

# The Effect upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel

Muriel Médard, *Member, IEEE*

**Abstract**—We present a model for time-varying communication single-access and multiple-access channels without feedback. We consider the difference between mutual information when the receiver knows the channel perfectly and mutual information when the receiver only has an estimate of the channel. We relate the variance of the channel measurement error at the receiver to upper and lower bounds for this difference in mutual information. We illustrate the use of our bounds on a channel modeled by a Gauss–Markov process, measured by a pilot tone. We relate the rate of time variation of the channel to the loss in mutual information due to imperfect knowledge of the measured channel.

**Index Terms**—Channel uncertainty, multiple-access channels, mutual information, time-varying channels, wireless communications.

## I. INTRODUCTION

THE increasing applications of wireless communications have spawned much research and debate about the best manner to utilize the available spectrum, and have prompted more spectrum to be opened to commercial uses. The study of the intrinsic limitations of wireless communications is very important because most of the factors are outside the engineering domain: the power used by mobile users is limited by safety regulations, the carrier frequency and the bandwidth are determined by commercial availability under regulatory constraints, and the nature of the channels is governed by weather, car traffic, physical obstacles, etc. It is therefore necessary to take such constraints into account in determining the capacity of these channels. The goal of this paper is to determine how the time variations cause uncertainty about the channel at the receiver, and how this uncertainty affects capacity. We first consider the capacity of a known time-varying channel and second we use a perturbation approach on a perfectly known channel case to account for the channel estimation error. We consider one or several senders and a single receiver.

The issues of propagation in mobile channels have been extensively studied and can be found in many textbooks (see, for

instance, [29], [38], [64], and [66]). Treatment of many information-theoretic issues relating to mobile communications can be found in [62], [47], and [27]. Three main issues which arise from the characterization of the mobile channels are intersymbol interference (ISI), time variation of channels, and multiple access. The effect of time-invariant intersymbol interference on capacity for a single-user ([24], [70], [28]) and multiple-user systems ([7], [74], [75]) has been well studied. The time variation of the channels proves much more difficult to model and circumscribe. A commonly taken approach is to consider that the ISI is constant and known for certain durations and thus decompose the channel into time blocks [43], [13], [35]. The issue on which we concentrate in this paper is the effect of the lack of knowledge of the channel on the achievable mutual information in systems with time-varying channels. Of particular interest is the issue of the effect of imperfect knowledge on multiple-user communications. The capacity of time-varying channels has been often studied in the context of Markov channels. Different cases have been considered, among them: the state of the channel is known at the sender and receiver, or the receiver only ([78], [15], [21]), or only knowledge of the statistics of the state transition probabilities exists ([22], [46]), or solely the set of possible states is known ([12]). For non-Markov channels that decorrelate in time, a strong coding theorem [44] and weak coding theorems, which use information stability results [50], [68], have been established [31], [49]. The last issue we have mentioned, the multiple-access problem, has been well studied in the memoryless case. The capacity region in the memoryless case was established with synchronicity ([41], [1]) and asynchronicity ([9], [51]). Overviews of the multiple-access problem are given in [16], [20], [71], and [45].

In this paper, we consider the issues of ISI, time-variation, and multiple access in the context of an error about the channel measurement available at the receiver. The time variation of the channel and the error on the estimate of the channel are tightly linked. The time variation entails that the channel will always remain unknown to some extent. Indeed, a time-invariant channel could be measured arbitrarily well by using a sounding probe signal. The time variation and the ISI together are responsible for the fading seen on the channel. The questions we seek to investigate in this paper are as follows.

- For a given channel measurement error variance, what is the effect on mutual information of not knowing the channel? In particular, what is the effect of imperfect

Manuscript received July 30, 1997; revised August 29, 1999. The material in this paper was presented in part at the 1995 IEEE Vehicular Technology Conference, at the 1995 IEEE International Symposium on Information Theory, and at the 1996 IEEE Vehicular Technology Conference.

The author is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: medard@mit.edu).

Communicated by S. Shamai, Associate Editor for Shannon Theory.  
Publisher Item Identifier S 0018-9448(00)02892-3.

channel knowledge on mutual information in the multiple access case?

- For a specific channel fading model, how does the channel variation affect the mutual information?

The question concerning the effect of imperfect channel knowledge on mutual information in the multiple-access case is particularly relevant, because the corners of the Cover–Wyner region for the multiple-access capacity region in memoryless channels can be attained using interference cancellation. Other points along the Cover–Wyner region can be obtained using rate splitting and interference cancellation ([58], [59], [23], [57], [56]). The error on the channel of a user will affect the mutual information of another user if interference cancellation is used. It is not immediately clear whether, when the channel is not perfectly known, it is preferable to consider each user to be interference to all other users rather than attempt multiple-user interference cancellation. This issue is the subject of debate ([73], [17], [77], [18]).

## II. CHANNEL MODEL AND MUTUAL INFORMATION WITH PERFECT KNOWLEDGE OF THE CHANNEL AT THE RECEIVER

### A. Channel Model

Mobile communication channels are subject to many changes which cannot be controlled or predicted. Obstacles appear and disappear as the mobile moves in its environment, leading to partial blocking (shadowing) and to echoes (multipath). We present a channel model based on a general multipath situation. We establish equivalent continuous and discrete-time models for this general multipath situation. In Section II, we present a channel model for a time-varying multipath environment and results for the maximum mutual information of time-varying channels in the single-user and multiple-access cases when the channel is perfectly known both at the sender(s) and the receiver. In Section III, we establish the framework for studying the mutual information when the channel is known with some error at the receiver. For both the single-user and multiple-access cases, we use this framework to establish upper and lower bounds for the loss in mutual information due to an error of known variance for the channel at the receiver. In Section IV, we consider explicitly the effect of time variations upon the channel error for a Gauss–Markov model of variations. In Section V, we draw conclusions from these results and propose directions for further research.

Let us look at the channel response at baseband for a multipath channel. We assume that the passband signal is contained in a bandwidth  $W'$  centered around a carrier frequency  $f_0$ . A discrete data stream from the user is passed through a modulator, whose output is a continuous signal. We denote the channel by a linear time-varying filter of impulse response  $g$ . In this paper, unless the context indicates otherwise, lower case variables denote deterministic variables and upper case letters denote random variables. If we have reflections off objects in motion relative to the transmitter or receiver, the received signals will experience a Doppler spread in frequency ([14], [34], [38]).

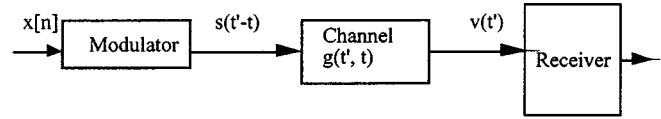


Fig. 1. Continuous-time multipath system.

The impulse response of the multipath channel seen at time  $t'$  for an input at  $t' - t$  is

$$g(t', t) = \sum_{\text{all paths } m} g^m(t', t) \quad (1)$$

where  $g^m(t', t)$  is the response of the  $m$ th path at time  $t'$  for an input at  $t' - t$ . Thus  $g^m(t', t) = a^m(t')\delta(\tau^m(t') - t)$ , where  $a^m(t')$  is the complex multiplicative factor associated with the  $m$ th path (accounting for fading and phase shift). The Doppler shift associated with path  $m$  is  $B^m$  and the time-varying delay associated with path  $m$  is  $\tau^m(t) = \tau^m + (B^m/f_0)t$ . The spread among these shifts in frequency is the Doppler spread, denoted by  $B_D$ . The difference between the smallest and the largest  $\tau^m$  is called the time spread, denoted by  $T_s$ . Since  $|a^m(t')|$  changes slowly with the delay, we may use the common assumption ([53, p. 705]) that the amplitude  $|a^m(t')|$  changes much more slowly than the phase. In terms of the Bello functions for the description of time-varying channels,  $g(t', t)$  is the input delay-spread function [4], [25, p. 128]. Fig. 1 shows our continuous-time model.

In order to derive a discrete-time model for the system, we must determine how to sample the channel output. In general, we know that the bandwidth of the output of a linear time-varying filter will be band-limited to the sum of the input bandwidth plus the filter variation bandwidth (a more general result for stochastic channels was established in [33]). Therefore, the output may be band-limited to  $W = W' + B_D$  at baseband. We assume that the input and the noiseless response at baseband are band-limited to  $[-(W/2), +(W/2)]$ . The response may be sampled at time intervals of  $1/W$ , by the Nyquist sampling theorem. We define the baseband output  $y(t')$  of a multipath channel with additive white Gaussian noise (AWGN) to be the noiseless baseband output with the addition of a term  $n(t')$ , where  $n(t')$  is a sample function of band-limited complex white Gaussian noise of double-sided spectral density  $N_0$  at baseband  $[-(W/2), (W/2)]$ . We consider the discrete-time sequences defined as  $v[k] = v(k/W)$  and  $s[k] = s(k/W)$ . The output  $y[k]$  is  $v[k] + n[k]$ , where  $n[k]$  is the sampled version of  $n(t)$ . Fig. 2 shows the sampled system. We have that

$$y[k] = \sum_{\text{all paths } m} \sum_{n=-\infty}^{+\infty} s[k-n]g^m[k, n] + n[k] \quad (2)$$

$$g^m[k, n] = a^m\left(\frac{k}{W}\right) \text{sinc}\left\{\pi W\left(\frac{n}{W} - \tau^m\left(\frac{k}{W}\right)\right)\right\}. \quad (3)$$

We see from (3) that there are infinitely many terms  $g^m[k, n]$ , whereas in the continuous case there is, for any given  $t'$ , at most a single  $t$  which yields a nonzero value for  $g^m(t', t)$ . For our

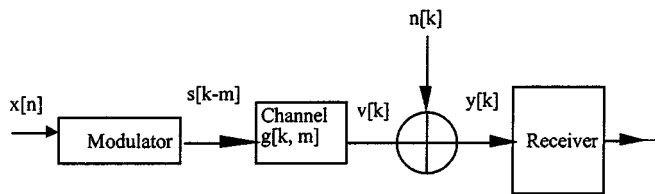


Fig. 2. Discrete-time model.

purposes, we wish to establish an expression in terms of finite matrices in the discrete-time case. Expression (3) shows that the terms  $g^m[k, n]$  become very small with  $n$  for  $k$  fixed. Therefore, we may choose some artificial cutoff in  $n$  such that  $g^m[k, n]$  is approximated to be zero beyond that point. We have used the sinc function for our sampling reconstruction, but we could instead use a function with significantly faster rolloff in time. A detailed treatment of the issues concerning time and band-limiting assumptions can be found in [63], [80], and [79]. Note that the spreading in both time and frequency is crucial ([32], for instance, does not take this into account).

In what follows, a subscript  $k$  after a vector  $\underline{v}$  indicates that we are considering  $[v[1], \dots, v[k]]^T$ . If we are considering a matrix, the superscript will indicate the range of the columns in the same manner. We assume that  $s[n]$  for any  $n \leq 0$  is zero. Thus  $y[1], \dots, y[k]$  will depend only on inputs  $s[1], \dots, s[k]$ . Let us choose some cutoff  $\Delta > WT_s$ . We approximate the sampled channel output as

$$\underline{y}_k = \underline{f}^k \underline{s}_k + \underline{n}_k \quad (4)$$

where  $\underline{f}^k$  is the complex matrix with entries

$$f[j, i] = \begin{cases} \sum_{\text{all paths } m} g^m[j, j-i], & \text{for } 0 \leq j-i \leq \Delta \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

Although the vector expression is approximate, we may make this approximation arbitrarily good by the choice of  $\Delta$ .

