On Optimal Signal Sets for Digital Communications with Finite Precision and Amplitude Constraints

Michael L. Honig, *Member, IEEE,* Stephen P. Boyd, *Member, IEEE,* B. Gopinath, *Member, IEEE,* and Erik Rantapaa

Abstract—**Given a linear, time-invariant, dispersive channel, a receiver that samples the channel output to within an accuracy of $\pm d$ where $d > 0$, and a transmitter with an output amplitude constraint, what is the maximum data rate that can be reliably communicated? For any dispersive channel the maximum rate depends on d, and is finite. The transmitted waveforms must be designed so that two channel outputs associated with two distinct transmitted signals are separated in amplitude at a particular time by d. It is shown that given any channel impulse response with rational Laplace transform, there exists an optimal set of inputs that are $\pm A$ everywhere where A is the maximum allowable amplitude. Furthermore, in any finite time interval each input changes sign a finite number of times. If the channel impulse response is a single decaying exponential, it is shown that simple binary signaling, in which A or $-A$, depending on the current message bit, is transmitted during each symbol interval, maximizes the data rate.**

I. Introduction

THE problem of designing transmitted signals for digital communications is reexamined here under different constraints from those typically assumed. For a given channel the primary limitation on the maximum data rate that can be reliably communicated is assumed to be the precision with which the receiver measures the channel output. Our motivation originates from communication systems in which impairments at the receiver cause the maximum achievable data rate to be considerably less than the Shannon capacity of the channel, assuming only additive thermal (Gaussian) noise.

A particular channel of interest is the subscriber loop, which typically consists of twisted wire-pairs. A single twisted-pair (ignoring crosstalk) is accurately modeled as a linear, time-invariant system. Furthermore, the amount of additive thermal noise introduced by the channel is very small, and does not pose a major limitation on achievable data rate. Rather, the main limitation is most likely due to inaccuracies introduced by a particular transmitter and receiver implementation, such as VLSI nonlinearities, timing inaccuracy, and the precision of the analog to digital (A/D) converter.

One approach to estimating the capacity of channels in the presence of the aforementioned receiver impairments is to model these impairments as additive noise with specified statistics, and subsequently attempt to compute the Shannon capacity of the resulting channel model. The primary difficulty with this approach is that an accurate statistical model of the preceding receiver impairments is generally unavailable, and appears to be difficult to construct. Furthermore, unless the noise statistics are assumed to be Gaussian, evaluating the Shannon capacity, subject to an appropriate input constraint, is likely to be a formidable task.

Here we take a simpler, and perhaps more useful approach to estimating the maximum data rate for the preceding types of channels. The channel is taken to be a linear, time-invariant system, and two channel outputs are assumed to be distinguishable at the receiver if and only if they are sufficiently separated in an appropriate metric space. A set of N inputs to the channel must therefore be designed so that the minimum distance between channels outputs is at least some prespecified amount. For a given minimum distance and input constraint, the maximum achievable data rate, or *maximum channel throughput*, is then the largest rate at which $\log N$ can grow with time. This is a deterministic notion of maximum achievable data rate, in contrast to the preceding statistical approach.

The metric used to distinguish the channel outputs should depend on the type of receiver impairments considered. For example, quantization error due to the A/D converter can be modeled as an additive noise which is bounded in amplitude by some constant $d/2$. Consequently, if this is the only impairment, it is appropriate to design channel inputs so that any two distinct channel outputs are separated in amplitude by at least d at a particular time instant. In this case the corresponding metric space in which the channel outputs are to be separated is L_∞. This is the case studied in this paper. That is, it is implicitly assumed that all receiver impairments can be modeled as an additive noise which is bounded in amplitude by $d/2$ almost surely. No additional assumptions will be made concerning the statistical properties of the receiver impairments. In addition, we will assume that the transmitter output is constrained by the dynamic range of the electronics, which implies a maximum input amplitude constraint.

Different assumptions about the receiver impairments lead to different metric spaces in which the channel outputs should be separated. For example, if it is assumed that the receiver impairments can be modeled as an additive noise which has bounded power, then the appropriate metric space for the channel outputs is L_2. In each case the statistical properties of the noise are lumped into a single constant representing the maximum amount

Paper approved by the Editor for Signal Design, Modulation, and Detection of the IEEE Communications Society. Manuscript received July 13, 1988; revised February 12, 1990. This paper was presented in part at the 1987 GLOBECOM Conference, Tokyo, Japan.

M. L. Honig is with Bell Communications Research, Morristown, NJ 07960.

S. Boyd is with the Department of Electrical Engineering, Stanford University, Stanford, CA.

B. Gopinath is with the Department of Electrical Engineering, Rutgers University, Piscataway, NJ 08855.

E. Rantapaa is with the Department of Mathematics, University of Minnesota, Minneapolis, MN 55455.

IEEE Log Number 9040897.

0090-6778/91/0200-0249$01.00

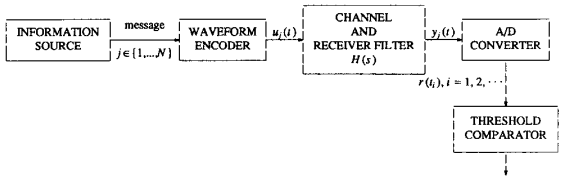

Fig. 1. Communications system model.

the noise can perturb the channel outputs. This type of input signal design is therefore worst case in the sense that a higher data rate might be achievable by exploiting additional statistical properties of the receiver impairments.

The communications system model considered in this paper is shown in Fig. 1. We wish to transmit one of N messages, corresponding to the transmitted signal $u_j(t)$, $1 \le j \le N$, in a finite time interval $[0, T]$. The channel has impulse response $h(t)$, transfer function $H(s)$, and maps the input to the output $y_j(t) = h * u_j(t)$ where "$*$" denotes convolution. (Any linear processing of the channel output, which the receiver may perform, is assumed to be part of the channel transfer function.)

