Optimization of Signal Sets for Partial-Response Channels—Part II: Asymptotic Coding Gain

Michael L. Honig, *Member, IEEE*

***Abstract*—For a linear, time-invariant, discrete-time channel with transfer function $H(f)$, and information rate R bits/T, where T is the symbol interval, an optimal signal set of length K is defined to be a set of 2^{RK} inputs of length K that maximizes the minimum l_2 distance between pairs of outputs. This paper studies the minimum distance between outputs, or equivalently, the coding gain of optimal signal sets as $K \to \infty$. For large K this coding gain, relative to single-step detection, can approximately be decomposed into the coding gain of an optimal signal set of length K for the identity channel, plus the gain of a "baseline" coding scheme for the channel $H(f)$. The baseline signal set is selected from the multidimensional integer lattice, where the basis vectors of the space are taken to be the eigenvectors of $H'H$, and H is the Toeplitz matrix that maps channel inputs to channel outputs. The coding gain of the baseline scheme can be computed explicitly as $K \to \infty$ in terms of $|H(f)|$ and R. The minimum distance between channel outputs for optimal signal sets as $K \to \infty$ is determined by the ϵ-rate of the channel. Existing upper and lower bounds on the ϵ-rate are used to compute bounds on the maximum asymptotic coding gains achievable for some partial response channels. These asymptotic coding gains are compared with the coding gains corresponding to signal sets found by numerical optimization techniques. A comparison of bounds on ϵ-rates for the identity and $1-D$ channels indicates that for a given large K, the squared minimum distance of an optimal signal set for the $1-D$ channel is 2 dB more than the squared minimum distance of an optimal signal set for the identity channel at a rate of 1 bit/T. For rates greater than 2 bits/T, however, this comparison indicates that optimal signal sets of length K for these two channels have nearly the same minimum distance.**

***Index Terms*—Coding, partial-response channels, intersymbol interference, ϵ-rate, multidimensional signal sets.**

I. Introduction

Given an arbitrary channel that maps inputs to outputs, the objective in l_2/l_2 signal design [1] is to construct N inputs of length K, where the inputs are bounded in l_2, to maximize the minimum l_2 distance between pairs of channel outputs. Motivation for this type of signal design arises from situations where it is difficult to characterize additive noise, or disturbances, to the received signal statistically, but it is safe to assume that these disturbances are bounded almost surely in l_2. Here we consider only channels that can be accurately modeled as linear, time-invariant, discrete-time systems. In Part I of this paper [1], numerical techniques for finding locally optimal solutions to the l_2/l_2 signal design problem were presented and used to find good signal sets, or codes, for the identity, $1-D$, and $1-D^2$ channels. Here, we study the maximum achievable minimum distance between channel outputs for large input lengths K.

It was shown in [1] that the l_2/l_2 signal design problem with N inputs of length $K < N$ is equivalent to packing N points in an ellipsoid in $\mathbb{R}^K$ to maximize the minimum Euclidean distance between pairs of points. Each point corresponds to one of the N channel outputs and the collection of points is referred to as the "output signal constellation." The axes of the ellipsoid are coincident with the orthogonal eigenvectors of the channel linear operator, and the lengths of the axes are the corresponding singular values. Strictly speaking, this interpretation applies only to the *hard input constraint* (HIC) problem, in which each of the N inputs is assumed to be bounded in l_2 norm. For moderate to large input rates (i.e., ≥ 2 bits/T), however, this interpretation also applies to the *soft input constraint* (SIC) problem, in which the average l_2 norm over the entire input signal set is bounded.

In Section II, after reviewing notation in Section II, we show how to estimate the coding gain, relative to single-step detection, of an optimal signal set of length K when K is large. This coding gain can be decomposed into two additive components. The first component is the gain of the optimal signal set relative to a K-dimensional signal constellation constructed from the integer (cubic) lattice [2] within a cube. The second component is the gain of this "cubic," or *baseline* signal constellation, where the basis vectors of the signal space are taken to be the channel eigenvectors, relative to single-step detection. The latter gain depends only on the channel transfer function and information rate, and can be explicitly computed. For large information rates R, nearly optimal signal sets can be constructed from dense lattices [1]. The former gain can therefore be decomposed further into the gain due to the dense lattice, plus the shaping gain obtained by selecting the input constellation points from within the sphere rather than the cube [2]–[6]. The spectrum of the baseline constellation is also computed as $K \to \infty$, and for large R is shown to be approximately constant over a subset of the channel bandwidth.

Manuscript received December 10, 1989; revised March 26, 1991. This work was presented in part at the 1990 Conference on Information Sciences and Systems, Princeton University, Princeton, NJ, March 1990.
The author is with Bellcore, Room MRE 2L-343, 445 South Street, Morristown, NJ 07960-0910.
IEEE Log Number 9101605.

0018-9448/91/0900-1342$01.00

In Section IV, the numerical results in [1], consisting of coding gains for different rates and input lengths, are compared to upper and lower bounds on ϵ-rate [7], [8], which for a linear, time-invariant channel and given information rate specifies the maximum achievable minimum distance of a signal set, or coding gain, as $K \to \infty$. It is shown that the bounds on ϵ-rate given in [7]–[10], which assume the HIC, also apply to ϵ-rate assuming the SIC; however, an open question is whether or not ϵ-rate assuming the HIC is always equal to ϵ-rate assuming the SIC.

A comparison of ϵ-rates is given for the identity, $1-D$, and $1-D^2$ channels. For a required minimum distance between outputs, this comparison indicates that longer codes are needed for the identity channel at low rates (i.e., at 1 bit$/T$), but that at high rates (i.e., 2 bits$/T$ or more) the lengths of optimal signal sets that achieve a given minimum distance for these channels are nearly the same. This conclusion depends, however, on how the channels are normalized. That is, if the PR channels are normalized by the energy in the impulse response, then codes for PR channels must be longer than codes designed for the identity channel. Plots showing bounds on maximum achievable coding gain divided by input length, as $K \to \infty$, vs. information rate are also shown for the preceding PR channels.

II. Notation

Given an input vector $\boldsymbol{u}$, the channel output is

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{u}, \tag{2.1}$$

where $\boldsymbol{H}$ is the convolution matrix formed from the channel impulse response $h[k]$. It is assumed that $h[k]=0$ for $k<0$ and $k>\tau-1$, where $\tau<\infty$, and that $|h[k]\|<\infty$ for $k=0,\cdots,\tau-1$. If the input in (2.1) has length K, then $\boldsymbol{H}$ is a $(K+\tau-1)\times K$ matrix, and $\boldsymbol{y}$ has dimension $K+\tau-1$.

Given a set of inputs $\boldsymbol{u}_i$, $i=1,\cdots,N$, then the minimum distance between pairs of outputs is

$$d = \min_{i \neq j} \| \boldsymbol{H}(\boldsymbol{u}_i - \boldsymbol{u}_j) \|, \tag{2.2}$$

where the norm is the l_2 (Euclidean) norm evaluated over the time interval $[1, K+\tau-1]$. Since $\boldsymbol{H}'\boldsymbol{H}$ is symmetric and Toeplitz, we can write

$$\boldsymbol{H}'\boldsymbol{H} = \Phi\Lambda\Phi', \tag{2.3}$$

where Φ is the $K \times K$ orthonormal matrix whose columns are eigenvectors of $\boldsymbol{H}'\boldsymbol{H}$, and

$$\Lambda = \operatorname{diag}[\lambda_1, \lambda_2, \cdots, \lambda_K], \tag{2.4}$$

where λ_k, $k=1,\cdots,K$, are the eigenvalues of $\boldsymbol{H}'\boldsymbol{H}$, assumed to be arranged in nonincreasing order. We assume that $\boldsymbol{H}'\boldsymbol{H}$ is nonsingular, so that these eigenvalues are real and strictly positive.