For the multiple-access model, each source has its own time-varying channel. Fig. 3 shows the multiple-user model. We let  $K$  denote the number of users. It is reasonable to assume that the channels of the different sources are mutually independent. The signal from each modulator passes through a different channel. The input bandwidth, of size  $W'$ , is shared by all users. This does not necessarily mean that all users simultaneously transmit over the same portions of the spectrum. The received signal is the sum of all these modulated inputs, as received through their respective filters. Let  $P_i$  be the number of paths in the channel seen by user  $i$ ,  $a_i^m$  be the complex multiplicative constant associated with the  $m$ th path of user  $i$  (accounting for fading and phase shift), and  $\tau_i^m$  be the delay associated with the  $m$ th path of user  $i$ . Let  $s_i(t)$  be the complex signal transmitted by user  $i$ . It is sufficient to take  $W$  to be  $W' + \max_{i=1, \dots, K} \{B_D^i\}$  where  $B_D^i$  is the Doppler spread of the  $i$ th channel. By extending the discussion for the single-user case, we may write that

$$y[k] = \sum_{i=1}^K \sum_{m=1}^{P_i} \sum_{n=-\infty}^{+\infty} s_i[k-n] g_i^m[k, n] + n[k] \quad (6)$$

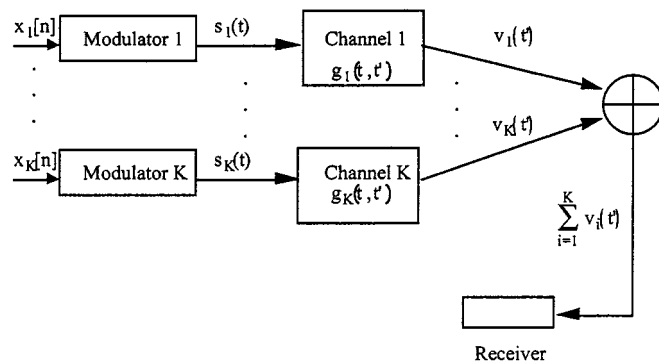


Fig. 3. Continuous time multipath system with multiple users.

where

$$g_i^m[k, n] = a_i^m \left( \frac{k}{W} \right) \text{sinc} \left\{ \pi W \left( \frac{n}{W} - \tau_i^m \left( \frac{k}{W} \right) \right) \right\}.$$

As in the single-user case, our initial assumption that  $W$  is large enough so that  $y(t)$  may be fully recovered from  $y[k]$  ensures that (6) gives a complete characterization of the channel response. As in the single-user case, we choose some arbitrary cutoff in  $n$  such that  $g_i^m[k, n]$  is taken to be zero beyond that point. We may then write, with the obvious extensions of notation from the single user case, that

$$\underline{y}_k = \sum_{i=1}^K \underline{f}_{i,k} \underline{s}_{i,k} + \underline{n}_k. \quad (7)$$

### B. Maximum Mutual Information in the Case of a Perfectly Known Channel at the Sender and the Receiver

In this section, we consider the case where both the receiver and the sender have perfect knowledge of the channel for all times. In the case of a time-varying channel, this would require knowledge of the future behavior of the channel, which is not a reasonable assumption. The purpose of this section is simply to establish how we can compute, for a particular realization of the channel, an upper bound on the mutual information achievable over some time. Note that we establish a mutual information and not an achievable rate *per se*, since we do not look at limits over time. We first review the results for the single-user case and the multiple-user case when the users cooperate. We then present a new result concerning the maximum mutual information in the case of independent multiple users. For complex transmission over time  $T$  using an input bandwidth  $W$ , the mutual information between input and output is

$$I(\underline{Y}_k; \underline{S}_k) = h(\underline{Y}_k) - h(\underline{N}_k) \quad (8)$$

where  $k = \lfloor TW \rfloor$  and  $h$  denotes differential entropy

$$h(Y) = - \int_y p_Y(y) \ln(p_Y(y)) dy. \quad (9)$$

We actually have  $2k$  degrees of freedom since we are dealing with complex random variables. In order to simplify our manipulations, we shall use the random vectors  $\underline{S}'_{2k}$ ,  $\underline{Y}'_{2k}$ , and  $\underline{N}'_{2k}$ ,

whose first  $k$  components and last  $k$  components are, respectively, the real and imaginary parts of the corresponding vectors  $\underline{S}_k$ ,  $\underline{Y}_k$ , and  $\underline{N}_k$ . We use such a notation instead of the more compact complex notation because, in the following sections, we restrict ourselves to real vectors, for the sake of simplicity. Thus in our paper, all operations are in the real domain and there is no ambiguity. Let us denote by  $\tilde{S}[k']$  the sample at time  $k'/W'$  of the continuous input  $S(t)$ . We may write

$$Y[i] = \sum_{\text{all paths}} \sum_{m=-\infty}^{+\infty} \tilde{S}[i-n] \tilde{g}^m[i, n] + N[i] \quad (10)$$

where

$$\tilde{g}^m[i, n] = \text{sinc} \left\{ \pi W' \left( \frac{i}{W} - \tau^m(t) - \frac{i-n}{W'} \right) \right\} a^m \left( \frac{i}{W} \right). \quad (11)$$

Therefore, making assumptions of approximate simultaneous time and bandwidth limiting, if we take  $S'$  to be 0 before sample 1 and we choose a  $\tilde{\Delta}$  large enough, we may write that

$$[Y[1] \cdots Y[k]]^T = \underline{f}_{k'}^k [\tilde{S}[1] \cdots \tilde{S}[k']]^T + [N[1] \cdots N[k]]^T \quad (12)$$

where  $k' = \lfloor W'T \rfloor$ ,  $k = \lfloor W(T + \tilde{\Delta}) \rfloor$  and

$$\begin{cases} \tilde{f}[i, j] = \tilde{g}[i, i-j], & \text{for } 0 \leq i-j \leq \tilde{\Delta} \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

The matrix  $\underline{f}_{k'}^k$  gives, in terms of a linear transformation, both the effect of sampling input and output at different rates and the effect of the channel. Let us define the real vector  $\underline{S}'_{2k'}$  constructed from the complex vector  $\underline{S}_{k'}$  similarly to the way we constructed  $\underline{S}'_{2k}$  from  $\underline{S}_k$ . We have that

$$\underline{Y}_{2k} = \underline{f}'_{2k'} \underline{S}'_{2k'} + \underline{N}'_{2k} \quad (14)$$

where  $\underline{f}'_{2k'}$  can be easily expressed using (11) and (13).

Let us consider the  $2k'$  by  $2k'$  matrix

$$\left( \underline{f}'_{2k'} \right)^T \underline{f}'_{2k'}.$$

Let  $\lambda_1, \dots, \lambda_{2k'}$  be its eigenvalues. These eigenvalues are real [65, p. 222] and nonnegative. Using [15, Theorem 8.4.1] and water-filling arguments similar to those of [10], we may establish that maximum mutual information per second is

$$\frac{1}{2T} \sum_{i=1}^{2k'} \ln \left( 1 + \frac{u_i \lambda_i}{\frac{WN_0}{2}} \right)$$

where  $u_i$  is given by

$$u_i = \left( \gamma - \frac{WN_0}{2\lambda_i} \right)^+ \quad (15)$$

(using  $^+$  to denote the positive part) and

$$\sum_{i=1}^{2k'} u_i = TPW. \quad (16)$$

Let us consider the multiple-access case. We place a constraint  $P$  on the sum of all the  $K$  users' powers. We first look at the case where the users may cooperate, and therefore act as an antenna

array. Such a model is only reasonable if the users are colocated or linked to each other in some fashion. In particular, several cooperating users with several antennas may achieve a rate greater than the sum of the rates of the users. Again, we consider mutual information rather than capacity and the techniques we use are similar to the single-user case. Note that different types of decomposition might be used ([3]). For the memoryless case, without the ISI we consider here, the capacity region for multiple-access channels with arbitrarily correlated sources is given in [8] and the worst case capacity for  $s$  senders and  $r$  receivers is given in [72]. For issues concerning multiple-antenna systems in fading channels, see [67], [62], and [42]. The problem is similar to the one-dimensional problem, except that we now consider that the channel has more degrees of freedom at the input than at the output. There are  $M = 2Kk'$  input degrees of freedom and  $2k$  output degrees of freedom. We have

$$[Y[1] \cdots Y[2k]] = \underline{f}'_{2k'} [\hat{S}[1] \cdots \hat{S}[M]]^T + [N[1] \cdots N[2k]] \quad (17)$$

where we have defined

$$\begin{aligned} [\hat{S}[1] \cdots \hat{S}[M]] &= [\tilde{S}_1[1] \cdots \tilde{S}_1[2k'], \tilde{S}_2[1] \cdots \tilde{S}_2[2k'], \dots, \\ &\quad \tilde{S}_K[1] \cdots \tilde{S}_K[2k']] \\ \underline{f}'_{2k'} &= [\underline{f}'_{1,2k'}, \underline{f}'_{2,2k'}, \dots, \underline{f}'_{K,2k'}]. \end{aligned} \quad (18)$$

$\underline{f}'_{2k'} \underline{f}'_{2k'}^T$  has  $M$  eigenvalues, all of which are real and nonnegative and at most  $2k$  of which are nonzero. Let us assume that there are  $\kappa$  positive eigenvalues, which we denote  $\hat{\lambda}_1, \dots, \hat{\lambda}_\kappa$ . We have decomposed our multiple-access channels into  $\kappa$  channels which may be interpreted as parallel independent channels. The input has  $M - \kappa$  additional degrees of freedom, but those degrees of freedom do not reach the output. The maximization along the active  $\kappa$  channels may now be performed using water-filling techniques. Let  $T$  be the duration of the transmission. We choose

$$u_i = \left( \gamma - \frac{N_0 W}{2\hat{\lambda}_i} \right)^+ \quad (19)$$

for  $\hat{\lambda}_i \neq 0$ , where  $\gamma$  satisfies

$$\sum_{i \text{ such that } \hat{\lambda}_i \neq 0} \left( \gamma - \frac{N_0 W}{2\hat{\lambda}_i} \right)^+ = TPW \quad (20)$$

and  $u_i$  satisfies

$$\sum_{i=1}^{2k} u_i = TPW. \quad (21)$$

We have reduced several channels, each with its own user, to a single channel with a composite user. The sum of all the mutual informations averaged over time is upper-bounded by

$$\frac{1}{T} \sum_{i \text{ such that } \hat{\lambda}_i \neq 0} \frac{1}{2} \ln \left( 1 + \frac{\left( \gamma - \frac{N_0 W}{2\hat{\lambda}_i} \right)^+ \hat{\lambda}_i}{\frac{N_0 W}{2}} \right).$$

Let us now consider that each source has an individual power constraint, that the signal of each source is independent of every

other source, but that we still seek to maximize the sum of all the mutual informations averaged over time. Such a model is reasonable for mobile users which do not cooperate. Note that this is different from the case considered in [35], where the channel is essentially assumed to be block-fading. We seek to maximize

$$I\left(\left\{S_{i_{2k'}}\right\}_{i=i}^K; \underline{Y}_{2k}\right) = h(\underline{Y}_{2k}) - h(\underline{N}_{2k}). \quad (22)$$

Maximizing (22) is equivalent to maximizing  $h(\underline{Y}_{2k})$ . For any covariance matrix  $\underline{\Delta}_{Y_{2k}}$  for  $\underline{Y}_{2k}$ , the entropy is maximized by  $\underline{Y}_{2k}$  Gaussian, which can be obtained by Gaussian inputs. Let us denote the autocorrelation of a matrix  $\underline{M}$  by  $\underline{\Lambda}_M$ . Then the problem, for the two-user case, is to maximize

$$\max \ln \left( \underline{f}_1^{2k} \underline{\Lambda}_{S_1} \underline{f}_1^{2kT} + \underline{f}_2^{2k} \underline{\Lambda}_{S_2} \underline{f}_2^{2kT} + \underline{\Lambda}_{N_{2k}} \right) \quad (23)$$

where we have omitted the  $2k$  subscript in  $\underline{\Lambda}_{S_i}$ . The maximization of (23) is subject, for  $i = 1, 2$ , to  $\underline{\Lambda}_{S_i}$  being positive-semidefinite and having a trace upper-bounded by  $PWT$ .