The receiver samples the channel output at prespecified times t_i, $i = 1, 2, \cdots$, to within an accuracy of $\pm d$, and selects the transmitted message $\hat{j}$ based on threshold comparisons. The sampling times $\{t_i\}$ are times at which two channel outputs are known to be separated by d. Specifically, let $r(t_i)$, $i = 1, 2, \cdots$ denote the received samples, and suppose that $y_1(t_1) - y_2(t_1) > d$. Then if $r(t_1) > [y_1(t_1) + y_2(t_1)]/2$, the receiver rejects message 2 as the final estimate. Otherwise, if "$>$" is replaced by "$<$," then message 1 is rejected. The estimate $\hat{j}$ therefore corresponds to the input u_j such that $|r(t_i) - y_j(t_i)| < d/2$ for each i.

Suppose that the A/D converter in Fig. 1 is replaced by an additive noise $n(t)$ where $|n(t)| < d/2$ for all t, so that $r(t) = y(t) + n(t)$. Then the estimate $\hat{j}$ corresponds to the input $u_{\hat{j}}(t)$ which produces the channel output $y_{\hat{j}}(t)$ closest to $r(t)$ in the L_∞ sense. In practice, a channel estimator, or equalizer, can be used at the receiver to subtract out intersymbol interference, and thereby center the threshhold comparisons about zero. Note that this type of receiver is often used in practice, with the added restriction that the sampling times are uniformly spaced.

The problem studied here was considered as early as 1928 by Hartley [1]; however, the authors are unaware of any other closely related work. Hartley's paper considers only channels with an exponential impulse response. A similar problem to that posed here, at least in appearance, in which the outputs must be separated in the L_2 sense, and the inputs are constrained in L_2 norm, was studied by Root [2] and by Wyner [3].

In the next section the problem outlined in this section is stated precisely. Section III presents the main results, and Section IV mentions some related problems.

II. Problem Statement

Some notation is first defined. The L_∞ norm of a continuous, real-valued function f over an interval $[0, T]$ is given by

$$\|f\|_{T,\infty} \equiv \sup_{0 \le t \le T} |f(t)|. \tag{2.1}$$

Since we will not use any other norms in this paper, we will simply write this as $\|f\|_T$. (If f is not continuous, then "sup" is replaced by "essential sup" [4].) The channel is assumed to be linear and time-invariant with real-valued, bounded impulse response $h(t)$ where $h(t) = 0$ for $t < 0$. The channel output in response to input $u(t)$ is therefore given by

$$y(t) = h * u(t) \equiv \int_0^t h(s)u(t-s)\,ds. \tag{2.2}$$

We also assume that $\int_0^\infty |h(t)|\,dt < \infty$ (i.e., $h(\cdot) \in L_1$), so that any bounded input produces a bounded output.

Throughout the rest of the paper we will assume a transmitter amplitude constraint. Specifically, any input to the channel is assumed to be less than or equal to one in magnitude ($\|u_j\|_T \le 1$ for each j and any $T > 0$).

The following problems are precise versions of those outlined in Section I.

P1) Given some time interval $[0, T]$ and some small constant $d > 0$, find inputs $u_1(t), \cdots, u_N(t)$ such that $\min_{i \ne j} \|y_j - y_i\|_T \ge d$, with N as large as possible. Let $N_{\max}(T, d)$ denote the largest possible N for fixed time interval $[0, T]$ and discrimination d.

P2) Given the discrimination d and the number of messages N, find N inputs $u_1(t), \cdots, u_N(t)$ that minimize the time T such that $\min_{i \ne j} \|y_j - y_i\|_T \ge d$. Let $T_{\min}(N, d)$ denote the minimum time.

P3) Given the interval $[0, T]$ and the number of messages N, find N inputs $u_1(t), \cdots, u_N(t)$ that maximize the minimum output separation $d = \min_{i \ne j} \|y_j - y_i\|_T$.

Of course, these three problems are related, for example,

$$T_{\min}(N, d) = \inf\{T \mid N_{\max}(T, d) = N\} \tag{2.3}$$

and

$$N_{\max}(T, d) = \max\{N \mid T_{\min}(N, d) \le T\}. \tag{2.4}$$

Note that $N_{\max}(T, d) = 0$ for all $T > 0$ if $\int_0^\infty |h(t)|\,dt < d/2$. If $\int_0^\infty |h(t)|\,dt = d/2$, then $N_{\max}(T, d) \le 2$ for any $T > 0$.

The discrete-time versions of problems P1)–P3) are also of interest. In this case $\boldsymbol{h} * \boldsymbol{u}_j = \boldsymbol{H}\boldsymbol{u}_j$, where $\boldsymbol{h}$ is the vector of impulse response coefficients, the inputs $\{\boldsymbol{u}_j\}$ are vectors in $\mathbb{R}^T$, the outputs $\{\boldsymbol{y}_j\}$ are vectors in $\mathbb{R}^T$, and the rows of the matrix $\boldsymbol{H}$ are shifted versions of $\boldsymbol{h}$. The discrete time l_∞ norm is defined by $\|\boldsymbol{u}\|_\infty \equiv \max_j |[\boldsymbol{u}]_j|$.

We now make two remarks concerning P1)–P3). In order to obtain a reliable estimate of the transmitted message, it is assumed that the receiver samples the output at times where two outputs differ by at least d. Of course, in practice the receiver cannot sample at precisely the correct time instant. However, for any allowable input and impulse response, the output will be continuous. Two outputs separated by d at some time t_0 will therefore be separated by at least $d - \epsilon$ for some small ϵ in a neighborhood of t_0. Compensation for timing errors can be therefore be made by choosing d large enough so that the distance between outputs is sufficiently large when the sampling times are shifted by small amounts.