Letting $\tilde{\boldsymbol{u}}_i = \Phi'\boldsymbol{u}_i$, then d can be rewritten as

$$d = \min_{i \neq m} \| \Lambda^{1/2}(\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_m) \|. \tag{2.5}$$

The objective in l_2/l_2 signal design is to maximize d subject to the *hard input constraint* (HIC),

$$\|\boldsymbol{u}_i\|^2 = \|\tilde{\boldsymbol{u}}_i\|^2 \le PK, \qquad i=1,\cdots,N, \tag{2.6}$$

where P is the average transmitted power. We also consider the more conventional *soft input constraint* (SIC),

$$\frac{1}{N}\sum_{i=1}^{N} \|\tilde{\boldsymbol{u}}_i\|^2 \le PK. \tag{2.7}$$

A set of N vectors $\tilde{\boldsymbol{u}}_1,\cdots,\tilde{\boldsymbol{u}}_N$ will be referred to as the "input signal constellation." The set of vectors $\tilde{\boldsymbol{y}}_i = \Lambda^{1/2}\tilde{\boldsymbol{u}}_i$, $i=1,\cdots,N$, will be referred to as the "output signal constellation," and "signal constellation" will be used when a specific reference to input or output points is unnecessary. If the $\tilde{\boldsymbol{u}}_i$'s are within a sphere of radius r in K dimensions, then the output signal constellation is bounded by an ellipsoid with axes having lengths $r\lambda_1^{1/2},\cdots,r\lambda_K^{1/2}$.

Given any signal constellation, its *coding gain* is defined as

$$CG = 10\log_{10}\left(\frac{d^2/P}{d_{ss}^2/P_{ss}}\right) \tag{2.8}$$

in dB, where d_{ss} and P_{ss} are the minimum distance and average power, respectively, assuming single-step detection. Single-step detection refers to a one-dimensional, multilevel signaling scheme in which the receiver makes a decision on a (scalar) transmitted symbol based on a single channel output. In this case the normalized squared distance between outputs is

$$\frac{d_{ss}^2}{P_{ss}} = \frac{12\alpha^2}{4^R - 1}, \tag{2.9}$$

where R is the information rate in bits$/T$, $1/T$ being the symbol rate, and α is the ratio of the minimum distance between any two possible outputs at a particular time to the spacing between input levels. For the $1-D$ and $1-D^2$ channels $\alpha = 1$.

Strictly speaking, (2.9) makes sense only when $R = \log_2 L$, where L is a positive integer representing the number of levels that can be transmitted at each symbol interval. However, we will use (2.9) to compute the coding gain of signal sets at arbitrary values of R, which may correspond to noninteger values of L.

It is also of interest to compare the performance of coding schemes with that of a system with an ideal decision feedback equalizer (DFE). That is, the transmitter again transmits one of $L > 2$ uniformly spaced levels per baud (T), but the receiver now makes a decision on the (scalar) output of an ideal DFE, which is assumed to eliminate intersymbol interference caused by the channel impulse response coefficients $h[k]$, $k \neq 0$. In this case the ratio of minimum distance between two possible outputs to the spacing between input levels is $h[0]$. Note that $h[0]$ can be substituted for α in (2.9) to give the normalized minimum distance as a function of R. For the PR channels used to generate numerical results in this paper

$\alpha = h[0] = 1$, so that for these PR channels (2.9) also applies to the ideal DFE system just described. Consequently, the computed coding gains that appear in the paper are with respect to either single-step detection or the ideal DFE system.

III. A Baseline Signal Constellation (BSC)

In this section we estimate the coding gain, relative to single-step detection, of an input signal constellation obtained by selecting points from a scaled dense lattice within a sphere. This estimate can be accomplished in two steps, namely, first compute the gain of the signal constellation with respect to a baseline signal constellation (BSC), and then the gain of the BSC with respect to single-step detection. The BSC is defined so that for large R, the gain in the first step can be partitioned into the gain due to the dense lattice, plus the shaping gain obtained from selecting the points in the input space from within the sphere rather than the cube, which has been studied elsewhere [2]–[6]. As pointed out in [1], and in [11], a trellis code can be defined with respect to an output signal constellation, as defined here. The coding gain of the trellis code is then added to the gain of the signal constellation. We add that computations of coding gains for PR channels, assuming different coding schemes from the one considered here, also appear in [11] and [12].

To motivate the choice of BSC consider first the identity channel. The baseline signaling scheme for this channel to which coding schemes are typically compared is multilevel signaling, in which a transmitted symbol at each time instant is chosen independently from a set of uniformly spaced, discrete levels centered at the origin. A block of M transmitted symbols therefore corresponds to a point selected from a section of an M-dimensional integer lattice, which is bounded by a cube, and is centered at the origin.

To generalize this to PR channels, consider the same constellation, where the basis vectors of the signal space are taken to be the channel eigenvectors. That is, each constellation point corresponds to an input vector $\tilde{\boldsymbol{u}}_i$. If the input set $\{\tilde{\boldsymbol{u}}_i\}$ lies within the unit cube in $\mathbb{R}^n$, then the output signal constellation $\{\tilde{\boldsymbol{y}}_i\}$ lies within a rectangular parallelepiped in $\mathbb{R}^n$. Notice also that if the input points are uniformly spaced with minimum distance d along each axis, then the outputs are uniformly spaced along the ith axis with minimum distance $\lambda_i^{1/2}d$. What is desired, however, is that the output signal constellation be uniformly spaced along each axis with the *same* minimum distance. Consequently, the input signal constellation must be uniformly spaced along each axis, but the ratio of minimum distance between points, or signal levels, along the ith axis relative to the minimum distance between points along the first axis must be $(\lambda_1/\lambda_i)^{1/2}$. Alternatively, the *density* of points along the ith axis relative to the density of points along the first axis is $(\lambda_i/\lambda_1)^{1/2}$. More signal levels, or bits, are therefore allocated to those dimensions associated with larger eigenvalues.

The input BSC is therefore obtained from the integer lattice by "stretching" the ith dimension by the factor $(\lambda_1/\lambda_i)^{1/2}$. The maximum magnitude of each coordinate, or signal level, is assumed to be less than or equal to $A/2$. The input BSC is therefore bounded by a cube of length A on each side, and centered at the origin. Note, however, that if p_1 is the number of points along the first axis, associated with λ_1, then the number of points along the ith axis should be $(\lambda_i/\lambda_1)^{1/2}p_1$, which is generally not an integer.

To ensure that an integer number of points are assigned to each dimension, the input BSC is defined as follows. We start with the stretched integer lattice, as previously defined. Next, select p_1 to be some positive integer greater than or equal to two. The choice of p_1 determines the rate in bits/T, as will be seen shortly. The number of points along the ith axis is then defined to be $p_i = \lfloor(\lambda_i/\lambda_1)^{1/2}p_1\rfloor$, provided that this number is not less than two, where $\lfloor x \rfloor$ is the greatest integer less than or equal to x. If $(\lambda_i/\lambda_1)^{1/2}p_1 < 2$, then $p_i = 0$, implying that this dimension is not used. Finally, the entire constellation is scaled to satisfy the average energy constraint.