Fortunately, the problem is convex  $\cap$  and the region over which we optimize is convex. Let us first show that the region over which we optimize is convex and then show that the problem is convex  $\cap$ . Let  $\underline{Z}_{2k}$  be the output of the multi-access channel before noise is added, so the covariance of  $\underline{Z}_{2k}$  is

$$\underline{f}_1^{2k} \underline{\Lambda}_{S_1} \underline{f}_1^{2kT} + \underline{f}_2^{2k} \underline{\Lambda}_{S_2} \underline{f}_2^{2kT} = \underline{\Lambda}_{Z_{2k}} \quad (24)$$

where  $\underline{\Lambda}_{S_1}$  and  $\underline{\Lambda}_{S_2}$  are a pair of covariance matrices that satisfy our conditions. Similarly, letting  $\underline{\Lambda}_{S'_1}$  and  $\underline{\Lambda}_{S'_2}$  be another pair of covariance matrices that satisfy our conditions, let  $\underline{Z}'_{2k}$  be the corresponding output before noise so that

$$\underline{f}_1^{2k} \underline{\Lambda}_{S'_1} \underline{f}_1^{2kT} + \underline{f}_2^{2k} \underline{\Lambda}_{S'_2} \underline{f}_2^{2kT} = \underline{\Lambda}_{Z'_{2k}}. \quad (25)$$

Let us consider an AWGN channel with noise  $\underline{N}_{2k}$ . For any  $\theta$ ,  $0 \leq \theta \leq 1$ , and for each  $i$ , let  $S''_{i_{2k}}$  be a random variable whose probability density function (pdf) is the sum of  $\theta$  times the pdf of  $S_{i_{2k}}$  and  $1 - \theta$  times the pdf of  $S'_{i_{2k}}$ . Then, since the  $S_i$ s are independent

$$\underline{\Lambda}_{S''_i} = \theta \underline{\Lambda}_{S_i} + (1 - \theta) \underline{\Lambda}_{S'_i}. \quad (26)$$

Note that  $S''_{i_{2k}}$  satisfies the trace and positive semidefiniteness constraints. Thus the region over which we maximize is convex.

Let  $\underline{Z}''_{2k}$  be the corresponding output before noise. Then

$$\underline{\Lambda}_{Z''_{2k}} = \theta \underline{\Lambda}_{Z_{2k}} + (1 - \theta) \underline{\Lambda}_{Z'_{2k}}. \quad (27)$$

Then, from [15, Theorem 4.4.2], we know that

$$\theta h(\underline{Z}_{2k} + \underline{N}_{2k}) + (1 - \theta) h(\underline{Z}'_{2k} + \underline{N}_{2k}) \leq h(\underline{Z}''_{2k} + \underline{N}_{2k}). \quad (28)$$

The variance of  $\underline{Z}''_{2k} + \underline{N}_{2k}$  is given by  $\theta \underline{\Lambda}_{Z_{2k}} + (1 - \theta) \underline{\Lambda}_{Z'_{2k}} + \underline{\Lambda}_{N_{2k}}$ . Since entropy is maximized by a Gaussian distribution for any given covariance matrix

$$h(\underline{Z}''_{2k} + \underline{N}_{2k}) \leq \ln((2\pi e)^{2k} |\theta \underline{\Lambda}_{Z_{2k}} + (1 - \theta) \underline{\Lambda}_{Z'_{2k}} + \underline{\Lambda}_{N_{2k}}|). \quad (29)$$

Choosing each  $S_{i_{2k}}$  and each  $S'_{i_{2k}}$  to be Gaussian, we may combine (28) and (29) to yield

$$\theta \ln \left( |\underline{\Lambda}_{Z_{2k}} \underline{\Lambda}_{N_{2k}}| \right) + (1 - \theta) \ln(|\underline{\Lambda}_{Z'_{2k}} + \underline{\Lambda}_{N_{2k}}|) \leq \ln(|\theta \underline{\Lambda}_{Z_{2k}} + (1 - \theta) \underline{\Lambda}_{Z'_{2k}} + \underline{\Lambda}_{N_{2k}}|). \quad (30)$$

Therefore, the function in (22) is convex  $\cap$ . Since the region over which we maximize is convex, any kind of hill-climbing technique can be used to perform the maximization. In particular, we may use water-filling alternately over the two users. For any given covariance matrix for user 1, we may optimize the distribution for user 2. Then, we may take that distribution for user 2 as being fixed and optimize the distribution for user 1, and so on. A stationary point found by this technique is guaranteed to be a global maximum by our convexity results.

### III. THE EFFECT OF A CHANNEL ERROR OF KNOWN VARIANCE

In Section II, we did not establish how we would determine the channel at both the receiver and the sender. In this section, we assume that the channel is unknown at the sender and partially unknown at the receiver. Such is the situation if, for instance, we are transmitting without feedback from the receiver to the sender(s). The senders have no way of knowing the channel. The receiver must either estimate the channel(s) from the signals sent or from some separate sounding signals. If the channels are not perfectly known, it is reasonable to model them as a known part with a probabilistic additive component. As before, the input at each time is a random variable  $S_i[n]$  whose joint distribution can be chosen subject to a power constraint.

#### A. Single-User Case

We look at a fixed constraint length  $k$ , or equivalently  $k'$ . Whether or not a coding theorem applies depends on how the channel varies asymptotically. See [76] for a general single-user coding theorem and [26] for a general coding theorem for the multiple-access case. In the following, replacing sample values with random variables and using our real notation for complex vectors, we have

$$\underline{Y} = \underline{F} \underline{S} + \underline{N} \quad (31)$$

where we omit the  $2k$  and  $2k'$  for ease of notation. The mutual information between input and output is, by definition

$$I(\underline{Y}; \underline{S}) = h(\underline{S}) - h(\underline{S} | \underline{Y}). \quad (32)$$

Similarly, the mutual information given perfect knowledge of the channel is

$$I(\underline{Y}; \underline{S} | \underline{F}) = h(\underline{S} | \underline{F}) - h(\underline{S} | \underline{Y}, \underline{F}). \quad (33)$$

Note that our model assumes that the conditional probability of  $Y$  given  $S$  is known. In Section IV, we show an example of a channel model where this is indeed the case. If this conditional probability is not known, we would need to contend with issues of channel mismatch ([6]).

Since  $\underline{S}$  and  $\underline{F}$  are independent, we may write  $h(\underline{S} | \underline{F}) = h(\underline{S})$ . Therefore,

$$I(\underline{Y}; \underline{S} | \underline{F}) - I(\underline{Y}; \underline{S}) = I(\underline{S}; \underline{F} | \underline{Y}). \quad (34)$$

Although the above expression is simple, computing it in any particular case may be difficult. However, we can make some general statements about  $I(\underline{Y}; \underline{S} | \underline{F}) - I(\underline{Y}; \underline{S})$ . We have the intuition that the difference in how well we can use the channel between the known and unknown channel cases depends on how well we can measure the channel. In other words, we suspect that  $I(\underline{Y}; \underline{S} | \underline{F}) - I(\underline{Y}; \underline{S})$  will depend on how well the channel can be measured at the output solely from  $\underline{Y}$ . Indeed, in the limit where the channel can be measured perfectly, the difference  $I(\underline{Y}; \underline{S} | \underline{F}) - I(\underline{Y}; \underline{S})$  is 0, from (34).

Expression (34) has a further interpretation. We might, in general, question whether it matters that we do not know the channel, for we may perform equalization. Equalization uses the symbols already decoded to help decode the next symbols. However, equalization usually supposes that the past input can be perfectly detected from the output [55]. Such a supposition is generally thought to be reasonable, because if we did not decode the data properly anyway, we have a failure in the system. If the input is perfectly determined from the output, then the right-hand side (RHS) of (34) is nil. Hence, if we claim that we may perform equalization perfectly, we are in essence already claiming that we do not suffer, in terms of mutual information, from not knowing the channel. Therefore, in order to perform equalization, we may be sending at a lower rate than the maximum achievable rate. Note that our results differ from [40], which uses Jensen's inequality in the wrong direction.

Let us consider the case where the channel is known at the receiver within some mean-square error. We may thus consider the fundamental detrimental effect of a measurement error without restricting ourselves to a specific measuring scheme. We consider, for simplicity, that the measurement has some known error power. We first look at the single real sample case to be able to view with ease the effect of the measurement error without having to contend with matrices.

We first consider the single-symbol case. The channel model is depicted in Fig. 4. This channel model is characterized by the following random variables:

- $S$  the channel input, constrained by  $E[S^2] \leq \sigma_S^2$ ;
- $N$  AWGN with variance  $\sigma_N^2$ ;
- $Y$  the output of the channel;
- $F$  the channel strength, so that  $Y = SF + N$ .

We assume that  $S, N$ , and  $F$  are statistically independent. We break  $F$  into  $\bar{F}$  and  $\tilde{F}$ , where  $E[F] = \bar{F}$  and  $E[\tilde{F}] = 0$ . Intuitively, we view  $\bar{F}$  as the measurement of the channel and  $\tilde{F}$  as the zero-mean measurement error of the channel at the receiver, with variance  $\sigma_{\tilde{F}}^2$ .

The mutual information between the output  $Y$  and the input  $S$  is

$$I(Y; S) = h(Y) - h(Y | S) \quad (35)$$

where the entropies are differential entropies. Let us examine how we may maximize (35). Let us first consider whether the

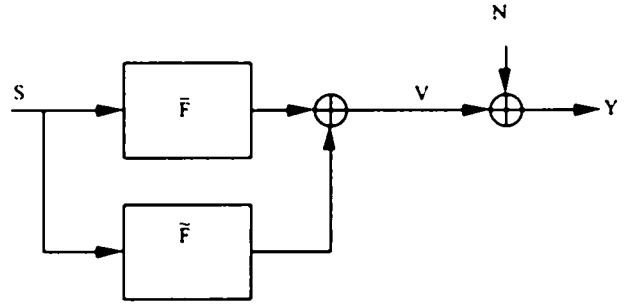


Fig. 4. System model for channel known with error.

methods that we use to find the capacity of an additive Gaussian noise channel may be applied here. Since  $Y = S\bar{F} + S\tilde{F} + N$ , for a given value of  $S$  we may write that

$$h(Y | S = s) = h(s\tilde{F} + N) \quad (36)$$

since entropy is invariant under translation. Therefore,

$$h(Y | S) = \int_{-\infty}^{+\infty} p_S(s) h(s\tilde{F} + N) ds. \quad (37)$$

The above expression cannot be calculated if we do not have a distribution for  $\tilde{F}$ . Even if we do know the distribution for  $\tilde{F}$ , whether (37) can be easily calculated depends on the form of the distribution of  $\tilde{F}$ . If the distribution is not known, then mismatch issues [6] must be considered. We now find a lower bound and an upper bound to the maximum achievable mutual information  $I(Y; S)$  for the model given in Fig. 4.

To find a lower bound on the maximum achievable  $I(Y; S) = h(S) - h(S | Y)$ , we may choose  $S$  to be Gaussian, even though the Gaussian distribution may not be the one that maximizes mutual information for the specific measurement noise distribution. Thus we fix the value of  $h(S)$ . We next find an upper bound on  $h(S | Y)$ , which holds for all possible distributions for  $S$ . The difference between  $h(S)$  and the bound on  $h(S | Y)$  immediately yields a lower bound on  $I(Y; S)$ .