The second remark is that P1)–P3) state that the distance between outputs is measured with respect to the interval $[0, T]$. For many channels, however, two outputs can be first separated by d when $t > T$ even though the associated inputs are zero for $t > T$ (i.e., when the channel has a group delay). In this case P1)–P3) can be reformulated so that the outputs must be separated by d on the interval $[0, T + \tau]$ where $\tau > 0$ and is independent of T. Since the impulse response $h \in L_1[0, \infty)$, we can assume that τ is finite. That is, because the inputs are zero

for $t > T$, there exists a finite $\tau > 0$ such that for any T, two outputs cannot become separated by d at any time $t > T + \tau$. In what follows it is convenient to first assume that $\tau = 0$ before considering the case $\tau > 0$.

For a given allowable channel impulse response $h(\cdot)$ and discrimination d, the *maximum channel throughput* (MCT) is defined as

$$\mathrm{MCT}(d) = \sup_{T>0} \frac{\log N_{\max}(T, d)}{T}. \tag{2.5}$$

We remark that

$$\sup_{T>0} \frac{\log N_{\max}(T, d)}{T} = \sup_{N} \frac{\log N}{T_{\min}(N)}.$$

To show that the left side is less than or equal to the right side, fix $T > 0$ and $d > 0$, and let $N = N_{\max}(T, d)$. For this choice of N, $[\log N_{\max}(T, d)]/T \le (\log N)/T_{\min}(N, d)$, i.e., for any fixed $T > 0$, there exists an N such that the preceding inequality holds. Conversely, fixing N and letting $T = T_{\min}(N, d)$ establishes that the right-hand side is less than or equal to the left-hand side.

Fact 1:

$$\mathrm{MCT}(d) = \lim_{T\to\infty} \frac{\log N_{\max}(T, d)}{T}. \tag{2.6}$$

In particular, the limit exists.

Roughly speaking, this establishes that $N_{\max}(T)$ asymptotically grows exponentially with T, with exponential coefficient equal to the MCT. We remark here that $\log N_{\max}(T)/T$ is *not* monotonic in T. This is because $N_{\max}(T)$ increases only at a discrete set of times. Fact 1 follows from two simple lemmas.

Lemma 1:

$$N_{\max}(T + S) \ge N_{\max}(T) N_{\max}(S). \tag{2.7}$$

Thus in time $T + S$ we can transmit at least $N_{\max}(T)N_{\max}(S)$ messages.

Proof of Lemma 1: Let $N = N_{\max}(T)$ and $v_1, \cdots, v_N$ be input signals such that $\| h * v_i - h * v_j \|_T \ge d$ for $i \ne j$. Similarly, let $M = N_{\max}(S)$ and $w_1, \cdots, w_M$ be a set of input signals such that $\| h * w_i - h * w_j \|_s \ge d$ for $i \ne j$. Define

$$u_{ij}(t) \equiv \begin{cases} v_i(t) & 0 \le t \le T \\ w_j(t - T) & T < t \le T + S \end{cases} \qquad i = 1, \cdots, N,\ j = 1, \cdots, M \tag{2.8}$$

which are just the concatenations of the signals w_j with the signals u_i. We now show that $\| h * u_{ij} - h * u_{kl} \|_{T+S} \ge d$ for $i \ne k$ or $j \ne l$, which will establish the lemma. To see this, note that if $i \ne k$ then

$$d \le \| h * u_{ij} - h * u_{kl} \|_T \le \| h * u_{ij} - h * u_{kl} \|_{T+S}. \tag{2.9}$$

If on the other hand $i = k$ but $j \ne l$, then $h * u_{ij}(t) = h * u_{kl}(t)$ for $0 \le t \le T$ and

$$h * u_{ij}(t) - h * u_{kl}(t) = h * w_j(t - T) - h * w_l(t - T) \qquad \text{for } T \le t \le T + S \tag{2.10}$$

so that

$$d \le \| h * w_{ij} - h * w_{kl} \|_T = \| h * u_{ij} - h * u_{kl} \|_{T+S}. \tag{2.11}$$

□

Lemma 2: Let $f: [0, \infty) \to [0, \infty)$ be superadditive, that is,

$$f(a + b) \ge f(a) + f(b) \qquad \text{for all } a, b \ge 0. \tag{2.12}$$

Then

$$\sup_{x>0} \frac{f(x)}{x} = \lim_{x\to\infty} \frac{f(x)}{x}. \tag{2.13}$$

By Lemma 1, $\log N_{\max}(T)$ is a superadditive function of T, so Lemma 2 will establish Fact 1.

Proof of Lemma 2: Clearly, f is nondecreasing. From the superadditive property, for any a and any nonnegative integer n we have $f(na) \ge nf(a)$. Given $x \ge 0$, let $x = na + r$ where $0 \le r < a$. Using the monotonicity of f and the fact cited above, we have

$$\frac{f(x)}{x} \ge \frac{f(na)}{na + r} \ge \frac{na}{na + r} \frac{f(a)}{a}. \tag{2.14}$$

Note that for x large enough, the right-hand side of the above inequality exceeds $f(a)/a - \epsilon$ for any positive ϵ. Hence,