As defined, the input BSC lies within a cube in $\mathbb{R}^n$, $n \le K$, centered at the origin, with side lengths equal to $A = (p_1 - 1)d/\lambda_1^{1/2}$. This is because p_1 uniformly spaced points along a line with minimum distance ξ results in a distance of $(p_1 - 1)\xi$ between the furthest points, and the minimum distance between the points along the ith axis in the input BSC is $d/\lambda_i^{1/2}$. The *output* BSC is therefore a section of the integer lattice with minimum distance d that is contained within a parallelepiped centered at the origin, where the length of the side along the ith axis is $A\lambda_i^{1/2}$. The definition of p_i implies that the length of this ith side when $i \ne 1$ is generally greater than $d(p_i - 1)$, the distance between furthest points along the ith axis. Note, however, that any point in the integer lattice that is not in the output BSC must lie outside of this parallelepiped.

By definition, $p_i = 0$, if and only if $A\lambda_i < d$. That is, a dimension is "discarded," if and only if λ_i is small enough so that two points separated by A in the input constellation along the ith axis are less than d apart in the output space. Since the λ_i's are nonincreasing with i, $p_n = 0$ therefore implies that $p_k = 0$, for $k > n$. That is, $[\tilde{\boldsymbol{u}}_i]_k = 0$ for $n \le k \le K$, and each $i = 1, \cdots, N$. We later show that this causes the transmitted energy, as $K \to \infty$, to be concentrated in a subset of the frequency band $[0,1]$ where the channel attenuation is relatively small.

For fixed K, the total number of points in the BSC is

$$N = \prod_{k=1}^{n} p_k(K), \tag{3.1a}$$

where

$$n = \max\left\{k : p_1\left(\frac{\lambda_k}{\lambda_1}\right)^{1/2} \ge 2\right\}, \tag{3.1b}$$

and is also a function of K. The information rate $(\log_2 N)/K$ is therefore determined solely by p_1 and the channel matrix $\boldsymbol{H}$. In general, this rate will be some irrational number. However, for large K, p_1 can be selected so that the resulting rate is close to some desired rate, and a small number of points can subsequently be added to or deleted from the BSC to achieve the desired rate without significantly changing the coding gain.

Another way to define a BSC is to fix p_i for some i, and define $p_j = \lfloor p_i(\lambda_j/\lambda_i)^{1/2} \rfloor$, for all j such that $p_j \geq 2$. The total number of points for a particular K, as given by (3.1), and the coding gain would then depend on i. However, for channels of interest this dependence will diminish as p_1 increases. One can also compute an approximation to the coding gain of a BSC by allowing noninteger values of p_i. That is, one could approximate $p_i = (\lambda_i/\lambda_1)^{1/2} p_1$. As will be seen, however, it is relatively simple to compute numerically the coding gain of the BSC previously defined without using this type of approximation.

We now compare the BSC just defined with the constellations proposed for vector coding [11]. In both cases, the constellations are taken from the integer lattice, where the basis vectors of the space are the eigenvectors of $\boldsymbol{H}'\boldsymbol{H}$. In vector coding, however, the number of points along each dimension (p_i for $1 \leq i \leq K$) is constrained to be a power of two, or zero. (For specific values of K, this is necessary if one is to use the constellation with known multidimensional trellis codes, such as those listed in [4].) The number of bits allocated to each dimension ($\log_2 p_i$, $i = 1, \cdots, K$) is then selected to minimize a cost function that approximates the average energy of the signal set. This type of optimization therefore imposes a more spherical shape on the boundary of the input constellation, as compared with the cubic shape of the input BSC defined here.

The BSC can be used to estimate the coding gain of optimal signal sets for large K. One attractive property of the BSC is that the average energy of the BSC can be computed explicitly as a function of p_1, assuming fixed minimum distance d between channel outputs. As $K \to \infty$, this average energy is easily evaluated in terms of the channel frequency response, and can be used to compute asymptotic coding gains of the BSC relative to single-step detection. This is in contrast to the computation of coding gain performed in [11], which relies on the approximation that the signal points are continuously distributed throughout the interior of the boundary region (see also [5]). This "continuous" approximation becomes inaccurate at low rates (i.e., $R \leq 1$ bit$/T$).

A. Decomposition of Coding Gain

For a given PR channel let $(d^2/P)_{\text{PR-SSD}}$ and $(d^2/P)_{\text{PR-OSS}(K)}$ denote normalized squared minimum distances for single-step detection (i.e., (2.9)), and an optimal signal set of length K, respectively. Then

$$\frac{(d^2/P)_{\text{PR-OSS}(K)}}{(d^2/P)_{\text{PR-SSD}}} = \frac{(d^2/P)_{\text{PR-OSS}(K)}}{(d^2/P)_{\text{BSC}(K)}} \frac{(d^2/P)_{\text{BSC}(K)}}{(d^2/P)_{\text{PR-SSD}}}, \tag{3.2}$$

where BSC (K) refers to the BSC of length K. The ratio to the far right can be computed explicitly. The middle ratio is the additional coding gain that can be obtained relative to the BSC. For example, one way to achieve additional coding gain is to use a dense lattice, "stretched" in each dimension by $(\lambda_1/\lambda_i)^{1/2}$, and to shape the input signal constellation so that the points are uniformly distributed throughout a sphere, which has the least average energy for fixed volume of any region. The gain due to selecting the input constellation points from within a sphere rather than a cube can be estimated by using the "continuous" approximation just described [5]. This gain approaches 1.53 dB from below as the number of dimensions increases to infinity. Techniques for achieving substantial shaping gains for moderate K are given in [6], and can also be applied to the BSC defined here.

Assume, then, that the coding gain of an optimal signal set of length K is accurately approximated by the coding gain of a constellation consisting of points from the densest lattice within a sphere in $\mathbb{R}^K$, scaled appropriately in each dimension. This was observed in [1], where signal sets obtained from scaled lattice constructions were often found to be nearly as good, if not better, as signal sets obtained from a gradient search algorithm. The resulting signal constellation will generally span $m < K$ dimensions [1]; however, we assume that starting with a K-dimensional lattice, rather than an m-dimensional lattice, is nearly optimal. As R increases, it can be shown by using a volume argument that m increases to K. (See [1] and Section IV of this paper.) Since the BSC is based on the scaled K-dimensional integer lattice, the middle ratio in (3.2) can be approximated as the gain of the densest K-dimensional lattice relative to the K-dimensional integer lattice, plus the shaping gain in K dimensions. According to our previous assumption, this is approximately the coding gain of an optimal signal set of length K for the identity channel. We, therefore, rewrite (3.2) as

$$CG_{\text{PR-OSS}(K)} \approx CG_{\text{ID-OSS}(K)} + CG_{\text{BSC}(K)}, \tag{3.3}$$

where $CG_{\text{ID-OSS}(K)}$ is the coding gain of an optimal signal set of length K for the identity channel, and the other terms are defined in the obvious way.