By definition

$$h(S | Y) = \int h(S | Y = y) p_Y(y) dy. \quad (38)$$

Since adding a constant does not change differential entropy

$$h(S | Y = y) = h(S - \alpha y | Y = y) \quad (39)$$

thus

$$h(S | Y) = h(S - \alpha Y | Y) \quad (40)$$

for any real  $\alpha$ . Since conditioning always decreases entropy, we have that

$$h(S - \alpha Y | Y) \leq h(S - \alpha Y). \quad (41)$$

Using the fact that the entropy of a random variable with given variance is upper-bounded by the entropy of a Gaussian random variable with the same variance

$$\begin{aligned} h(S | Y) &\leq h(S - \alpha Y) \\ &\leq \frac{1}{2} \ln(2\pi e \text{Var}(S - \alpha Y)) \end{aligned} \quad (42)$$

for any  $\alpha$ . Therefore, (42) also holds when we minimize the RHS over  $\alpha$ . Hence, we pick  $\alpha$  so that  $\alpha Y$  is the linear minimum mean-square error (LMMSE) estimate of  $S$  in terms of  $Y$  [2, p. 93]

$$\alpha = \frac{E[SY]}{E[Y^2]} = \frac{\bar{F}\sigma_S^2}{\bar{F}^2\sigma_S^2 + \sigma_F^2\sigma_S^2 + \sigma_N^2} \quad (43)$$

since  $S$  and  $Y$  are zero mean. Therefore, the variance of  $S - \alpha Y$ , minimized over  $\alpha$ , is

$$\text{Var}(S - \alpha Y) = \frac{\sigma_S^4\sigma_F^2 + \sigma_N^2\sigma_S^2}{\bar{F}^2\sigma_S^2 + \sigma_F^2\sigma_S^2 + \sigma_N^2}. \quad (44)$$

Therefore, we have that

$$h(S|Y) \leq \frac{1}{2} \ln \left( 2\pi e \frac{\sigma_S^4\sigma_F^2 + \sigma_N^2\sigma_S^2}{\bar{F}^2\sigma_S^2 + \sigma_F^2\sigma_S^2 + \sigma_N^2} \right). \quad (45)$$

The mutual information between  $S$  and  $Y$  may be lower-bounded by

$$\begin{aligned} I(S; Y) &\geq \frac{1}{2} \ln(2\pi e\sigma_S^2) - \frac{1}{2} \ln \left( 2\pi e \frac{\sigma_S^4\sigma_F^2 + \sigma_N^2\sigma_S^2}{\bar{F}^2\sigma_S^2 + \sigma_F^2\sigma_S^2 + \sigma_N^2} \right) \\ &= \frac{1}{2} \ln \left( 1 + \frac{\bar{F}^2\sigma_S^2}{\sigma_F^2\sigma_S^2 + \sigma_N^2} \right). \end{aligned} \quad (46)$$

Expression (46) may be interpreted as saying that the worst effect that the measurement noise can have is to behave as AWGN. We see that the above bound is equal to the capacity of the channel that would result from sending a Gaussian signal with variance  $\bar{F}^2\sigma_S^2$  in an AWGN channel with noise variance  $\sigma_F^2\sigma_S^2 + \sigma_N^2$ . Related issues to this one are addressed in [36].

We obtain an upper bound to  $I(Y; S)$  by using the following simple consequence of (34):

$$I(S; Y) \leq I(S; Y|F). \quad (47)$$

For  $F$  known to be  $f$ , the maximum mutual information  $I(Y; S|F = f)$ , over all possible choices for  $S$ , is given by  $(1/2) \ln(1 + (f^2\sigma_S^2/\sigma_N^2))$ . Hence

$$I(S; Y|F) = \frac{1}{2} E_F \left[ \ln \left( \frac{\bar{F}^2\sigma_S^2 + \tilde{F}^2\sigma_S^2 + \sigma_N^2}{\sigma_N^2} \right) \right]. \quad (48)$$

Therefore, using the convexity  $\cap$  of the  $\ln$  function, we may write that

$$I(S; Y) \leq \ln \frac{1}{2} \left( \frac{\bar{F}^2\sigma_S^2 + \sigma_F^2\sigma_S^2 + \sigma_N^2}{\sigma_N^2} \right). \quad (49)$$

The above bound is equal to the capacity of the channel that would result from sending a Gaussian signal with variance  $\bar{F}^2\sigma_S^2 + \sigma_F^2\sigma_S^2$  in an AWGN channel with noise variance  $\sigma_N^2$ . We may interpret (49) as an upper bound given by the case where the effect of the measurement error is only that of extra transmission power. Intuitively, we may see the upper bound as the case where the measurement noise is altogether useful and the lower bound as the case where the measurement error is wholly detrimental.

We may make three remarks from (46) and (49). The first is that we can upper-bound the loss in mutual information due to not knowing the channel as

$$I(S; Y|F) - I(S; Y) \leq \frac{1}{2} \ln \left( 1 + \frac{\sigma_F^2\sigma_S^2}{\sigma_N^2} \right). \quad (50)$$

The second remark is that

$$\lim_{\sigma_F \rightarrow 0} I(Y; S) = \frac{1}{2} \ln \left( 1 + \frac{\bar{F}^2\sigma_S^2}{\sigma_N^2} \right) \quad (51)$$

i.e., the mutual information converges to the case where there is no measurement error. Such behavior was to be expected and the bounds (46) and (49) simply offer an indication as to how that convergence occurs. Finally, we may address how our bound is related to other bounds concerning mutual information. For instance, the mutual information between two possibly correlated random variables is considered in [61]. The result in [61] considers the minimization of  $I(\Phi; U)$ , where  $\Phi = U + \Psi + N$  for  $N$  zero-mean Gaussian and  $\Psi$  the jamming signal, which may depend on  $N$  and  $U$ . The result obtained in [61] is that

$$\rho_{\Psi N} = \sqrt{1 - \rho_{\Psi U}^2}.$$

Our bound concentrates on the effect of the multiplicative part of the channel, but is related to the model  $\Phi = U + \Psi + N$  if we consider that  $\Psi = (F - 1)U$ .

The extension of the previous results to the multiple-symbol case is fairly straightforward. Such an extension is useful to analyze our model as given in matrix form. Indeed, unless we consider several time intervals, we cannot hope to capture the memory inherent in multipath systems. Let us assume that we know the covariance matrix of the input  $\underline{S}$  and that the covariance matrix is not singular. We have, as in Section II, that the dimension of the input vector is  $k'$  and the dimension of the output vector is  $k$  if we use complex notation. If we consider complex random variables expressed as real vectors, the dimensions are, respectively,  $2k'$  and  $2k$ . We omit these subscripts for ease of exposition.

Let us first determine a lower bound to the achievable maximum mutual information. Assume we have a Gaussian input with mutually independent Gaussian components. Such a distribution will yield a lower bound on the maximum achievable mutual information between input and output. If  $\underline{S}_{2k'}$  is  $2k'$ -dimensional Gaussian, then ([10, p. 234])

$$h(\underline{S}) = \frac{1}{2} \ln((2\pi e)^{2k'} |\Lambda_{\underline{S}}|) \quad (52)$$

where  $||$  denotes the absolute value of the determinant of a matrix and, as in the previous section,  $\Lambda_{\underline{S}}$  is the covariance matrix of  $\underline{S}$ . We denote the cross-covariance matrix of  $\underline{S}$  and  $\underline{R}$ , say, by  $\Lambda_{(\underline{S}, \underline{R})}$ . We take the covariance matrix of  $\underline{S}$  to be given. We have that  $\underline{F}$  is a  $2k \times 2k'$  known matrix.  $\tilde{\underline{F}}$  is a zero-mean random matrix representing the measurement noise on each component of  $\underline{F}$  which is not set to 0 by our approximation of simultaneous time and frequency limitation. To find a lower bound on the maximum achievable  $I(\underline{Y}; \underline{S})$ , we proceed as in the single-use case with some modifications. We shall consider  $2k$  and  $2k'$  to

be constant in the derivation of our lower bound. Therefore, we shall omit  $2k$  and  $2k'$  as subscripts and superscripts in order to simplify the notation. We still have that, as in (42)

$$h(\underline{S} | \underline{Y}) = h(\underline{S} - \underline{\alpha} \underline{Y} | \underline{Y}) \leq \frac{1}{2} \ln((2\pi e)^{2k'} |\Lambda_{\underline{S} - \underline{\alpha} \underline{Y}}|) \quad (53)$$

for  $\underline{\alpha}$  any real  $2k' \times 2k$  matrix. Using the same reasoning as in the one-dimensional case, we wish to find that  $\underline{\alpha}$  for which  $\underline{\alpha} \underline{Y}$  is the LMMSE estimate of  $\underline{S}$  in terms of  $\underline{Y}$ . We shall derive the information form of the LMMSE error because it provides insight into the meaning of our bounds. Expression (43) becomes [2, p. 93]

$$\underline{\alpha} = \Lambda_{(\underline{S}, \underline{Y})} \Lambda_{\underline{Y}}^{-1}. \quad (54)$$

Using the fact that the signal, the noise in the estimate of the channel, and the additive Gaussian noise are all mutually independent and have zero mean

$$\underline{\alpha} = \Lambda_{(\underline{S}, \underline{F} \underline{S})} (\Lambda_{\underline{F} \underline{S}} + \Lambda_{\underline{F} \underline{S}} + \Lambda_{\underline{N}})^{-1}. \quad (55)$$

The matrix  $\Lambda_{\underline{F} \underline{S}} + \Lambda_{\underline{F} \underline{S}} + \Lambda_{\underline{N}}$  must be invertible, because  $\underline{N}$  is white.

The estimation error of the input from the output is given by  $\underline{E} = \underline{S} - \underline{\alpha} \underline{Y}$ . Thus in (53), we can replace  $\underline{S} - \underline{\alpha} \underline{Y}$  in the RHS by  $\underline{E}$ . The covariance matrix of  $\underline{E}$  is given by [2, p. 83]

$$\Lambda_{\underline{E}} = \Lambda_{\underline{S}} - \Lambda_{(\underline{S}, \underline{F} \underline{S})} (\Lambda_{\underline{F} \underline{S}} + \Lambda_{\underline{F} \underline{S}} + \Lambda_{\underline{N}})^{-1} \Lambda_{(\underline{F} \underline{S}, \underline{S})}. \quad (56)$$

The above matrix  $\Lambda_{\underline{E}}$  cannot be singular, again because  $\underline{N}$  is white. We may also write that

$$\Lambda_{\underline{E}}^{-1} = \Lambda_{\underline{S}}^{-1} + \underline{\bar{F}}^T (\Lambda_{\underline{F} \underline{S}} + \Lambda_{\underline{N}})^{-1} \underline{\bar{F}}. \quad (57)$$

For a given covariance matrix, entropy is maximized for a Gaussian distribution. Therefore, from (53)

$$\begin{aligned} h(\underline{S} | \underline{Y}) &\leq h(\underline{E}) \\ &\leq k \ln(2\pi e) + \frac{1}{2} \ln(|\Lambda_{\underline{E}}|) \quad \text{using (57)} \\ &\leq k \ln(2\pi e) - \frac{1}{2} \ln \left( \left| \Lambda_{\underline{S}}^{-1} + \underline{\bar{F}}^T (\Lambda_{\underline{F} \underline{S}} + \Lambda_{\underline{N}})^{-1} \underline{\bar{F}} \right| \right). \end{aligned} \quad (58)$$

The mutual information between  $\underline{S}$  and  $\underline{Y}$  may therefore be lower-bounded by the following vector analog to (46). We manipulate our expression so as to obtain a form which will easily be comparable to the upper bound

$$\begin{aligned} I(\underline{S}; \underline{Y}) &\geq \frac{1}{2} \ln \left( |\Lambda_{\underline{S}}| \left| \Lambda_{\underline{S}}^{-1} + \underline{\bar{F}}^T (\Lambda_{\underline{F} \underline{S}} + \Lambda_{\underline{N}})^{-1} \underline{\bar{F}} \right| \right) \\ &\quad \text{using the equality } |AB| = |A||B| \\ &= \frac{1}{2} \ln(|\underline{I} + \underline{\bar{F}}^T (\Lambda_{\underline{F} \underline{S}} + \Lambda_{\underline{N}})^{-1} \underline{\bar{F}} \Lambda_{\underline{S}}|). \end{aligned} \quad (59)$$