$\forall a > 0, \forall \epsilon > 0, \exists B$ such that

$$x \ge B \text{ implies } \frac{f(x)}{x} \ge \frac{f(a)}{a} - \epsilon,$$

from which lemma 2 follows. □

Consider now P1)–P3) where the distance between outputs is measured with respect to the time interval $[0, T + \tau]$ where τ is a finite, positive constant. The maximum channel throughput can be redefined as

$$\mathrm{MCT}'(d) = \lim_{T\to\infty} \frac{\log N'_{\max}(T, d)}{T + \tau}, \tag{2.15}$$

assuming the limit exists where $N'_{\max}$ is the maximum number of distinguishable outputs with respect to the time interval $[0, T + \tau]$, assuming the inputs are zero for $t < 0$ and $t > T$. Note that

$$\frac{\log N_{\max}(T, d)}{T + \tau} \le \frac{\log N'_{\max}(T, d)}{T + \tau} \le \frac{\log N_{\max}(T + \tau, d)}{T + \tau} \tag{2.16}$$

for any finite $T > 0$. Now as $T \to \infty$, the left and right hand expressions converge to $\mathrm{MCT}(d)$, which is assumed to be finite. Consequently, Fact 1 implies that the limit in (2.15) does exist, and that $\mathrm{MCT}'(d) = \mathrm{MCT}(d)$. That is, the definitions (2.5) and (2.15) are equivalent. It is easily verified that the results in the next section are valid for any $\tau \ge 0$, assuming that the MCT is replaced by MCT′ defined by (2.15).

III. The Main Results

Problems P1)–P3) have not yet been solved for general $h(\cdot)$; however, here we give some partial results, stated by Theorems 1–4. The proofs of Theorems 2 and 3 are given in the Appendix.

Assume that the channel has the state-space characterization,

$$\dot{x}(t) = Ax(t) + bu(t) \qquad y(t) = cx(t) \tag{3.1}$$

where $u(t)$ and $y(t)$ are the input and output of the channel, respectively, and $x(t)$ is an n-vector representing the state of the channel where n is finite. Of course, A is an $n \times n$ matrix, and b and c are n element column and row vectors, respectively, and $H(s) = c(sI - A)^{-1}b$ is the channel transfer function.

Theorem 1: For the channel with transfer function $H(s)$ given in the preceding paragraph, there exists a solution to P2) such that $|u_i(t)| = 1$ for each $i = 1, \cdots, N$ and $0 \le t \le T$. Furthermore, in any finite time interval, each u_i changes sign a finite number of times.

This theorem is a consequence of the following lemma.

Lemma 3: Assume that the system (3.1) is a controllable realization of $H(s)$, and that there exists an input $u(t)$ which drives the state from $x(t_i)$ to $x(t_f)$ where $t_f > t_i$. Then there exists an input $\tilde{u}(t)$ where $|\tilde{u}(t)| = 1$, which also drives the state from $x(t_i)$ to $x(t_f)$. If the eigenvalues of the $n \times n$ matrix A in (3.1) are real, then the number of times $\tilde{u}(t)$ changes sign is at most $n - 1$ whereas if the eigenvalues of A are complex, then the number of times $\tilde{u}(t)$ changes sign is finite, but depends upon the initial and final state.

Proof: Define $T(x_i, x_f)$ as the minimum time to go from the initial state x_i to the final state x_f. Then $t_f - t_i \ge T[x(t_i), x(t_f)]$. Assume that $\tilde{u}(t) = 1$ (or -1) for $t_i \le t \le t^* \le t_f$. We wish to show that there exists a t^* such that

$$t^* + T[x(t^*), x(t_f)] = t_f - t_i. \tag{3.2}$$

$T(x_i, x_f)$ is a continuous function of x_i (see [5, Section 6-7), which implies that $T[x(t^*), x(t_f)]$ is a continuous function of t^*. Consequently, a t^* which satisfies (3.2) exists by the intermediate value theorem. Furthermore, for $t^* \le t \le t_f$, $\tilde{u}(t)$ becomes the minimum time control that drives the state from $x(t^*)$ to $x(t_f)$. The lemma therefore follows from Theorem 6-8 in [5] (and the following discussion), which gives the number of times the minimum time control changes sign. □

Proof of Theorem 1: Let $u_1(t), \cdots, u_N(t)$ be a solution to P2), and assume that at time t_{ij}, $|y_i(t_{ij}) - y_j(t_{ij})| \ge d$. Without loss of generality, pick t_{1j} so that $t_{11} = 0 \le t_{12} \le \cdots \le t_{1N}$. Associated with $u_1(t)$ is therefore the sequence of states $x_1(t_{12}), \cdots, x_N(t_{1N})$. Lemma 3 implies that there exists an input $\tilde{u}_1(t)$, which switches between $+1$ and -1 during each interval $(t_{1,j-1}, t_{1j})$, $j = 2, \cdots, N$, such that the same sequence of states will be visited if $u_1(t)$ is replaced by $\tilde{u}_1(t)$. Replacing each input $u_j(t)$ by an input $\tilde{u}_j$ in this manner guarantees that $|h * \tilde{u}_i(t_{ij}) - h * \tilde{u}_j(t_{ij})| \ge d$, which gives the result. □

Notice that the preceding proof and Lemma 3 imply that if the eigenvalues of A are negative real, then there exists a solution to P2) such that the magnitude of each input is one for all $0 \le t \le T$, and each input changes sign a maximum of $(N - 1)(n - 1)$ times.

Consider now the specific impulse response

$$h(t) = Ae^{-\alpha t} \tag{3.3}$$

where $\alpha > 0$, and assume that the number of messages N and the discrimination d are fixed. Let b_{jk} be the kth digit in the binary expansion of the integer j representing message j, $0 \le j \le N - 1$ where "0" is replaced by "-1," and $k \le \lceil \log_2 N \rceil$, the smallest integer greater than or equal to $\log_2 N$. For example, if $N = j = 5$, then $\{b_{51}, b_{52}, b_{53}\} = \{1, -1, 1\}$.