Although the validity of (3.3) has been argued only for the case of large R, the numerical results in this section and the next suggest that (3.3) is accurate for moderate values of R as well (i.e., $R \geq 1$ bit$/T$). For large K it will be convenient to approximate the right-most term in (3.3) by the asymptotic gain $\lim_{K \to \infty} CG_{\text{BSC}(K)} = CG^*_{\text{BSC}}$, which is easily computed given the channel transfer function $H(f)$. Numerical results in the next section indicate that this approximation is accurate for $K \geq 10$.

B. Asymptotic Coding Gain

We now compute the coding gain of the BSC with respect to single-step detection as $K \to \infty$. From (3.1) the asymptotic rate as a function of d is

$$\begin{aligned} R^* &= \lim_{K\to\infty} \frac{\log_2 N(K,d)}{K} \\ &= \lim_{K\to\infty}\left(\frac{1}{K}\sum_{i=1}^{n}\log_2 p_i(K)\right) \\ &= \lim_{K\to\infty}\left\{\frac{1}{K}\sum_{i=1}^{n}\log_2\left\lfloor p_1\left(\frac{\lambda_i}{\lambda_1}\right)^{1/2}\right\rfloor\right\} \text{ bits}/T, \quad (3.4)\end{aligned}$$

where $n(K)$ is given by (3.1b). Throughout this paper, the superscript "*" indicates that the associated variable is defined by letting $K \to \infty$.

To compute R^* for fixed p_1 we apply Szegö's theorem regarding the asymptotic distribution of eigenvalues of a Toeplitz matrix [13]. This theorem states that for any continuous function $F(x)$,

$$\lim_{K\to\infty}\left(\frac{1}{K}\sum_{k=1}^{K}F[\lambda_k(K)]\right) = \int_0^1 F\left[|H(f)|^2\right]df, \quad (3.5)$$

where $\lambda_1(K),\cdots,\lambda_K(K)$ are the eigenvalues of $\boldsymbol{H}'\boldsymbol{H}$ at time K arranged in decreasing order, and

$$H(f) = \sum_{k=0}^{\tau-1} h[k]e^{-j2\pi fk}. \quad (3.6)$$

Throughout this paper f denotes normalized frequency (i.e., the analog frequency times the symbol length T). Now the summand in (3.4) is not a continuous function of $\lambda_1,\cdots,\lambda_n$. However, for each K we can construct a sequence of continuous functions that converges to the summand almost everywhere. Applying (3.5) to this sequence of functions, and using the fact that the summand in (3.4) is a uniformly bounded function of the λ_i's for all K, it is easily shown that

$$\begin{aligned} R^* &= \int_{B(f;p_1)} \log_2\left\lfloor \frac{|H(f)|}{M}p_1\right\rfloor df \\ &= \sum_{i=2}^{p_1}(\log_2 i)\,\text{meas}\left(f: i \le \frac{|H(f)|}{M}p_1 < i+1\right), \quad (3.7)\end{aligned}$$

where

$$B(f;p_1) = \left\{f: \frac{|H(f)|}{M}p_1 \ge 2,\ 0 \le f \le 1\right\} \quad (3.8)$$

and

$$M = \lim_{K\to\infty}\lambda_1^{1/2}(K) = \sup_f |H(f)|. \quad (3.9)$$

The set $B(f;p_1)$ is the set of frequencies in which the signal energy is concentrated as $K \to \infty$.

To compute the average energy in the BSC, first note that the total energy in p_i uniformly spaced points centered around the origin in one dimension with minimum distance ξ_i is

$$E_i = \frac{\xi_i^2}{12}p_i(p_i^2-1). \quad (3.10)$$

Consequently, the total energy of the BSC is

$$\begin{aligned} E_{\text{tot}}(K) &= \sum_{i=1}^{N}\|\tilde{\boldsymbol{u}}_i\|^2 \\ &= \sum_{i=1}^{n}\left\{\frac{\xi_i^2}{12}p_i(p_i^2-1)\left(\prod_{k\ne i}p_k\right)\right\} \\ &= \left(\prod_{i=1}^{n}p_i\right)\left(\sum_{i=1}^{n}\frac{\xi_i^2}{12}(p_i^2-1)\right). \quad (3.11)\end{aligned}$$

The energy per point is therefore

$$\frac{E_{\text{tot}}(K)}{N(K,d)} = \frac{d^2}{12}\sum_{i=1}^{n}\frac{1}{\lambda_i}\left(\left\lfloor p_1(\lambda_i/\lambda_1)^{1/2}\right\rfloor^2 - 1\right), \quad (3.12)$$

where the fact that $\xi_i^2 = d^2/\lambda_i$ has been used. Since the summand is a piecewise continuous function of $\lambda_1,\cdots,\lambda_n$, and uniformly bounded for all K, we can again apply Szegö's theorem to get an expression for the asymptotic transmitted power,

$$\begin{aligned} P^* &= \lim_{K\to\infty}\frac{E_{\text{tot}}(K)}{KN(K,D)} \\ &= \frac{d^2}{12}\int_{B(f;p_1)}\frac{\lfloor p_1|H(f)|/M\rfloor^2 - 1}{|H(f)|^2}df \\ &= \frac{d^2}{12}\sum_{i=2}^{p_1}\left\{(i^2-1)\int_{\{f:\, i\le p_1|H(f)|/M<i+1\}}\frac{df}{|H(f)|^2}\right\}. \quad (3.13)\end{aligned}$$

For a given channel transfer function $H(f)$ and number of levels p_1, the asymptotic rate can therefore be computed from (3.7), and the asymptotic normalized squared minimum distance is

$$\frac{d^2}{P^*} = \frac{12}{\sum_{i=2}^{p_1}\left\{(i^2-1)\int_{\{f:\, i\le p_1|H(f)|/M<i+1\}}\frac{df}{|H(f)|^2}\right\}}. \quad (3.14)$$

As a simple example, consider the identity channel. In this case $|H(f)| = 1$ for $0 < f < 1$. From (3.7),

$$R^* = \log_2 p_1 \quad (3.15)$$

and from (3.14),

$$\frac{d^2}{P^*} = \frac{12}{p_1^2 - 1} = \frac{12}{4^{R^*}-1}, \quad (3.16)$$

which agrees with the single-step formula (2.9), since $\alpha = 1$.

Consider now the extreme cases $p_1 = 2$ and $p_1 \to \infty$. The former case gives the minimum value of R^*, and the latter case corresponds to $R^* \to \infty$. In the former case from (3.7),

$$R^*_{\min} = \text{meas}\left\{f: \frac{|H(f)|}{M} \geq 1\right\}, \qquad (3.17)$$

and from (3.14), when $R^*_{\min} > 0$,

$$\frac{d^2}{P^*} = \frac{12}{3R^*_{\min}/M^2} = \frac{4M^2}{R^*_{\min}}. \qquad (3.18)$$

Consequently, the asymptotic coding gain of the BSC in this case is

$$CG^* = \frac{M^2}{3\alpha^2}\frac{4^{R^*_{\min}} - 1}{R^*_{\min}}. \qquad (3.19)$$

As $R^*_{\min} \to 0$, $CG^* \to (M^2/3\alpha^2)\log_e 4$. For piecewise continuous $|H(f)|$, this expression can be used to estimate the asymptotic coding gain of the BSC for small rates even when $R^*_{\min}$ defined by (3.17) is zero. Specifically, if $R^*_{\min} = 0$ for given $|H(f)|$, we can instead consider the magnitude transfer function

$$|G(f)| = \begin{cases} |H(f)|, & \text{when } |H(f)| < M' < M, \\ M', & \text{when } |H(f)| \geq M'. \end{cases} \qquad (3.20)$$

Applying (3.7) and (3.14) to $|G(f)|$ when $p_1 = 2$, and letting $M' \to M$ again gives (3.19) where $R^*_{\min} \to 0$. For the $1-D$ channel, $\alpha = 1$ and $|H(f)|^2 = 2(1-\cos 2\pi f)$, so that $M^2 = 4$. The asymptotic coding gain of the BSC corresponding to the modified channel $|G(f)|$ given by (3.20) therefore approaches $(4\log_e 4)/3$, or 2.7 dB as $M' \to M$.