The interpretation of the above bound is as in the single-sample case. The RHS of (59) is the mutual information between the input and the output of a known channel  $\underline{\bar{F}}$  with independent additive Gaussian noise  $\underline{\hat{N}}$  of correlation  $\Lambda_{\underline{F} \underline{S}} + \Lambda_{\underline{N}}$ . Indeed,

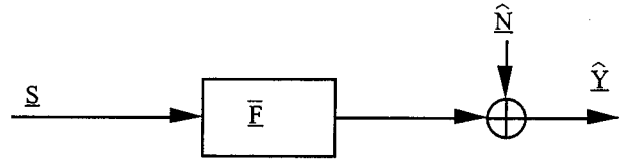


Fig. 5. Channel with mutual information between input and output equal to the RHS of (64).

let us call  $\underline{\hat{Y}}$  the output from that channel, shown in Fig. 5. Note that we may also write the mutual information between  $\underline{S}$  and  $\underline{\hat{Y}}$  as

$$\begin{aligned} I(\underline{S}; \underline{\hat{Y}}) &= h(\underline{\hat{Y}}) - h(\underline{\hat{Y}} | \underline{S}) \\ &= \frac{1}{2} \ln(|\underline{\bar{F}} \Lambda_{\underline{S}} \underline{\bar{F}}^T + \Lambda_{\underline{F} \underline{S}} + \Lambda_{\underline{N}}|) - \frac{1}{2} \ln(|\Lambda_{\underline{F} \underline{S}} + \Lambda_{\underline{N}}|) \\ &= \frac{1}{2} \ln(|(\Lambda_{\underline{F} \underline{S}} + \Lambda_{\underline{N}})^{-1} \underline{\bar{F}} \Lambda_{\underline{S}} \underline{\bar{F}}^T + \underline{I}|). \end{aligned} \quad (60)$$

Therefore, from our interpretation of (59) as a mutual information, we may rewrite the RHS as

$$I(\underline{S}; \underline{Y}) \geq \frac{1}{2} \ln(|(\Lambda_{\underline{F} \underline{S}} + \Lambda_{\underline{N}})^{-1} \underline{\bar{F}} \Lambda_{\underline{S}} \underline{\bar{F}}^T + \underline{I}|). \quad (61)$$

Obtaining an upper bound to  $I(\underline{S}; \underline{Y})$  is a straightforward extension of the single sample case. Expression (47) holds with the appropriate modifications to take into account that we are in the multidimensional case. The expression analogous to (49) is

$$\begin{aligned} I(\underline{S}; \underline{Y}) &\leq \frac{1}{2} \ln \left( \frac{|\Lambda_{(\underline{F} + \underline{\hat{F}}) \underline{S}} + \Lambda_{\underline{N}}|}{|\Lambda_{\underline{N}}|} \right) \\ &= \frac{1}{2} \ln \left( |\Lambda_{\underline{N}}^{-1} \Lambda_{(\underline{F} + \underline{\hat{F}}) \underline{S}} + \underline{I}| \right) \\ &\quad \text{rewriting } \Lambda_{(\underline{F} + \underline{\hat{F}}) \underline{S}} \\ &= \frac{1}{2} \ln \left( |\Lambda_{\underline{N}}^{-1} \underline{\bar{F}} \Lambda_{\underline{S}} \underline{\bar{F}}^T + \Lambda_{\underline{N}}^{-1} \Lambda_{\underline{F} \underline{S}} + \underline{I}| \right). \end{aligned} \quad (62)$$

We see that the above bound is equal to the capacity of the channel that would result from sending a Gaussian signal with covariance matrix  $\Lambda_{(\underline{F} + \underline{\hat{F}}) \underline{S}}$  in an AWGN channel with noise variance  $\Lambda_{\underline{N}}$ . Therefore, as in the single-dimensional case, we may interpret the upper bound (62) as the case where the only effect of the measurement error is that of extra transmission power.

In the limit as  $\Lambda_{\underline{F} \underline{S}} \rightarrow 0$ , from (61) and (62), we have that

$$I(\underline{S}; \underline{Y}) \rightarrow \frac{1}{2} \ln \left( \left| \Lambda_{\underline{N}}^{-1} \underline{\bar{F}} \Lambda_{\underline{S}} \underline{\bar{F}}^T + \underline{I} \right| \right). \quad (63)$$

Note that the condition  $\Lambda_{\underline{F} \underline{S}} \rightarrow 0$  may seem to be a fairly strong condition. However, an equivalent condition is that  $\text{tr}(\Lambda_{\underline{F} \underline{S}}) \rightarrow 0$ . Since  $\Lambda_{\underline{F} \underline{S}}$  is positive-semidefinite, requiring that  $\text{tr}(\Lambda_{\underline{F} \underline{S}}) \rightarrow 0$  is equivalent to requiring that each individual term in the diagonal of  $\Lambda_{\underline{F} \underline{S}}$  go to 0. Indeed, if the diagonal terms of  $\Lambda_{\underline{F} \underline{S}}$  are 0 then, since  $\Lambda_{\underline{F} \underline{S}}$  is a covariance matrix, the offdiagonal terms must go to 0 also. In the following section, we shall use these results for channels with a measurement error to determine some general properties regarding the



loss of mutual information from not knowing the channel at the receiver.

### B. Multiple-Access Case

In this section, we seek to extend the results concerning channel measurement uncertainty to the multiple-user case and consider the effect of spreading when we have several users. The use of broadband signals under multipath channel conditions is often advocated because of the possibility of resolving the multiple paths at the receiver and then combining them in some approximately optimal way ([39]), such as by a rake receiver ([52]). When we deal with multiple-access situations, such as the situation of mobile communications, where the users share bandwidth under some spread-spectrum protocol, the issue of spreading is further complicated by multiple-user interference. The most commonly proposed approach ([54]) is to decode each user as if all other users were noise. The advantage of spreading under these conditions is to reduce the power spectral density of the users, when the users are all under a maximum power constraint, so that the system becomes limited by the additive Gaussian noise in the channel rather than being interference-limited. Such an approach is, from an information-theoretic point of view, quite suboptimal (unless we use FDMA [48]) when we can measure the channels for the various users ([60]). We, therefore, wish to show that interference cancellation is still possible when we do not know the channel as long as we may measure the channel for each user. We know that under certain ergodic conditions, if decoding a user is possible, then canceling the interference from that user is possible with a small added error due to the channel uncertainty. However, we consider explicitly the loss in mutual information without any assumption about decoding individual users. Our model is somewhat different from that in [19] in that we divide the channel  $F$  into a known component  $\bar{F}$  and a fluctuation  $\tilde{F}$  without considering the relation between  $\bar{F}$  and the sample observation. We bound the effect of channel estimation error upon interference cancellation.

The inequalities we have found for the single-user case hold for every inequality which defines that feasible rate region. We shall show the effect of knowing the channels of the users with a certain error of given variance. We shall give our examples for two users, since we may give a graphical interpretation of our bounds. The signals of the users and of the channels are assumed to be mutually independent. To obtain upper bounds, we consider that all received signal components other than the AWGN are part of the input signal. Thus we may write that

$$\begin{aligned}
I(Y; S_1 | S_2) &\leq I(Y; S_1 | S_2, F_1, F_2) \\
&= h(Y | S_2, F_1, F_2) - h(Y | S_1, S_2, F_1, F_2) \\
&= h(Y | S_2, F_1, F_2) - h(N) \\
&\quad \text{since a Gaussian distribution maximizes} \\
&\quad \text{entropy for a given variance} \\
&\leq \frac{1}{2} E_{F_1} [\ln(\sigma_N^2 + \sigma_{S_1}^2 F_1^2)] - \frac{1}{2} \ln(\sigma_N^2) \\
&\quad \text{since the } \ln \text{ function is convex } \cap \\
&\leq \frac{1}{2} \ln \left( 1 + \frac{\bar{F}_1^2 \sigma_{S_1}^2 + \sigma_{S_1}^2 \sigma_{F_1}^2}{\sigma_N^2} \right). \quad (64)
\end{aligned}$$

A similar inequality holds for  $I(Y; S_2 | S_1)$ . A bound for  $I(Y; (S_1, S_2))$  may be obtained by considering the sum of the signals of the two users as being a single signal. Thus we obtain the following upper bounds on the region of feasible mutual informations for the single symbol case with real inputs and real channels (the subscripts denote users):

$$I(Y; S_1 | S_2) \leq \frac{1}{2} \ln \left( 1 + \frac{\bar{F}_1^2 \sigma_{S_1}^2 + \sigma_{S_1}^2 \sigma_{F_1}^2}{\sigma_N^2} \right) \quad (65)$$

$$I(Y; S_2 | S_1) \leq \frac{1}{2} \ln \left( 1 + \frac{\bar{F}_2^2 \sigma_{S_2}^2 + \sigma_{S_2}^2 \sigma_{F_2}^2}{\sigma_N^2} \right) \quad (66)$$

$$\begin{aligned}
I(Y; (S_1, S_2)) \\
\leq \frac{1}{2} \ln \left( 1 + \frac{\bar{F}_2^2 \sigma_{S_2}^2 + \sigma_{S_2}^2 \sigma_{F_2}^2 + \bar{F}_1^2 \sigma_{S_1}^2 + \sigma_{S_1}^2 \sigma_{F_1}^2}{\sigma_N^2} \right). \quad (67)
\end{aligned}$$

Each bound of (65)–(67) may be interpreted as in the single-user case. We have that

$$\begin{aligned}
I(Y; S_1 | S_2) &= h(S_1 | S_2) - h(S_1 | Y, S_2) \\
&\quad \text{since } S_1 \text{ and } S_2 \text{ are independent} \\
&= h(S_1) - h(S_1 | Y, S_2) \\
&= h(S_1) - h(S_1 | (Y - \bar{F}_2 S_2), S_2) \\
&\quad \text{since conditioning reduces entropy} \\
&\geq h(S_1) - h(S_1 | (Y - \bar{F}_2 S_2)). \quad (68)
\end{aligned}$$

We may use the LMMSE estimate of  $S_1$  from  $Y - \bar{F}_2 S_2$  to obtain

$$I(Y; S_1 | S_2) \geq \frac{1}{2} \ln \left( 1 + \frac{\bar{F}_1^2 \sigma_{S_1}^2}{\sigma_N^2 + \sigma_{S_1}^2 \sigma_{F_1}^2 + \sigma_{S_2}^2 \sigma_{F_2}^2} \right) \quad (69)$$

from (68). We may obtain in a similar manner the following lower bounds to the region of achievable rates:

$$I(Y; S_2 | S_1) \geq \frac{1}{2} \ln \left( 1 + \frac{\bar{F}_2^2 \sigma_{S_2}^2}{\sigma_N^2 + \sigma_{S_2}^2 \sigma_{F_2}^2 + \sigma_{S_1}^2 \sigma_{F_1}^2} \right) \quad (70)$$

$$I(Y; (S_1, S_2)) \geq \frac{1}{2} \ln \left( 1 + \frac{\bar{F}_2^2 \sigma_{S_2}^2 + \bar{F}_1^2 \sigma_{S_1}^2}{\sigma_N^2 + \sigma_{S_2}^2 \sigma_{F_2}^2 + \sigma_{S_1}^2 \sigma_{F_1}^2} \right). \quad (71)$$

Again, the bounds in (69)–(71) bear a similar interpretation to the one we gave for the single-user case. The effect of the channel uncertainty can be no worse than to contribute to the AWGN. Each user must contend not only with the uncertainty of its own channel but also with the uncertainty of the other users' channels. In particular, when we consider the mutual information between user 1 and the output conditioned on user 2, we must take into account the fact that there is a residual effect upon user 1 of the uncertainty on the channel of user 2. Indeed, when we perform interference cancellation, the fact that we are not able to cancel a user exactly because of error in its channel measurement means that there is extra noise beyond the AWGN. Expressions (70) and (71) may be obtained similarly.