Theorem 2: A solution to P2) is given by

$$u_j(t) = b_{jk}, \quad (k-1)t_0 \le t < kt_0, \tag{3.4a}$$

where

$$t_0 = -\frac{1}{\alpha}\ln\left(1 - \frac{d}{2A}\right), \tag{3.4b}$$

and $1 \le k \le \lceil \log_2 N \rceil$.

The optimal inputs in this case are therefore ± 1, corresponding to whether a zero or one is the current source bit, for the fixed duration t_0 ("bit-by-bit" or "binary" signaling). Theorem 2 implies that for the impulse response (3.3),

$$\mathrm{MCT}(d) = -\frac{\alpha}{\ln\left(1 - \frac{d}{2A}\right)}. \tag{3.5}$$

The inverse of the right-hand side is the time it takes to transmit one bit.

So far, the impulse response (3.3) is the only nontrivial example for which $T_{\min}(N, d)$ can be explicitly computed for all N and d. For the class of functions h defined in the next Theorem, however, it is possible to compute $T_{\min}(N, d)$ for $N \le 4$.

Theorem 3: If h is an integrable, nonincreasing function defined for $t \ge 0$, then

$$T_{\min}(4, d) = 2t_0(d), \tag{3.6}$$

where

$$\int_0^{t_0(d)} h(t)\, dt = \frac{d}{2}. \tag{3.7}$$

By taking $\alpha = 0$, it is apparent that Theorem 2 also applies to the integrator impulse response $h(t) = a$ where a is a constant and $t_0 = |d/2a|$. This observation is used to derive the following upper bound on MCT in terms of the total variation of the channel impulse response.

Theorem 4: Assume that $h(\cdot)$ is differentiable on $(0, \infty)$, and let K denote the total variation of h, that is, $K = |h(0)| + \int_0^\infty |dh/dt|\, dt$. Then the MCT of a channel with impulse response h satisfies

$$\frac{1}{t_0(d)} \le \mathrm{MCT}(d) \le \frac{2K}{d} \tag{3.8}$$

where $t_0(d)$ is any time such that

$$\int_0^{t_0(d)} |h(t)|\, dt = \frac{d}{2}. \tag{3.9}$$

Proof: To derive the lower bound let

$$u_1(t) = \begin{cases} \mathrm{sgn}\{h[t_0(d) - t]\} & 0 \le t \le t_0(d) \\ 0 & \text{elsewhere} \end{cases} \tag{3.10}$$

and $u_2(t) = -u_1(t)$. Then we have

$$y_1[t_0(d)] - y_2[t_0(d)] = \int_0^{t_0(d)} |h(t)|\, dt = d \tag{3.11}$$

so that $\|y_1 - y_2\|_{t_0(d)} \ge d$. It follows that $(\log N_{\max}[t_0(d)])/t_0(d) \ge 1/t_0(d)$ (in fact equality must hold), so of course $\mathrm{MCT} \ge 1/t_0(d)$.

The upper bound is established by showing that the MCT of a channel with impulse response of total variation K must be less than the MCT of a channel with impulse response $h(t) = K$. The fact that bit-by-bit signaling, as defined in Theorem 2, is optimal for $h(t) = K$ then completes the proof.

Let $u_1(t), \cdots, u_N(t)$ be a solution to Problem P2) for a channel having impulse response $h(t)$. We define the *ij*th *sampling time* as

$$t_{ij} = \min\{t_0 \text{ such that } \|y_i - y_j\|_{t_0} = d\}. \tag{3.12}$$

For fixed N there are $N(N-1)/2$ sampling times; however, many sampling times may coincide. Let t_l, $l = 1, \cdots, L$, be the distinct sampling times in ascending order where $L \le N(N-1)/2$. Set $t_0 = 0$. For $i = 1, \cdots, N$ define

$$v_i(t) \equiv \frac{y_i(t_{l+1}) - y_i(t_l)}{K(t_{l+1} - t_l)} \qquad \text{for } t_l \le t < t_{l+1}. \quad (3.13)$$

Since the total variation of h is K and $\|u_i\| \le 1$, we have $|y_i(t) - y_i(s)| \le K|t - s|$. Hence, the signals v_i defined above satisfy $\|v_i\| \le 1$. Furthermore, if we let $\tilde{y}_i = K * v_i$, then $\tilde{y}_i(t_l) = y(t_l)$, from which it follows that

$$\min_{i \ne j} \|K * v_i - K * v_j\| \ge d. \quad (3.14)$$

Thus for each T, $N_{\max}(T)$ for the channel with impulse response K is at least as big as $N_{\max}(T)$ for the channel with impulse response h. It follows that the MCT for the channel with impulse response K is at least as big as the MCT for the channel with impulse response h. □

The lower bound on MCT becomes equality when the impulse response is a single exponential, and the upper bound becomes equality when the impulse response is a constant. Evaluating K for $h(t) = e^{-t}$ gives

$$-\frac{1}{\ln(1 - d/2)} = \mathrm{MCT}(d) \le \frac{4}{d}. \quad (3.15)$$

As $d \to 0$, the MCT approaches $2/d$, so that the upper bound is twice the lower bound in this case. Both the lower and upper bounds on MCT presented in Theorem 4 have been improved recently [6].

IV. Open Issues

Problems P1)–P3) remain unsolved except for the specific case mentioned in Section III. One conjecture is that bit-by-bit signaling, as described in Theorem 2, is the solution for the class of impulse responses defined in Theorem 3.

All solutions to P1)–P3) may require that the inputs switch instantaneously between 1 and -1, or vice versa. Since this is impractical, it is of interest to reconsider P1)–P3) with additional constraints placed on the inputs $u_i(t)$. For instance, the magnitude of the derivatives of the inputs might be constrained.