To evaluate the asymptotic coding gain of the BSC as $p_1 \to \infty$, first note that (3.7) implies that

$$\int_{p_1|H(f)|M > 2} \log_2\left(\frac{|H(f)|}{M}p_1 - 1\right) df \leq R^* \leq \int_{p_1|H(f)|/M \geq 2} \log_2\left(\frac{|H(f)|}{M}p_1\right) df. \qquad (3.21)$$

Assuming $\int_0^1 \log_2|H(f)|\, df < \infty$, it follows that

$$R^* = \int_0^1 \log_2\left(\frac{|H(f)|}{M}p_1\right) df + o(p_1), \qquad (3.22)$$

where $o(p_1)$ is a correction term that goes to zero as $p_1 \to \infty$. If $H(f)$ is minimum phase,[1] then it can be shown that [14]

$$\int_0^1 \log_2|H(f)|\, df = \log_2 h[0], \qquad (3.23)$$

where $h[0]$ is the first impulse response coefficient. Although the $1-D$ and $1-D^2$ channels are not minimum phase, (3.23) is also valid for these channels. Assuming (3.23) is valid, then (3.22) can be rewritten as

$$R^* = \log_2\frac{p_1 h[0]}{M} + o(p_1). \qquad (3.24)$$

[1]A transfer function $H(f)$ is minimum phase, if and only if all of its zeros and poles are inside the unit circle.

Using an argument analogous to that used to derive (3.22), it can be shown that (3.13) implies

$$P^* = \frac{d^2}{12}\int_0^1 \frac{1}{|H(f)|^2}\left[\left(\frac{|H(f)|}{M}p_1\right)^2 - 1\right] df + O(p_1)$$
$$= \frac{d^2}{12}\left[\left(\frac{p_1}{M}\right)^2 - \int_0^1 \frac{df}{|H(f)|^2}\right] + O(p_1) \qquad (3.25)$$

provided that $\int_0^1 df/|H(f)|^2 < \infty$, where $O(p_1)$ is a correction term satisfying

$$\lim_{p_1 \to \infty} O(p_1)/p_1^\kappa = 0, \qquad \text{for any } \kappa > 1.$$

Consequently,

$$\frac{d^2}{P^*} = 12\left(\frac{M}{p_1}\right)^2 + O(1/p_1^4) = 12h[0]^2 4^{-R^*} + O(16^{-R^*}). \qquad (3.26)$$

Comparing this with single-step detection gives

$$\lim_{p_1 \to \infty} CG^* = \lim_{R^* \to \infty} 12h[0]^2 4^{-R^*}\frac{4^{R^*} - 1}{12\alpha^2} = \left(\frac{h[0]}{\alpha}\right)^2. \qquad (3.27)$$

It is easily verified that $\int_0^1 df/|H(f)|^2 < \infty$ for the $1-D$ channel, so that (3.27) implies that the BSC gives 0 dB coding gain with respect to single-step detection as $R^* \to \infty$. For large rates, an optimal signal set designed for the $1-D$ channel therefore has approximately the same coding gain as an optimal signal set designed for the identity channel.

The asymptotic coding gain of the BSC relative to the ideal DFE system as $R^* \to 0$ and $R^* \to \infty$ is again given by (3.19) and (3.27), respectively, where α is replaced by $h[0]$. As $R^* \to \infty$, this asymptotic coding gain is again 0 dB. Note that if a code designed for the identity channel is applied to a PR channel with an ideal DFE, then the coding gain remains the same, assuming $h[0] = 1$. Since $CG^* = 0$ as $R^* \to \infty$, the ideal DFE system combined with codes designed for the identity channel offers the same performance as comparable codes (i.e., optimal signal sets) designed explicitly for the PR channel. This has also been observed in [11] and [15]. (See also [16]. Because the eigenvectors of $\boldsymbol{H}'\boldsymbol{H}$ become sinusoidal as $K \to \infty$ [1], the multitone transmission scheme considered in [16] is analogous to the BSC considered here.) However, the previous analysis indicates that at low rates a code can be designed for a PR channel that performs better than a comparable code designed for the identity channel and subsequently applied to this PR channel with an ideal DFE.

C. Numerical Results

Fig. 1 shows plots of asymptotic coding gain of the BSC vs. information rate for $(1-D)(1+D)^m$ channels, $m = 0, 1, 2, 3$. Each "curve" actually consists of a set of discrete

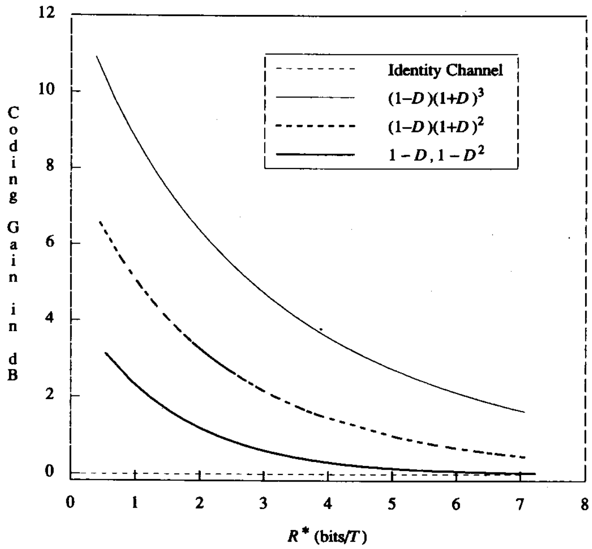

Fig. 1. Asymptotic coding gain of the BSC for some PR channels.

points corresponding to $p_1 = 3, 4, 5, \cdots$. In each case $p_1 = 2$ gives $R^* = 0$. The remaining points are close enough together so that smooth curves can be drawn. Note that the asymptotic coding gain for the $1 - D$ channel at the smallest value of R^* is not far from 2.7 dB, the approximation to low rate asymptotic coding gain given by (3.19). The asymptotic coding gain as $R^* \to \infty$ is zero, as predicted by (3.27).

Fig. 1 and the numerical results in [1] are now used to check the decomposition of coding gain (3.3). For the $1 - D$ channel Fig. 1 shows that the BSC gives an asymptotic coding gain of approximately 2.4 dB at $R^* = 1$ bit$/T$. Reference [1, Table II] shows a 4.38 dB coding gain corresponding to a 256/8 code ($N = 256$ points in eight dimensions) for the $1 - D$ channel (without padding). The corresponding 256/8 code for the identity channel gives a 2.10 dB gain (see [1, Table I]), and adding this to the BSC coding gain gives 4.5 dB, which is close to the preceding 4.38 dB coding gain. However, adding the gain of the 256/4 code for the identity channel (1.71 dB shown in [1, Table I]) to the gain of the $1 - D$ BSC at 2 bits$/T$ (1.19 dB) gives 2.9 dB, which is somewhat less than the 3.43 dB gain of the $1 - D$ 256/4 code, shown in [1, Table II]. This is presumably caused by the inaccuracy in estimating the gain of the four-dimensional BSC from Fig. 1. For large K and moderate to large information rates, the coding gains of optimal signal sets for PR channels can be estimated by adding the appropriate gains computed here for the BSC to the gains shown in [3] for dense lattices and optimal shaping. This will be demonstrated further in the next section.