Using the lower bounds of (69)–(71), we may obtain a lower bound to the feasible rate region. The effect of the channel error may be bounded as the effect of extra AWGN on the channel. Note that the corners of the feasible rate regions correspond to the rates that are achievable when the users transmit white sig-

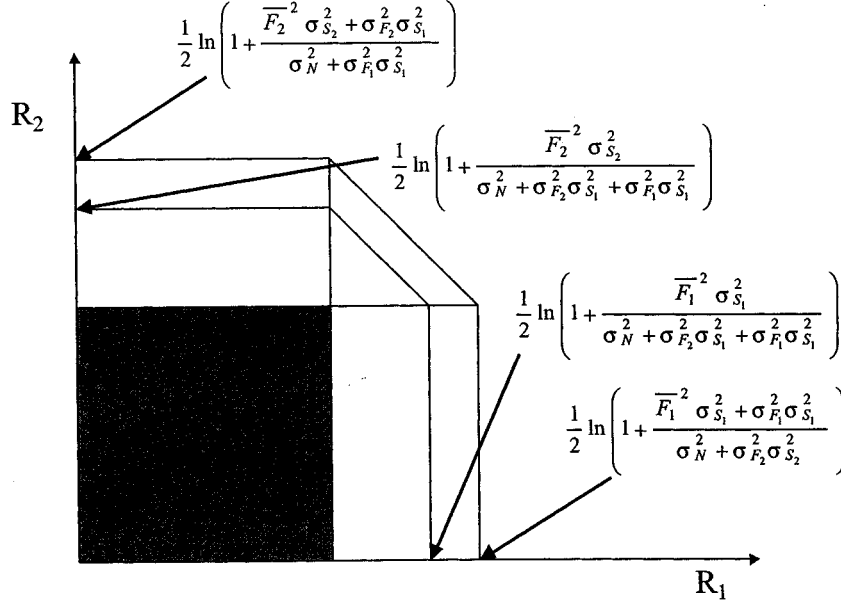


Fig. 6. Lower bound to the feasible rate region for channels known with a given error variance.

nals over the same bandwidth at the same time and the receiver performs interference cancellation. Fig. 6 shows our upper and lower bounds for  $R_1$  and  $R_2$ , the feasible rates for users 1 and 2, respectively. The gray square represents the mutual information region when we consider each user as noise to the other user, i.e., where we have an interference channel, for the channel model corresponding to the upper bound. For our example, we see that it is preferable to perform interference cancellation rather than to treat each user as noise to the other user.

For the multiple-symbol case, the bounds extend in the same way as for the single-symbol case. Let us simply write the lower bounds

$$I(\underline{S}_1; \underline{Y} | \underline{S}_2) \geq \frac{1}{2} \ln \left( \left( \Lambda_{\underline{F}_1 \underline{S}_1} + \Lambda_{\underline{F}_2 \underline{S}_2} + \Lambda_N \right)^{-1} \overline{F}_1 \Lambda_{\underline{S}_1} \overline{F}_1^T + \underline{I} \right) \quad (72)$$

$$I(\underline{S}_2; \underline{Y} | \underline{S}_1) \geq \frac{1}{2} \ln \left( \left( \Lambda_{\underline{F}_1 \underline{S}_1} + \Lambda_{\underline{F}_2 \underline{S}_2} + \Lambda_N \right)^{-1} \overline{F}_2 \Lambda_{\underline{S}_2} \overline{F}_2^T + \underline{I} \right) \quad (73)$$

$$I((\underline{S}_1, \underline{S}_2); \underline{Y}) \geq \frac{1}{2} \ln \left( \left( \Lambda_{\underline{F}_1 \underline{S}_1} + \Lambda_{\underline{F}_2 \underline{S}_2} + \Lambda_N \right)^{-1} \times \left( \overline{F}_2 \Lambda_{\underline{S}_2} \overline{F}_2^T + \overline{F}_1 \Lambda_{\underline{S}_1} \overline{F}_1^T \right) + \underline{I} \right). \quad (74)$$

Similarly, we may give the following upper bounds for the multiple-symbol case:

$$I(\underline{S}_1; \underline{Y} | \underline{S}_2) \leq \frac{1}{2} \ln \left( \Lambda_N^{-1} \left( \Lambda_{\underline{F}_1 \underline{S}_1} + \overline{F}_1 \Lambda_{\underline{S}_1} \overline{F}_1^T \right) + \underline{I} \right) \quad (75)$$

$$I(\underline{S}_2; \underline{Y} | \underline{S}_1) \leq \frac{1}{2} \ln \left( \Lambda_N^{-1} \left( \Lambda_{\underline{F}_2 \underline{S}_2} + \overline{F}_2 \Lambda_{\underline{S}_2} \overline{F}_2^T \right) + \underline{I} \right) \quad (76)$$

$$I((\underline{S}_1, \underline{S}_2); \underline{Y}) \leq \frac{1}{2} \ln \left( \Lambda_N^{-1} \left( \Lambda_{\underline{F}_1 \underline{S}_1} + \Lambda_{\underline{F}_2 \underline{S}_2} + \overline{F}_2 \Lambda_{\underline{S}_2} \overline{F}_2^T + \overline{F}_1 \Lambda_{\underline{S}_1} \overline{F}_1^T \right) + \underline{I} \right). \quad (77)$$

We do not give a graphical representation of the multiple symbol case, since such an extension would be difficult and would not improve our intuition.

#### IV. BOUNDS APPLIED TO RECURSIVE ESTIMATION

In the previous section, we have considered that we know the variance of the error on the channel estimate, but have not discussed how we determine such an estimate. The estimate of the channel will depend on the type of channel we consider and how it is measured. In this section, we illustrate the use of the bounds found in the previous section for the case of a single-user AWGN channel with a single tap which evolves in a Gauss–Markov fashion. The model for our channel is

$$F[i, i] = \alpha F[i-1, i-1] + \Xi[i-1, i-1] \quad (78)$$

where the  $\Xi$ 's are zero-mean mutually independent Gaussian random variables which are independent of the noise. The random variable  $\Xi$  is commonly known as the innovation term. We can relate the form of (78) to the physical characteristics of a wireless channel. In particular, the constant  $\alpha$  is related to the coherence time  $T_c$  of the channel.  $\alpha$  determines how fast the channel decorrelates, with

$$\alpha = \frac{E[F[i-1, i-1], F[i, i]]}{E[F[i, i]]}.$$

We select *a priori* the level of decorrelation  $\phi$ , which we deem sufficient to indicate that the channel has decorrelated in  $m$  time samples, i.e.,

$$\phi = \frac{E[F[i-m, i-m], F[i, i]]}{E[F[i, i]]}.$$

Thus  $T_c = m/W$ . From (78), we may write

$$\alpha^{T_c W} = \phi. \quad (79)$$

There are many ways to select  $\phi$ . In general, depending on how we define the coherence time, we have  $T_c$  roughly inversely proportional to  $B_D$ , with a proportionality constant which depends

on the level of decorrelation we deem necessary in our definition of coherence time. In the literature, the correlation coefficient is taken to vary from 0.9 ([11]) to 0.37 ([5]) for a time separation of  $T_c$ .

We sound the channel using a probe signal. The probe signal occupies its own band, the probe band, and does not overlap in frequency with the payload signal  $Y$ . The probe signal characterizes the single tap, as long as the probe signal is in the same coherence bandwidth as the payload. Note that a single tap model is well-suited to the case of a large coherence bandwidth, which corresponds to a small time spread, hence to the case of unresolved paths whose aggregate effect is represented by a single tap. If we had more resolved paths, then we would have ISI and the multiple-symbol case would be applicable. Note that our model does have memory, since there is memory in the tap. The probe signal is  $u$  and the received version of the probe signal is  $V$ . So as not to devote too large a bandwidth to the probe signal, the probe signal may be sent over a narrow probe band and at a lower rate than  $S$ . Therefore, the sampling of the probe signal will be at a lower rate  $W_p$  than the sampling rate  $W$  for the transmission payload  $Y$ . Let us consider that the sampling of the probe signal is  $\nu$  times slower than the sampling of  $S$ , i.e.,  $W_p\nu = W$ . For simplicity, we take  $\nu$  to be an integer. In the sampling, we have

$$F[i, i] = \beta F[i-1, i-1] + \Xi'[i-1, i-1] \quad (80)$$

where  $\beta = \alpha^\nu$ . Moreover,  $\sigma_{\Xi'}^2 = (1 - \beta^2)\sigma_F^2$ . We estimate the channel  $F[i, i]$  sampled at rate  $W$  by the sample  $F[[i/\nu]]$ .

Until we state otherwise, let us consider sampling at rate  $W_p$  in the probe band. Since we have a Gauss–Markov channel model, the maximum-likelihood (ML) estimate of the channel  $F[i]$  from  $\underline{V}_i$  and  $\underline{u}_i$  is obtained by using a Kalman filter [2]. The Kalman filter estimate in this case is also the MMSE and the LMMSE estimate. We denote the variance of the error of the estimate of  $F[i]$  by  $\lambda_i$ . The recursive expression for  $\lambda_i$  is

$$\begin{aligned} \frac{1}{\lambda_1} &= \frac{1}{\sigma_F^2 + \sigma_\xi^2} + \frac{u[1]^2}{\sigma_N^2} \\ \frac{1}{\lambda_{j+1}} &= \frac{1}{\alpha^2 \lambda_j + \sigma_\xi^2} + \frac{u[j+1]^2}{\sigma_N^2}. \end{aligned} \quad (81)$$

Note that, although (81) does not depend on the values  $\underline{V}_j$ , it does depend on the values  $\underline{u}_j$ . Let us suppose that we have an average energy constraint on  $u$ , so that the sample average value of  $u[i]^2$  over all samples must be no greater than  $\sigma_U^2$ . Therefore, the probe need not be constant. For instance, the probe may be sent at regular intervals, or may be transmitted at random times; for instance, alongside some training sequences for synchronization. If the probe is a pilot tone, then it will be constant. We do not address here the manner in which to allocate resources to the probe ([30]). It is useful for us to determine in what way our probing should be performed so as to satisfy our constraint while reducing the variance of the error as much as possible. To this end, let us define  $\lambda'_j$  by the recursive expression

$$\begin{aligned} \frac{1}{\lambda'_1} &= \frac{1}{\sigma_N^2 + \sigma_{\Xi'}^2} + \frac{\sigma_U^2}{\sigma_N^2} \\ \frac{1}{\lambda'_{j+1}} &= \frac{1}{\beta^2 \lambda'_j + \sigma_{\Xi'}^2} + \frac{\sigma_U^2}{\sigma_N^2}. \end{aligned} \quad (82)$$

The term  $\lambda'_{j+1}$  is the variance of the estimate when each probe signal  $u[i]$  is equal to  $\sigma_U$ . Let us suppose that we have some arbitrary ergodic distribution, so that the probe signal is a random variable  $U$  whose expected value satisfies the energy constraint. We shall use the following result, whose proof is relegated to the Appendix. For any  $\gamma > 0$

$$\ln(1 + \gamma \lambda'_j) \leq E_{\underline{U}_j}[\ln(1 + \gamma \lambda_j)]. \quad (83)$$

The limit as  $i \rightarrow \infty$  of  $\lambda'_i$  is  $\lambda'$

$$\begin{aligned} \lambda' &= \frac{\sigma_N^2 \beta^2 - \sigma_N^2 - \sigma_U^2 \sigma_{\Xi'}^2}{2\sigma_U^2 \beta^2} \\ &+ \frac{\sqrt{(\sigma_N^2 \beta^2 - \sigma_N^2 - \sigma_U^2 \sigma_{\Xi'}^2)^2 + 4\sigma_U^2 \sigma_{\Xi'}^2 \sigma_N^2 \beta^2}}{2\sigma_U^2 \beta^2}. \end{aligned} \quad (84)$$

Note that  $\lambda' \rightarrow 0$  as  $\beta \rightarrow 1$  or, equivalently, as  $\alpha \rightarrow 1$ . Thus  $\lambda'$  goes to 0 as the channel becomes time-invariant, because we can then measure the channel arbitrarily well. Also,  $\lambda' \rightarrow 0$  as  $\sigma_U^2 \rightarrow \infty$ . We may interpret this limit as the fact that, if our probe signal is sufficiently strong, we can measure the channel with almost no error.