Problems P1)–P3) are easily generalized to the case where data is to be transmitted over multiple *coupled* channels. In this case the channel impulse response is a matrix, $\boldsymbol{H}(t)$, the (i, j)th entry being the output of channel j when an impulse is applied to channel i. The problem is then to design the maximum number of *vector* inputs $\boldsymbol{u}_1(t), \cdots, \boldsymbol{u}_N(t)$, each having M elements where M is the number of channels, in a given time interval $[0, T]$ so that $\|\boldsymbol{H} * \boldsymbol{u}_i(t) - \boldsymbol{H} * \boldsymbol{u}_j(t)\|_T \ge d$, $i \ne j$. Then L_∞ norm of a continuous vector time function on the interval $[0, T]$ is $\|\boldsymbol{x}(t)\|_T = \sup_{j,\, 0 \le t \le T} |x_j(t)|$ where $x_j(t)$ is the jth component of $\boldsymbol{x}(t)$. An interesting question is how does the MCT behave as a function of cross-coupling between channels?

Appendix

Proofs of Theorems 2 and 3

Proof of Theorem 2: Let $S = [0, \infty) \times \mathbb{R}$ denote the state space corresponding to a single control u. A member (t, y) of S will be called a *state* and corresponds to an output value $y = h * u$ at time t where $h(t) = Ae^{-\alpha t}$. Given two states (t_0, y_0) and (t_1, y_1) with $t_0 \le t_1$, we will say that (t_1, y_1) is *reachable from* (t_0, y_0) if there exists an input u such that $|u| \le 1$, $y(t_0) = y_0$ and $y(t_1) = y_1$. All the states reachable from a given state (t, y) will be denoted by $R_{(t, y)}$. Note that $R_{(t, y)} \subset R_{(0,0)}$ for any state (t, y) reachable from $(0, 0)$. $R_{(0,0)}$ is also sometimes called the set of all reachable states.

Lemma A.1: Let $(0, 0) = (t_0, y_0), (t_1, y_1), \cdots, (t_n, y_n)$ be a sequence of states such that $t_0 \le t_1 \le \cdots \le t_n$ and (t_{i+1}, y_{i+1}) is reachable from (t_i, y_i) for $i = 0, \cdots, n-1$. Then there exists an input u, $|u| \le 1$, such that the output $y(t_i) = y_i$ for $i = 1, \cdots, n$.

Proof: This follows immediately from considering the differential equation satisfied by the output: $dy/dt + \alpha y = u$. □

For notational convenience we will denote $\int_0^t h(s)\, ds$ as $h * 1(t)$.

Lemma A2: $R_{(0,0)} = \{(t, y) \mid -h * 1(t) \le y \le h * 1(t)\}$. Hence, the set of all reachable states is bounded above by $h * 1$ and below by $-h * 1$.

Proof: This follows directly from the fact that h is nonnegative and the inputs are constrained to be at most 1 in absolute value. □

Let $\chi_{[a, b)}$ denote the characteristic function of the interval $[a, b)$, i.e.

$$\chi_{[a, b)} = \begin{cases} 1, & t \in [a, b) \\ 0, & \text{otherwise} \end{cases}. \quad (A.1)$$

Lemma A.3: Suppose that $y_0 = h * 1(t_0)$. Then

$$R_{(t_0, y_0)} \supseteq \{(t, y) \mid t \ge t_0, \text{ and } h * (\chi_{[0, t_0)} - \chi_{[t_0, \infty)})(t) \le y \le h * (\chi_{[0, t_0)} + \chi_{[t_0, \infty)})(t)\}. \quad (A.2)$$

A similar result holds for the case $y_0 = -h * 1(t_0)$.

Proof: The two inputs $u_1 = \chi_{[0, t_0)} + \chi_{[t_0, \infty)}$ and $u_2 = \chi_{[0, t_0)} - \chi_{[t_0, \infty)}$ both pass through the state (t_0, y_0), and any intermediate state between y_1 and y_2 can be obtained by considering the input $u(t) = \chi_{[0, t_0)} + \lambda \chi_{[t_0, \infty)}$ for $-1 \le \lambda \le +1$. □

Given a discrimination $d > 0$, define the time t_0 as the minimum for which $h * 1(t_0) = d/2$. We now define two sets of states U_d and L_d as

$$U_d = \{(t, y) \mid -h * 1(t) + d \le y \le h * 1(t)\} \quad (A.3a)$$

$$L_d = \{(t, y) \mid -h * 1(t) \le y \le h * 1(t) - d\}. \quad (A.3b)$$

Lemma A.4: Let u_1, u_2 be two inputs with $|u_1|, |u_2| \le 1$. If at some time $t > 0$, $h * u_1(t) - h * u_2(t) = d$, then $(t, h * u_1(t)) \in U_d$ and $(t, h * u_2(t)) \in L_d$.

Proof: Just note that the two states involved are reachable and so Lemma A.2 applies. The necessary inequalities directly follow. □

By making the substitution $u_1 = +1$ and $u_2 = -1$, Lemma A.4 says that $(t_0, d/2) \in U_d$ and $(t_0, -d/2) \in L_d$. Moreover, it is clear that t_0 is the earliest time for any state in U_d or L_d.

Lemma A.5: $U_d \subset R_{(t_0, d/2)}$ and $L_d \subset R_{(t_0, -d/2)}$.