As discussed in [1, Section IV], a particular signal set, or block code, can be "padded" with $\tau - 1$ zeros, where τ is the length of the channel impulse response, in order to eliminate intersymbol interference (ISI) between successive blocks of channel outputs. The coding gain with padding is, of course, less than the coding gain without padding, ignoring the effect of ISI in the latter case. When comparing coding gains of finite length signal sets with asymptotic gains, however, no padding is assumed, since d, defined by (2.2), is the minimum l_2-distance between pairs of outputs over the *entire* time interval in which the outputs are nonzero. Consequently, when estimating coding gains of optimal signal sets in finite dimensions, we assume no padding, although as the input length goes to infinity, the gains with and without padding become the same. Of course, we can still compare the coding gain of a signal set with padding to an asymptotic gain by modifying the corresponding information rate. For example, the 256/7, $1 - D$ signal set with padding (one zero is appended to each input vector) shown in [1, Table II] corresponds to an effective rate of 8/7 bits$/T$, even though in practice it can be used at the 1 bit$/T$ rate without interblock interference.

D. Asymptotic Spectrum of the BSC

The shape of the transmitted spectrum corresponding to an optimized signal set for a PR channel was discussed in [1]. It was shown there that for large rates R and large input length K, the spectrum of an input signal constellation in which the points are uniformly distributed throughout a region $\boldsymbol{R}$, which is invariant with respect to permutation of axes, is approximately white over a frequency band $F(f) = \{f\colon |H(f)| \geq \beta\}$, where β is a constant that depends on the minimum distance d. As R increases, $\beta \to 0$. It was also observed that the spectrum for $f \notin F(f)$ is primarily determined by $\boldsymbol{R}$, i.e., the shape of the constellation, rather than the particular lattice from which the points are chosen.

These results depend on an assumption concerning the asymptotic properties of the matrix Φ, defined by (2.3). Let $\hat{u}_i(f)$ denote the Fourier series associated with the input $\boldsymbol{u}_i$,

$$\hat{u}_i(f) = \sum_{k=1}^{K} u_i[k] e^{-j2\pi f(k-1)}, \qquad (3.28)$$

where $u_i[k] = [\boldsymbol{u}_i]_k$. The assumption is that the components of $\Phi \boldsymbol{u}_i$ and $\hat{u}_i[k/(2K+2)]$, $k = 1, \cdots, K$, have the same distribution for large K. This is based on the fact that for large K the eigenvectors of $\boldsymbol{H}'\boldsymbol{H}$, which are the columns of Φ, can be approximated by sinusoidal vectors at frequencies that are uniformly distributed over the interval $[0, 1/2]$. In this section, we use this assumption to compute the asymptotic spectrum of the BSC.

The preceding assumption implies that for large K the kth eigenvector of $\boldsymbol{H}'\boldsymbol{H}$ can be approximated as a sinusoidal vector at the frequency $k/(2K+2)$, and the corresponding eigenvalue, λ_k, is approximately $|H[k/(2K+2)]|^2$. In what follows, we therefore do not assume that the eigenvalues $\lambda_1, \cdots, \lambda_K$ are nonincreasing, but rather, that they behave as uniformly spaced samples of $|H(f)|^2$ for f in the interval $[0, 1/2]$.

For a given K-dimensional signal set the average input spectrum is defined as

$$S(f) = \frac{1}{KN} \sum_{i=1}^{N} |\hat{u}_i(f)|^2. \qquad (3.29)$$

The previous discussion implies that the components of $\tilde{\boldsymbol{u}}_i = \Phi \boldsymbol{u}_i$ behave as uniformly spaced samples of $\hat{u}_i(f)$, that is,

$$\hat{u}_i\left(\frac{k}{2(K+1)}\right) \approx \tilde{u}_i[k], \qquad k = 1, \cdots, K. \quad (3.30)$$

The spectrum at uniformly spaced frequencies for large K is therefore approximately

$$S\left(\frac{k}{2(K+1)}\right) \approx \begin{cases} \dfrac{1}{KN} \displaystyle\sum_{i=1}^{N} |\tilde{u}_i[k]|^2, & k = 1, \cdots, N, \\ 0, & n < k \leq K, \end{cases} \qquad (3.31)$$

where n, given by (3.1b), is the number of dimensions spanned by the $\tilde{\boldsymbol{u}}_i$'s. From (3.10)–(3.12) the average energy per point of the BSC with respect to the kth axis is

$$\frac{1}{KN} \sum_{i=1}^{N} |\tilde{u}_i[k]|^2 = \frac{E_k}{Kp_k} = \frac{d^2}{12K\lambda_k}\left(\left\lfloor p_1\left(\frac{\lambda_k}{\lambda_1}\right)^{1/2}\right\rfloor^2 - 1\right), \qquad (3.32)$$

assuming $p_1(\lambda_k/\lambda_1)^{1/2} \geq d$. If $p_1(\lambda_k/\lambda_1)^{1/2} < d$, then $S[k/(2K+2)] = 0$, since the corresponding dimension is not used. Letting $K \to \infty$, and applying Szegö's theorem as before gives

$$S(f) \to \begin{cases} \dfrac{c}{|H(f)|^2}\left[\left\lfloor \dfrac{|H(f)|}{M} p_1 \right\rfloor^2 - 1\right], & f \in B(f; p_1), \\ 0, & f \notin B(f; p_1), \end{cases} \quad \text{as } K \to \infty, \quad (3.33)$$

where c is chosen to satisfy the average power constraint $\int_0^1 S(f)\,df = P$, and $B(f; p_1)$ is defined by (3.8).

The sense in which $S(f)$ converges to the right hand side of (3.33) depends on the sense in which the eigenvectors of $\boldsymbol{H}'\boldsymbol{H}$ converge to sinusoidal vectors. Suppose $\boldsymbol{v}_k$ is a sinusoidal vector of length K with frequency $k/(2K+2)$, and let ϕ_k be the kth eigenvector of $\boldsymbol{H}'\boldsymbol{H}$. If $\|\boldsymbol{v}_k - \phi_k\| \to 0$ as $K \to \infty$ for each $k = 1, \cdots, K$, then the difference between the left and right sides of (3.31) converges to zero as $K \to \infty$ for each k. This would imply that $S(f)$ converges pointwise to the right-hand side of (3.33).

As before, we again examine the cases $R^* \to 0$ and $R^* \to \infty$. When $p_1 = 2$, $S(f)$ becomes a constant for all f such that $|H(f)| = M$, and is zero elsewhere. Using the substitution (3.20) and letting $M' \to M$, it is apparent that for small R^*, the spectrum of the BSC shrinks to a small subset of the channel bandwidth. As $p_1 \to \infty$, then any $f \in [0,1]$ for which $|H(f)| \neq 0$ is eventually contained in the set $B(f; p_1)$. Consequently, in this case (3.33) and the power constraint imply that $S(f)$ becomes a constant for all f where $|H(f)| \neq 0$. Note, however, that because the integer part of $p_1|H(f)|/M$ appears in (3.33), the asymptotic spectrum is discontinuous for finite p_1.

