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. Inform. Theory*, vol. 34, no. 3, pp. 380–388, May 1988.

[16] L. H. Brandenburg and A. D. Wyner, "Capacity of the Gaussian channel with memory: The multivariate case," *Bell Syst. Tech. J.*, vol. 53, no. 5, pp. 745–778, May–June 1974.

[17] S. Verdú, "Capacity region of Gaussian CDMA channels: The symbol-synchronous case," in *Proc. Twenty-fourth Allerton Conf. Commun., Contr. Comput.*, Allerton, IL, Oct. 1986, pp. 1025–1034.

[18] T. Cover, *Some Advances in Broadcast Channels.* New York: Academic Press, 1975, vol. 4, pp. 229–260.

[19] A. V. Oppenheim and D. V. Schafer, *Digital Signal Processing.* Englewood Cliffs, NJ: Prentice Hall, 1975.

[20] R. E. Blahut, *Principles and Practice of Information Theory.* Reading, MA: Addison Wesley, 1987.