Proof: Since U_d and $R_{(t_0, d/2)}$ are just mirror images of L_d and $R_{(t_0, -d/2)}$, respectively, we shall only show $U_d \subset R_{(t_0, d/2)}$. By Lemma A.3, we only need to show that U_d is contained in the set on the right side of (A.2). Hence, given

$$-h * 1(t) + d \le y \le h * 1(t), \quad (A.4)$$

we need to show

$$h*(\chi_{[0,t_0)}-\chi_{[t_0,\infty)})(t)\le y\le h*(\chi_{[0,t_0)}+\chi_{[t_0,\infty)})(t). \tag{A.5}$$

Since $\chi_{[0,t_0)}+\chi_{[t_0,\infty)}\equiv 1$, the right inequality of (A.5) is immediately satisfied. The left inequality will be satisfied if we can show

$$h*(\chi_{[0,t_0)}-\chi_{[t_0,\infty)})(t)\le -h*1(t)+d,\quad t\ge t_0. \tag{A.6}$$

Adding $h*1(t)$ to both sides and writing $1=\chi_{[0,t_0)}+\chi_{[t_0,\infty)}$, we are reduced to showing

$$2h*\chi_{[0,t_0)}(t)\le d,\quad t\ge t_0. \tag{A.7}$$

This, however, always holds since h is nonincreasing and t_0 was chosen such that $h*1(t_0)=d/2$. □

Lemma A.6: Let $\{u_i\}$ be inputs and $\{t_{ij}\}$ be sampling times, defined by (3.12), with respect to a discrimination $d>0$. Let t_0 be determined by the equation $h*1(t_0)=d/2$. Then there exist inputs $\{\tilde{u}_i\}$ such that

a) on the interval $[0, t_0]$, the inputs $\tilde{u}_i$ are constant, assuming the value of either $+1$ or -1,

b) there exist two indexes i, j such that $|h*\tilde{u}_i(t_0)-h*\tilde{u}_j(t_0)|=d$, and

c) $h*\tilde{u}_i(t_{ij})=h*u_i(t_{ij})$ for every i, j.

Proof: Let t_i denote the first sampling time for input u_i. Note that $t_i\ge t_0$. At time t_i, the output $h*u_i$ is separated from another output by d, so by Lemma A.4, the first sampling state $(t_i, h*u_i(t_i))$ is either in U_d or L_d (or possibly both). If it is in U_d, then by Lemmas A.1 and A.5 we can replace u_i by an input which is $+1$ on the interval $[0, t_0]$ and still have it pass through all of the sampling states associated with u_i. A similar result follows if the first sampling state is in L_d. Consequently, conditions a) and c) can be satisfied. Condition b) can easily be met if we just choose one of the replacements to be $+1$ on the interval $[0, t_0]$ and another to be -1 on that same interval. □

Lemma A.7: A solution to Problems P1)–P3) for the case $h(t)=Ae^{-\alpha t}$ is given by bit-by-bit signaling.

Proof: Clearly the result is true when the number of inputs is one or two. Assume the result holds when the number of inputs is n or less. Given an optimal solution for $n+1$ inputs, Lemma A.6 says that there exists another solution in which the inputs split up into two groups: those which are $+1$ on the interval $[0, t_0]$ and those which are -1. Moreover, neither of these two groups are empty. At time t_0, the first group has separated from the second group so all that has to be done is the separation of inputs within each group. Since on the interval $[0, t_0]$ the inputs in a given group are the same, the way they separate on $[t_0, \infty]$ must itself be optimal. Hence, by the induction hypothesis, we can assume this is bit-by-bit signaling. Finally, noting that making the number of inputs in each of the two groups as close as possible minimizes the total separation time completes the proof. □

Proof of Theorem 3: The proof relies upon the following Lemma.

Lemma A.8: Let f be a nonnegative, integrable function and I a real number, $0<I\le\int f$. Consider the problem of minimizing $\int g$ subject to $\int fg\ge I$ and $0\le g\le 1$. Suppose that A is a set of the form $f^{-1}((c,\infty])\cup B$ where $B\subset f^{-1}(c)$ for some real number c. Further, suppose that $\int_A f=I$. Then χ_A is an optimal solution to this problem in the sense that if g is another such solution, then $\int g\ge\int\chi_A$.

Proof: Let g be any other solution. Then

$$\int_A f=I\le\int fg. \tag{A.8}$$

Writing $\int fg=\int_A fg+\int_{A^c}fg$ where A^c denotes the complement of A, we have

$$\int_A f(1-g)\le\int_{A^c}fg. \tag{A.9}$$

On A, $f\ge c$ whereas on A^c, $f\le c$. Thus, $\int_A 1-g\le\int_{A^c}g$ whence $\int\chi_A\le\int g$. □

As an immediate corollary, we have the following.

Lemma A.9: Let h be a nonnegative, nonincreasing impulse response, u a nonnegative input, and $d>0$ a discrimination. If t_0 is the minimum for which $h*1(t_0)=d/2$, and $h*u(t_0)\ge d/2$, then $\int_0^{t_0}u(t)\,dt\ge t_0$.

Proof: Just note the set $[0, t_0]$ is of the form $h^{-1}((c,\infty])\cup B$ where $B\subset h^{-1}(c)$ for some real number c. Hence Lemma A.8 applies. □

From now on, let h, d, and t be as described in Lemma A.9. Also, let u^+ denote the function $\max(u,0)$ and u^- denote the function $-\min(u,0)$. One then has that both u^+ and u^- are nonnegative functions and $u=u^+-u^-$.

Lemma A.10: Let u_1 and u_2 be two inputs such that at time t_0, $h*u_1(t_0)-h*u_2(t_0)\ge d$. Then $\int_0^{t_0}[(u_1-u_2/2)]^+\,dt\ge t_0$. Furthermore, the function $u=[(u_1-u_2/2)]^+$ is nonnegative and bounded by 1.