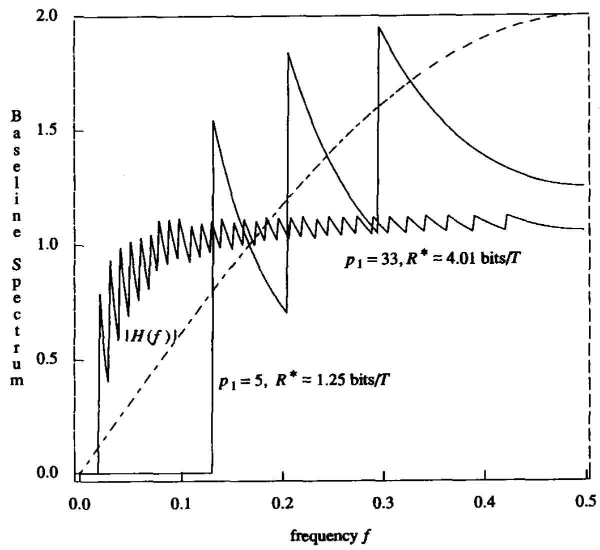

Fig. 2. Asymptotic spectra of the BSC for the $1-D$ channel at two different information rates, assuming $P = 1$.

The preceding discussion is illustrated in Fig. 2, which shows plots of average spectra of the BSC at two different information rates for the $1-D$ channel, assuming the average transmitted power $P = 1$. As predicted, as the information rate increases, the spectrum tends to a constant for frequencies where $|H(f)|$ is significantly greater than $2M/p_1$. In addition, there are discontinuities at frequencies f corresponding to jumps in $\lfloor p_1|H(f)|/M \rfloor$. As $R^*(p_1)$ increases, the number of discontinuities increases, and the size of the jumps decreases.

Intuitively, the transmitted spectrum associated with the BSC is undesirable because the discontinuities cause the transmitted energy to be unevenly allocated in frequency bands where the channel attenuation does not change significantly. Since the spectrum is determined by the average energy per dimension, $E_k/(Kp_k)$, this energy distribution can be improved by changing the distribution of points per dimension $p_1, \cdots, p_n$. Specifically, selecting $p_1, \cdots, p_n$ so that the constellation approximately lies within the sphere in n dimensions, rather than a cube, reduces the energy per constellation point, and smooths the resulting distribution of energy. This demonstrates the property observed in [1] that the transmitted spectrum is primarily determined by the shape of the constellation, rather than the lattice from which the points are chosen. This is simply because the density of any lattice in $\mathbb{R}^n$ is uniform throughout $\mathbb{R}^n$ (even after being scaled by the λ_k's in each dimension), and the average energy of a constellation with respect to each dimension is approximately independent of this density.

IV. Bounds on Asymptotic Coding Gain

Consider the generalization of the l_2/l_2 signal design problem in which N inputs, each restricted in l_q norm on the interval $[1, K]$, are to be designed to maximize the minimum l_p norm between pairs of channel outputs. Alternatively, we can fix the distance between outputs and the time interval $[1, K]$, and maximize N, the number of inputs. For a channel with impulse response $h[\cdot]$, let $N_{\max}(K, d)$ denote this maximum number of inputs. The *maximum channel throughput* for the channel $h[\cdot]$ is defined as $\mathrm{MCT}(d) = \lim_{K\to\infty}[\log_2 N_{\max}(K, d)]/K$, in bits/$T$. One interpretation of this quantity is the maximum asymptotic rate at which data can be reliably transmitted through the channel $h[\cdot]$, assuming an additive noise source about which nothing is known except that the l_p norm of the noise is bounded by $d/2$ almost surely. In the case of the l_2/l_2 problem, the MCT has been called "ϵ-capacity," or "ϵ-rate" [7], [8], where ϵ refers to the minimum distance between channel outputs. Since we have used d to denote minimum distance between channel outputs, we will refer to this quantity as the "d-rate" of the channel, and denote it as $C(d)$.

The d-rate can be defined assuming either the HIC or SIC, i.e.,

$$\|\boldsymbol{u}_i\|^2 \le P, \qquad \text{for each } i = 1, \cdots, N \qquad \text{(HIC)}$$

or

$$\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{u}_i\|^2 \le P. \qquad \text{(SIC)}$$

These input constraints do not allow the input energy to grow with time, as in (2.6) and (2.7). Alternatively, we can adopt either constraint (2.6) or (2.7), but then the squared minimum distance d^2 must be replaced by d^2K. That is, if the input energy is allowed to increase with K, then the minimum distance must also increase with K if the d-rate is to be finite for channels of interest. In what follows we assume the input constraints (2.6) and (2.7), so that the argument of $C(\cdot)$ is *time-normalized* minimum distance, denoted as $\bar{d}$, and refers to minimum distance between channel outputs normalized by $\sqrt{K}$, i.e., $\bar{d}^2 = d^2/K$.

For a given R and channel $|H(f)|$, it will be useful to define an *asymptotic* time-normalized minimum distance,

$$\bar{d}^*(R) = \lim_{K\to\infty}\bar{d}(K, R), \qquad (4.1)$$

assuming the limit exists, where $\bar{d}(K, R)$ is the time-normalized minimum distance of an optimal signal set of length K consisting of 2^{RK} points. For fixed R, there exists some $\bar{d}_0 \ge 0$ for which $R = C(\bar{d}_0)$. By the definition of d-rate, this means that the minimum distance of optimal signal sets grows asymptotically as $\bar{d}_0\sqrt{K}$, so that $\bar{d}_0 = \bar{d}^*(R)$. For large K, the d-rate can therefore be used to estimate the coding gain of an optimal signal set. In this section we compare the time-normalized minimum distance of signal sets, or block codes, found in [1] with $\bar{d}^*$ computed from the corresponding d-rates. We also compare d-rates for different channels. This comparison gives the relative input lengths of optimal signal sets for each channel needed to achieve asymptotically a specified minimum distance between outputs.

Upper and lower bounds on $C(\bar{d}^*)$ assuming the HIC are given for continuous-time channels in [7]–[10]. The analogous bounds for discrete-time channels are easily obtained, so that we omit the details. In general, the best lower bound on d-rate for a linear, time-invariant channel is given in [7], and is based on a simple volume argument. Modification of the bound in [7] so as to apply to discrete-time channels gives

$$C(\bar{d}^*) \ge \underline{C}(\bar{d}^*) = \int_{B(f)} \log_2\left(\frac{\sqrt{P}\,|H(f)|}{\bar{d}^*}\right) df, \qquad (4.2a)$$

where

$$B(f) = \left\{f: \sqrt{P}\,|H(f)| > \frac{\bar{d}^*}{2},\ 0 \le f \le 1\right\}. \qquad (4.2b)$$

This bound was derived assuming the HIC. However, since any signal set that satisfies the HIC also satisfies the SIC, it follows that $C(\bar{d}^*)$ assuming the HIC is less than or equal to $C(\bar{d}^*)$ assuming the SIC, so that (4.2) is valid assuming either constraint.