Let us consider that the transmitted signal  $S$  is stationary and ergodic, with variance  $\sigma_S^2$  per symbol. The mutual information we consider is between the vector of samples, sampled at rate  $W$ ,  $\underline{S}_k$ , and the pair  $(\underline{Y}_k, \underline{V}_{\lfloor k/\nu \rfloor})$ . Let us consider the mutual information over  $k$  samples

$$I(\underline{S}_k; (\underline{Y}_k, \underline{V}_{\lfloor k/\nu \rfloor})) = \sum_{i \leq k} I(S[i]; (\underline{Y}_k, \underline{V}_{\lfloor k/\nu \rfloor}) | \underline{S}_{i-1}) \quad (85)$$

$$\geq \sum_{i \leq k} I(S[i]; (\underline{Y}_k, \underline{V}_{\lfloor k/\nu \rfloor}) | \underline{S}_{i-1}) \quad (86)$$

$$\begin{aligned} &= \sum_{i \leq k} I(S[i]; \underline{V}_{\lfloor k/\nu \rfloor} | \underline{S}_{i-1}) + \sum_{i \leq k} \sum_{j \leq k} \\ &\quad \cdot I(S[i]; Y[j] | \underline{S}_{i-1}, \underline{Y}_{j-1}, \underline{V}_{\lfloor k/\nu \rfloor}) \end{aligned} \quad (87)$$

$$\geq \sum_{i \leq k} I(S[i]; Y[i] | \underline{S}_{i-1}, \underline{Y}_{i-1}, \underline{V}_{\lfloor k/\nu \rfloor}) \quad (88)$$

where (85) and (87) follow from the chain rule for mutual information. We obtain (86) by using the fact that

$$I(X; (Y, Z)) \geq I(X; Y).$$

We are interested in upper-bounding the loss in average maximum mutual information due to estimating the channel via our out-of-band probe rather than knowing the channel *a priori*. Note first that

$$I(\underline{S}_k; (\underline{Y}_k, \underline{V}_{\lfloor k/\nu \rfloor}) | \underline{E}_k) = I(\underline{S}_k; \underline{Y}_k | \underline{E}_k)$$

since  $\underline{V}_{\lfloor k/\nu \rfloor}$  is superfluous when the channel is known. Thus we seek to upper bound

$$\Delta_k = \max_{\underline{S}_k} [I(\underline{S}_k; \underline{Y}_k | \underline{E}_k)] - \max_{\underline{S}_k} \left[ I(\underline{S}_k; (\underline{Y}_k, \underline{V}_{\lfloor k/\nu \rfloor})) \right]. \quad (89)$$

Let us consider the first term of the RHS of (89)

$$\begin{aligned} I(\underline{S}_k; \underline{Y}_k | \underline{E}_k) &= h(\underline{Y}_k | \underline{E}_k) - h(\underline{Y}_k | \underline{E}_k, \underline{S}_k) \\ &\leq \sum_{i \leq k} h(Y[i] | F[i]) - h(N[i]) \end{aligned} \quad (90)$$

where (90) is derived from the chain rule for entropies and the fact that conditioning reduces entropy.

From (88) and (90), we have

$$\begin{aligned} \Delta_k \leq & \max_{\underline{S}_k} \left[ \sum_{i \leq k} I(Y[i]; S[i] | F[i]) \right] \\ & - \max_{\underline{S}_k} \left[ \sum_{i \leq k} I(S[i]; Y[j] | \underline{S}_{i-1}, \underline{Y}_{i-1}, \underline{V}_{\lfloor \frac{i}{\nu} \rfloor}) \right]. \end{aligned} \quad (91)$$

We consider for the second term in the sum of the RHS of (91) that we estimate the channel through  $\underline{V}_{\lfloor (i/\nu) \rfloor}$  only and that we do not use  $\underline{S}_{i-1}, \underline{Y}_{i-1}$  in our estimation of the channel. This assumption gives an upper bound on the variance of the channel estimate at sample  $i$ . Since up to  $\nu$  time samples at rate  $W$  can elapse between the time we obtain  $\underline{V}_{\lfloor (i/\nu) \rfloor}$  and the occurrence of  $F[i]$ , we can upper-bound the variance of the estimate of  $F$  from  $\underline{V}_{\lfloor (i/\nu) \rfloor}$  by the variance of the estimate of the channel from  $\underline{V}_{\lfloor (i/\nu) \rfloor}$  plus  $\nu\sigma_{\Xi}^2$ . Let us denote by  $\hat{F}[i]$  the estimate of  $F[i]$  obtained from  $\underline{V}_{\lfloor (i/\nu) \rfloor}$ . Let us denote by  $\tilde{F}[i]$  the error on the estimate  $\hat{F}[i]$ . This error has variance  $\sigma_{\tilde{F}[i]}^2 \leq \sigma_{\lambda_{\lfloor (i/\nu) \rfloor}}^2 + \nu\sigma_{\Xi}^2$ . From (81), we know that  $\lambda_{\lfloor (i/\nu) \rfloor}$  depends only on  $\underline{U}_i$ . Hence, from (50)

$$\Delta_k \leq \sum_{i \leq k} \frac{1}{2} E_{\underline{U}_i} \left[ \ln \left( 1 + \frac{\sigma_S^2 (\lambda_{\lfloor \frac{i}{\nu} \rfloor} + \nu\sigma_{\Xi}^2)}{\sigma_N^2} \right) \right] \quad (92)$$

$$\leq \sum_{i \leq k} \frac{1}{2} \ln \left( 1 + \frac{\sigma_S^2 (\lambda'_{\lfloor \frac{i}{\nu} \rfloor} + \nu\sigma_{\Xi}^2)}{\sigma_N^2} \right) \quad (93)$$

where (93) follows from (83). Thus the limit of the average loss in terms of maximum mutual information per symbol due dealing with an estimate of the channel rather than knowing the channel perfectly may be upper-bounded as

$$\lim_{k \rightarrow \infty} \frac{1}{k} \Delta_k \rightarrow \frac{1}{2} \ln \left( 1 + \frac{\sigma_S^2 (\lambda' + \nu\sigma_{\Xi}^2)}{\sigma_N^2} \right) \quad (94)$$

by using (84). Using our remarks following (84), we know that  $\lambda' \rightarrow 0$  as  $\alpha \rightarrow 1$  or as  $\sigma_U^2 \rightarrow \infty$ .

## V. CONCLUSIONS AND DIRECTIONS FOR FURTHER RESEARCH

We have investigated the effect of uncertainty in the knowledge of the channel at the receiver for a single-user time-varying channel. We have established that, in the case where the channel is known at the sender and the receiver, the mutual information can be found for a channel which induces spreading in both time and frequency. Therefore, the real difficulty in establishing capacity results stems from our imperfect knowledge of the channel rather than its time-varying nature, although it is the time variations that render estimation difficult. In the case where we know the covariance of the measurement noise, we have established upper and lower bounds which are independent of the channel and measurement distribution. These bounds are tight as the measurement error vanishes. For the multiple-access case, we have shown that, even in the presence of channel uncertainty, it may be advantageous to perform multiple-user interference cancellation rather than to consider all users to be noise

to all other users. We have applied our results to obtain explicit bounds to the loss of mutual information due to not knowing the channel when the taps may be described by a Gauss–Markov process.

An interesting question is whether incoherent interference cancellation is advantageous. Indeed, many spread-spectrum systems currently have an incoherent rake receiver for each user and detect all users separately. Since interference cancellation is computationally fairly simple (its time complexity is linear in the number of users), it might be of practical interest to establish when, if ever, the uncertainty in the channel due to not knowing the phase component is such that interference cancellation is still better than considering the multiple-access channel to be an interference channel.

## APPENDIX

### A. Proof of (83)

Let  $\gamma > 0$ . Define

$$\begin{aligned} z[i] &= \gamma^{-1} \left( \frac{z[i-1]}{\gamma^{-1}\beta^2 + \sigma_{\Xi}^2 z[i-1]} + \frac{u[i]^2}{\sigma_N^2} \right), \quad \text{for } i \geq 1 \\ z[0] &= \gamma^{-1} \frac{\beta^2}{\sigma_F^2}. \end{aligned} \quad (95)$$

For  $i \geq 1$ ,  $z[i]$  is  $\frac{1}{\gamma\lambda_i}$ . We define  $f_\gamma$  to be

$$f_\gamma(z) = \frac{z}{\gamma^{-1}\beta^2 + \sigma_{\Xi}^2 z}.$$

The function  $f$  is monotonically increasing and its second derivative is negative. Thus  $f$  is convex  $\cap$ . We have that

$$z[i] = \gamma^{-1} \left( f(z[i-1]) + \frac{u[i]^2}{\sigma_N^2} \right). \quad (96)$$

Define

$$\begin{aligned} z'[i] &= \gamma^{-1} \left( f(z[i-1]) + \frac{\sigma_U^2}{\sigma_N^2} \right), \quad \text{for } i \geq 1 \\ z'[0] &= \gamma^{-1} \left( \frac{\beta^2}{\sigma_F^2} \right). \end{aligned} \quad (97)$$

For  $i \geq 1$ ,  $z'[i]$  is  $\frac{1}{\gamma\lambda'_i}$ . Let us proceed by induction. We may verify that

$$E_{U_{[1]}}[z[1]] = z'[1].$$

Let us now assume that

$$E_{U_{i-1}}[z[i-1]] \leq z'[i-1].$$

Then, using the fact that  $f$  is convex  $\cap$ , we have

$$\begin{aligned} E_{U_i}[z[i]] &= E_{U_{i-1}}[\gamma^{-1} f(z[i-1])] + \gamma^{-1} \frac{\sigma_U^2}{\sigma_N^2} \\ &\leq \gamma^{-1} f(E_{U_{i-1}}[z[i-1]]) + \gamma^{-1} \frac{\sigma_U^2}{\sigma_N^2} \end{aligned} \quad (98)$$

$$\leq \gamma^{-1} \left( f(z'[i-1]) + \frac{\sigma_U^2}{\sigma_N^2} \right) \quad (99)$$

where (99) follows from our induction assumption and the fact that  $f$  is monotonically increasing. Hence, for all  $i \geq 1$ ,  $E_{U_i}[z[i]] \leq z'[i]$ . Thus it follows that

$$\frac{1}{\gamma\lambda'_j} \geq E_{U_j} \left[ \frac{1}{\gamma\lambda_j} \right].$$

Let us consider the function  $g$  defined by  $g(x) = 1/(1+(1/x))$ . For  $x > 0$ , we have that

$$g'(x) = \frac{1}{x^2(1+\frac{1}{x})^2} > 0$$

and

$$g''(x) = \frac{-2}{x^3(1+\frac{1}{x})^3} < 0.$$

Thus  $g$  is increasing and convex  $\cap$  for our region of interest. Hence

$$g(z'[i]) \geq g(E_{\underline{U}_i}[z[i]]) \geq E_{\underline{U}_i}[g(z[i])]. \quad (100)$$

Note that

$$g(z'[i]) = \frac{1}{1+\gamma\lambda'_i}$$

and

$$g(z[i]) = \frac{1}{1+\gamma\lambda_i}.$$

Thus (100) implies that

$$\ln\left(\frac{1}{1+\gamma\lambda'_i}\right) \geq E_{\underline{U}_i}\left[\ln\left(\frac{1}{1+\gamma\lambda_i}\right)\right].$$

Hence,  $\ln(1+\gamma\lambda'_i) \leq E_{\underline{U}_i}[\ln(1+\gamma\lambda_i)]$ .