Proof: Apply Lemma A.9 to the function $u=[(u_1-u_2/2)]^+$. Note that since u_1 and u_2 are bounded in absolute value by 1, u is bounded by 1. □

To prove Theorem 3 it suffices to show that $T_{\min}(3, d)=2t_0$ where t_0 is determined by (3.7). This is because $T_{\min}(N, d)$ is a nondecreasing function of N, and a set of three inputs that achieves $T_{\min}=2t_0$ is

$$u_1(t)=-u_3(t)=1,\qquad 0\le t\le t_0 \tag{A.10a}$$

$$u_2(t)=\begin{cases}1 & 0\le t<t_0\\ -1 & t_0\le t<2t_0\end{cases}. \tag{A.10b}$$

Adding the fourth input $u_4(t)=-u_3(t)$ gives $T_{\min}(4, d)=T_{\min}(3, d)$.

Lemma A.11: Let $\{u_i\}$, $i=1,\cdots,N$, be a solution to P2) for $N\ge 3$. Then there exist three inputs u_1, u_2, u_3 and two times t_1, t_2 such that $h*u_1(t_1)-h*u_2(t_1)=d$ and $h*u_2(t_2)-h*u_3(t_2)=d$.

Proof: Let $\{t_{ij}\}$ be the set of sampling times for the $\{u_i\}$. Consider the matrix A in which the element A_{ij} is $+1$ if $h*u_i(t_{ij})>h*u_j(t_{ij})$ and -1 otherwise. The diagonal elements are not important. Clearly, A is antisymmetric. For a group of three or more inputs, it is easy to see by inspection that it is impossible for all the rows of A to be only $+1$ or only -1. Hence, there exists a row which has both $+1$ and -1 elements. Consequently, there exists an output which at one sampling time is greater than the output it is separating from, and at another sampling time is less, which is what we wanted to show. □

Lemma A.10 and A.11 imply that for any set u_1, u_2, u_3 that is a solution to P2) for $N=3$,

$$\int_0^{t_1}\left(\frac{u_1-u_2}{2}\right)^+dt\ge t_0\quad\text{and}\quad\int_0^{t_2}\left(\frac{u_2-u_3}{2}\right)^+dt\ge t_0. \tag{A.11}$$

Extending both integrals to $T=\max(t_1, t_2)$ and adding them

yields

$$2t_0 \le \int_0^T \left[\left(\frac{u_1 - u_2}{2} \right)^+ + \left(\frac{u_2 - u_3}{2} \right)^+ \right] dt. \quad \text{(A.12)}$$

We now argue that the integrand is always non-negative and not more than 1. At every time between 0 and T, either both terms are zero, exactly one is nonzero or both are nonzero. In the first two cases, clearly the integrand is bounded above by 1. In the last case, $u_1(t) \ge u_2(t) \ge u_3(t)$, and so the integrand collapses to $(u_1 - u_3/2) \ge 0$ which is at most 1. Thus, $2t_0 \le T \le T_{\min}(3, d)$, $(u_1 - u_3)/2$ which establishes Theorem 3. □

References

[1] R. V. L. Hartley, "Transmission of information," *Bell Syst. Tech. J.*, 7, no. 3, pp. 535–563, July 1928.

[2] W. L. Root, "Estimates of ϵ-capacity for certain linear communication channels," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 361–369, May 1968.

[3] A. D. Wyner, "A bound on the number of distinguishable functions which are time-limited and approximately band-limited," *SIAM J. Appl. Math.*, vol. 24, no. 3, 1973.

[4] H. L. Royden, *Real Analysis*. New York: MacMillan, 1968.

[5] M. Athans and P. L. Falb, *Optimal Control*. New York: McGraw-Hill, 1966.

[6] M. Honig, K. Steiglitz, S. Boyd, and B. Gopinath, "Bounds on maximum throughput for digital communications with finite precision and amplitude constraints," *IEEE Trans. Inform. Theory*, vol. 36, pp. 472–484, May 1990.

Michael L. Honig (S'80–M'81) was born in Phoenix, AZ, in 1955. He received the B.S. degree in electrical engineering from Stanford University, Stanford, CA, in 1977, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1978 and 1981, respectively.

From July 1981 to October 1983 he was a Member of the Technical Staff at AT&T Information Systems, formerly part of Bell Laboratories, Holmdel, NJ, where he worked on the design and performance analysis of local area networks, and on voice-band data transmission. He subsequently transferred to the Systems Principles Research Division at Bell Communications Research, where he is currently working in the areas of data communications and signal processing.

Dr. Honig is a member of Tau Beta Pi and Phi Beta Kappa, and is currently an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS.

Stephen P. Boyd (S'82–M'85) received the A.B. degree in mathematics, summa cum laude, from Harvard University, Cambridge, MA, in 1980, and the Ph.D. degree in electrical engineering, and computer science from the University of California, Berkeley, in 1985.

Since 1985 he has been an Assistant Professor in the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA.

B. Gopinath (S'64–M'74) was born in Chengannoor, India, in 1944. In 1964, he received the Master of Science degree in mathematics and physics from the University of Bombay. In electrical engineering, he received Master's and Ph.D. degrees in 1965 and 1968, respectively, from Stanford University, Stanford, CA.

In 1968, he joined Bell Telephone Laboratories at Murray Hill, NJ, as a Member of Technical Staff in the Mathematics Research Center, and continued there until the breakup of AT&T in 1983. He then became the manager of Communications Principles Research Group at Bellcore and later, in 1985, became the Division Manager of Systems Principles Research Division. He has taught at the University of California, Berkeley (as a Gordon McKay Professor in 1980), Columbia University, New York (in 1987), and the University of Goettingen, Germany (as an Alexander von Humbolt Fellow in 1972). He is presently a Professor of Electrical and Computer Engineering at Rutgers University, New Brunswick. His current research interests are panoramic audio-visual environments, parallel object-oriented languages and databases, supercomputing, and signal processing.

Erik Rantapaa was born in St. Paul, MN, in 1965. He received the B.S. degree in mathematics from the University of Minnesota in 1990, and is currently pursuing the Ph.D. degree in mathematics at the University of Minnesota. His current research is in combinatorics.