#### ACKNOWLEDGMENT

The author wishes to thank R. G. Gallager for his extensive help and guidance throughout the thesis on which the bulk of this paper is based, as well as substantial comments in improving the paper. M. Trott and J. Shapiro provided helpful discussions and guidance. Finally, the author would like to acknowledge the help of S. Shamai and that of the anonymous reviewers.

#### REFERENCES

- [1] R. Ahlswede, "Multi-way communication channels," in *2nd Int. Symp. Information Theory*, 1971, pp. 23–52.
- [2] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice Hall, 1979.
- [3] S. Barbarossa and A. Scaglione, "On the capacity of linear time-varying channels," presented at the 1999 International Conference on Acoustics, Speech, and Signal Processing (ICASSP99).
- [4] P. Bello, "Characterization of randomly time-invariant channels," *IEEE Trans. Commun. Syst.*, vol. CS-11, pp. 361–393, Dec. 1963.
- [5] P. A. Bello and B. D. Nelin, "The effect of frequency selective fading on the binary error probabilities of incoherent and differentially coherent matched filter receivers," *IEEE Trans. Commun. Syst.*, vol. CS-11, no. 2, pp. 170–186, June 1963.
- [6] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.
- [7] R. S. Cheng and S. Verdú, "Gaussian multiaccess channels with capacity region and multiuser water-filling," *IEEE Trans. Inform. Theory*, vol. 39, pp. 773–785, May 1993.
- [8] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 648–657, Nov. 1980.
- [9] T. M. Cover, R. J. McEliece, and E. C. Posner, "Asynchronous multiple-access channel capacity," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 409–413, July 1981.
- [10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [11] D. C. Cox and R. P. Leck, "Correlation bandwidth and delay spread multipath propagation statistics for 910 MHz urban mobile radio," *IEEE Trans. Commun.*, vol. COM-23, pp. 1271–1280, Nov. 1975.
- [12] I. Csiszár and P. Narayan, "The capacity of the arbitrary varying channel," *IEEE Trans. Inform. Theory*, vol. 37, no. 1, pp. 18–26, Jan. 1991.
- [13] S. N. Diggavi, "Analysis of multicarrier transmission in time-varying channels," in *ICC'97*, Montreal, Que., Canada, pp. 1191–1195.
- [14] R. G. Gallager, "Characterization and measurement of time- and frequency-spread channels," MIT Lincoln Lab., Lexington, MA, Tech. Rep. 352, Apr. 30, 1964.
- [15] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [16] —, "A perspective on multiaccess channels," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 124–142, Mar. 1985.
- [17] —, "Multi access information theory for fading multipath channels," in *1996 Information Theory Workshop*, Haifa, Israel, p. 38.
- [18] —, "Residual noise after interference cancellation on fading multipath channels," in *Communications, Computation, Control, and Signal Processing*. Boston, MA: Kluwer, 1997, pp. 67–77.
- [19] —, "An inequality on the capacity region of multiaccess fading channels," in *Communications and Cryptography—Two Sides of One Tapestry*. Boston, MA: Kluwer, 1994.
- [20] A. El Gamal and T. M. Cover, "Multiple user information theory," *Proc. IEEE*, vol. 68, pp. 1466–1483, Dec. 1980.
- [21] M. Effros and A. Goldsmith, "Capacity definitions and coding strategies for general channels with receiver side information," in *Proc. 1998 Int. Symp. Information Theory*, Cambridge, MA, Aug. 16–21, 1998, p. 39.
- [22] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1986–1992, Nov. 1997.
- [23] A. Grant, B. Rimoldi, R. Urbanke, and P. Whiting, "Single user coding for the discrete memoryless multiple access channel," in *IEEE Int. Symp. Information Theory*, 1995, p. 448.
- [24] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 725–729, Nov. 1972.
- [25] D. Greenwood and L. Hanzo, "Characterization of mobile radio channels," in *Mobile Radio Communications*, R. Steele, Ed. Piscataway, NJ: IEEE Press, 1992.
- [26] T. S. Han, "An information-spectrum approach to capacity theorems for the general multiple-access channel," *IEEE Trans. Inform. Theory*, submitted for publication.
- [27] S. V. Hanly, "Information capacity of radio networks," Ph.D. dissertation, Univ. Cambridge, Cambridge, U.K., 1993.
- [28] W. Hirt and J. L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. Inform. Theory*, vol. 34, pp. 380–388, May 1988.
- [29] W. C. Jakes, Ed., *Microwave Mobile Communications*. Piscataway, NJ: IEEE Press, 1993.
- [30] S. Jayaraman and H. Viswanathan, "Simultaneous communication and detection over compound channels," in *1998 Int. Symp. Information Theory*, Cambridge, MA, Aug. 16–21, p. 40.
- [31] T. T. Kadota and A. D. Wyner, "Coding theorem for stationary, asymptotically memoryless, continuous-time channels," *Ann. Math. Statist.*, vol. 43, no. 5, 1972.
- [32] V. Kafedziski, "Capacity of frequency selective fading channels," in *1997 Int. Symp. Information Theory*, Ulm, Germany, p. 339.
- [33] T. Kailath, "Sampling Models for Linear Time-Variant Filters," MIT RLE, Cambridge, MA, Tech. Rep. 352, May 25, 1959.
- [34] R. S. Kennedy, *Fading Dispersive Communications Channels*. New York: Wiley-Interscience, 1969.
- [35] R. Knopp and P. A. Humblet, "Multiple-accessing over frequency-selective fading channels," in *IEEE Symp. Personal, Indoor and Mobile Radio Communications*, 1995, pp. 1326–1330.
- [36] A. Lapidoto and S. Shamai, "Fading channels: How perfect need 'Perfect side information' be," in *1999 IEEE Information Theory Workshop*, Kruger National Park, South Africa, June 20–25, 1999, pp. 36–39.
- [37] E. A. Lee and D. G. Messerschmitt, *Digital Communication*. Boston, MA: Kluwer, 1992.
- [38] W. C. Y. Lee, *Mobile Communications Engineering*. New York: McGraw-Hill, 1993.
- [39] J. S. Lehnert and M. B. Pursley, "Multipath diversity reception of spread-spectrum multiple-access communications," *IEEE Trans. Commun.*, vol. COM-35, Nov. 1987.
- [40] H. A. Leinhos, "Capacity calculations for rapidly fading communications channels," *IEEE J. Oceanic Eng.*, vol. 21, pp. 137–142, Apr. 1996.
- [41] H.-H. J. Liao, "Multiple-access channels," Ph.D. dissertation, Univ. Hawaii, Sept. 1972.
- [42] T. L. Marzetta and B. Hochwald, "Multiple antenna communications when nobody knows the Rayleigh fading coefficients," in *Proc. 35th Allerton Conf. Communications, Control and Computation*, Urbana, IL, 1997.

- [43] R. J. McEliece and W. E. Stark, "Channels with block interference," *IEEE Trans. Information Theory*, vol. IT-30, pp. 44–53, Jan. 1994.
- [44] M. Médard, "The capacity of time varying multiple user channels in wireless communications," Sc.D. dissertation, Mass. Inst. Technol., Cambridge, MA, Sept. 1995.
- [45] E. C. van der Meulen, "A survey of multi-way channels in information theory: 1961–1976," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 1–37, Jan. 1977.
- [46] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert-Elliott channel," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1277–1290, Nov. 1989.
- [47] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359–378, May 1994.
- [48] J. H. Park, "Two-user dispersive communication channels," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 502–505, July 1981.
- [49] E. Pfaffelhuber, "Channels with asymptotically decreasing memory and anticipation," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 379–385, July 1971.
- [50] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes* (in Russian, translated by A. Feinstein). San Francisco, CA: Holden-Day, 1964.
- [51] G. Sh. Poltyrev, "Coding in an asynchronous multiple-access channel," *Probl. Inform. Transm.*, July–Sept. 1983.
- [52] R. Price and P. E. Green, "Communication technique for multipath channel," *Proc. IRE*, vol. 46, pp. 555–570, Mar. 1958.
- [53] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1983.
- [54] Qualcomm, "An Overview of the Application of Code Division Multiple Access to Digital Cellular Systems and Personal Communications Networks," Qualcomm, Inc., Jan. 1991.
- [55] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, pp. 1350–1387, Sept. 1985.
- [56] B. Rimoldi, "Rate-splitting multiple access and cellular communication," in *IEEE Information Theory Workshop*, Haifa, Israel, June 9–13, 1996.
- [57] B. Rimoldi and Q. Li, "Potential impact of rate-splitting multiple access on cellular communications," in *IEEE Global Communication Conf. (GlobeCom 96)*, Nov. 18–22, 1996.
- [58] B. Rimoldi and R. Urbanke, "On single-user decodable codes for the Gaussian multiple access channel," in *IEEE Int. Symp. Information Theory*, 1994.
- [59] —, "A rate-splitting approach to the Gaussian multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 42, pp. 364–375, Mar. 1996.
- [60] M. Rupp, F. Tarköy, and J. L. Massey, "User-separating demodulation for code-division multiple-access systems," *IEEE Trans. Select. Areas Commun.*, vol. 12, pp. 786–795, June 1994.
- [61] S. Shamai (Shitz), S. Verdú, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. Inform. Theory*, vol. 44, pp. 564–579, Mar. 1998.
- [62] S. Shamai (Shitz) and A. D. Wyner, "Information-theoretic considerations for symmetric, cellular, multiple-access fading channels—Parts I, II," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1877–1911, Nov. 1997.
- [63] J. J. Spilker Jr., "On the characterization and measurement of randomly time varying filters," in *IEEE Trans. Circuit Theory*, vol. CT-12, May 1965, pp. 44–51.
- [64] R. Steele, Ed., *Mobile Radio Communications*. New York: IEEE Press, 1992.
- [65] G. Strang, *Linear Algebra and Its Applications*, 2nd ed. New York: Academic, 1980.
- [66] G. L. Stüber, *Principles of Mobile Communication*. Boston, MA: Kluwer, 1996.
- [67] E. Telatar, "Multi-antenna multi-user fading channels," AT&T Bell Labs. Tech. Memo, 1995.
- [68] H. K. Ting, "On the information stability of a sequence of channels," *Theory of Probability and Its Applications*, vol. VII, no. 3, 1962.
- [69] B. S. Tsybakov, "The capacity of a memoryless Gaussian vector channel," *Probl. Pered. Inform.*, vol. 1, no. 1, 1965.
- [70] —, "On the transmission capacity of a discrete-time Gaussian channel with filter," *Probl. Pered. Inform.*, vol. 6, pp. 26–40, 1970.
- [71] —, "Survey of USSR contributions to random multiple-access communications," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 143–155, Mar. 1985.
- [72] M. L. Ulrey, "The capacity region of a channel with  $s$  senders and  $r$  receivers," *Inform. Contr.*, vol. 29, pp. 185–203, 1975.
- [73] S. Vembu and A. J. Viterbi, "Two different philosophies in CDMA—A comparison," in *IEEE Vehicular Technology Conf.*, 1996, pp. 869–873.
- [74] S. Verdú, "Multiple-access channels with memory with and without frame synchronization," *IEEE Trans. Inform. Theory*, vol. 35, pp. 605–619, May 1989.
- [75] —, "The capacity region of the symbol-asynchronous Gaussian multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 35, pp. 733–751, July 1989.
- [76] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1149–1157, July 1994.
- [77] A. J. Viterbi, "Very low rate convolutional codes for maximum theoretical performance of spread-spectrum multiple-access channels," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 641–649, May 1990.
- [78] J. Wolfowitz, *Coding Theorems of Information Theory*. New York: Springer-Verlag, 1978.
- [79] A. D. Wyner, "The capacity of the band-limited Gaussian channel," *Bell Syst. Tech. J.*, vol. 45, pp. 359–395, Mar. 1966.
- [80] M. Zakai, "Band-limited functions and the sampling theorem," *Inform. Contr.*, vol. 8, pp. 143–158, 1965.