The best available upper bound on $C(\bar{d}^*)$ is given in [10], and is a modification of an upper bound originally given in [9]. This upper bound states that $C(\bar{d}^*)$ is less than or equal to the Shannon capacity of a power-limited channel consisting of the original channel, $H(f)$, followed by additive Gaussian noise with variance $\bar{d}^{*2}/4$, but otherwise having arbitrary power spectral density. It is shown in [10] that the noise spectrum that minimizes this bound is proportional to the input power spectral density, so that combining these two results gives

$$C(\bar{d}^*) \le \bar{C}(\bar{d}^*) = \frac{1}{2}\int_0^1 \log_2\left(\frac{4P|H(f)|^2}{\bar{d}^{*2}} + 1\right) df. \qquad (4.3)$$

Now the Shannon capacity of the additive Gaussian noise channel can be defined using either the HIC or the SIC, and is the same in either case [17]. Consequently, the bound (4.3) is valid assuming either constraint.

The preceding discussion naturally leads one to ask whether or not $C(\bar{d}^*)$ assuming the HIC is always equal to $C(\bar{d}^*)$ assuming the SIC. It appears that this problem is open; however, the following indicates that they are the same as the information rate $R \to \infty$. Assume that the optimal input constellation is uniformly distributed inside an n-dimensional sphere S with radius r. Note that the densities along each axis can be different; however, it is assumed that the density of points in n-space is a constant, ρ. The average energy per point of this signal constellation is approximately given by

$$\frac{E_{\text{tot}}}{N} \approx \frac{\int_S \rho\|\boldsymbol{x}\|^2\,d\boldsymbol{x}}{\int_S \rho\,d\boldsymbol{x}} = \frac{n}{n+2}r^2, \qquad (4.4)$$

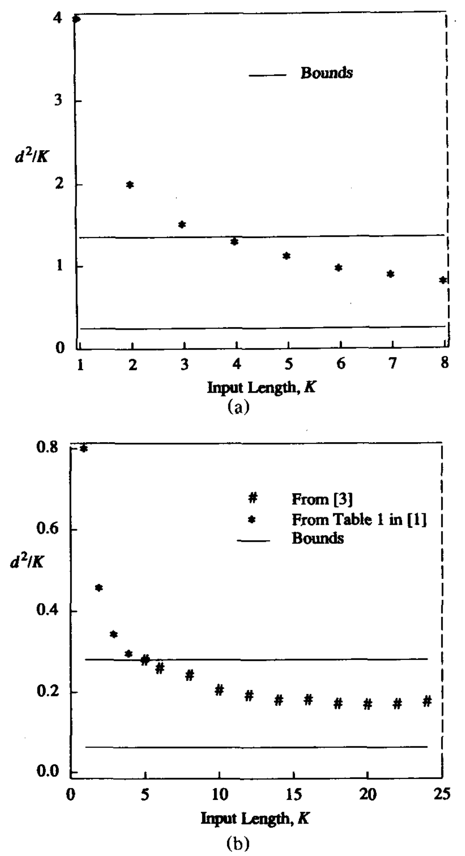

Fig. 3. (a) Normalized squared distance vs. input length for the identity channel signal sets represented in [1, Table I], corresponding to a rate of 1 bit/T. Also shown are upper and lower bounds on $\bar{d}^{*2}$. (b) Normalized squared distance vs. input length for identity channel signal sets at a rate of 2 bits/T. Bounds on $\bar{d}^{*2}$ are also shown.

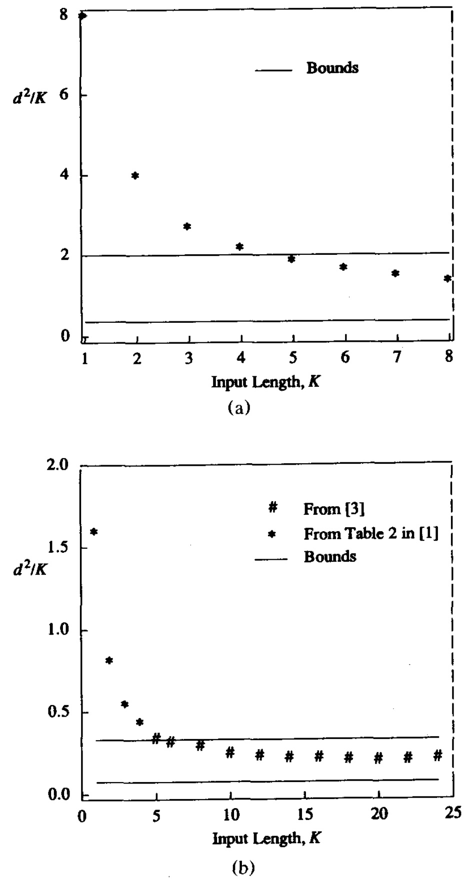

Fig. 4. (a) Normalized squared distance vs. input length for the $1-D$ channel codes in [1, Table II] (without padding) at the rate 1 bit/T. Also shown are upper and lower bounds on $\bar{d}^{*2}$. (b) Normalized squared distance vs. input length for $1-D$ channel signal sets shown in [1, Table II] at a rate of 2 bits/T. Bounds on $\bar{d}^{*2}$ are also shown.

which is independent of the density ρ. As $R \to \infty$, this expression becomes exact. If the HIC is assumed, then $r = \sqrt{PK}$, that is, each point has energy less than or equal to PK. If the SIC is assumed, then (4.4) implies that $r^2 \approx [(n+2)PK]/n$, which converges to PK as $n \to \infty$. Consequently, as $n \to \infty$, the number of points in the signal constellation assuming the SIC is the same as the number of points assuming the HIC, subject to the preceding integral approximation. Therefore, according to this approximation $C(\bar{d}^*)$ is the same in either case.

Figs. 3 and 4 show d^2/K vs. input length K for some of the signal sets found by the numerical search techniques discussed in [1, Section III]. Also shown for each channel are lower and upper bounds on the maximum asymptotic normalized distance obtained from (4.2) and (4.3), respectively. The code search results shown in Figs. 3(b) and 4(b) have been augmented by including additional points representing estimates of coding gains obtained from using dense lattices in higher dimensions than those considered in [1, Section IV]. These estimates are obtained by adding the coding gain of a dense lattice in dimension K, given in [3], to the corresponding asymptotic coding gain of the BSC computed in Section III. The shaping gain obtained by selecting the points from within a sphere rather than a cube in $\mathbb{R}^K$ is also included. The figures indicate that d^2/K approaches its asymptote for $K \geq 10$.

Fig. 5 shows plots of upper and lower bounds on $C(\bar{d}^*)$ vs. $4P/\bar{d}^{*2}$ in dB for the PR channels $1-D$, $1-D^2$, and $(1-D)(1+D)^2$, as well as for the identity channel. The bounds are the same for the $1-D$ and $1-D^2$ channels. For the channels considered, the bounds on d-rate increase with the energy in the channel impulse response. For fixed $4P/\bar{d}^{*2}$, the bounds on d-rates shown in Fig. 5 for the PR channels are therefore greater than the corresponding bounds on d-rate for the identity channel. This implies that asymptotically, the minimum distance of an optimal signal set for the identity channel is less than the minimum distance of an optimal signal set having the same input length and designed for one of these PR channels. For example, at a rate of 1 bit/T the bounds in

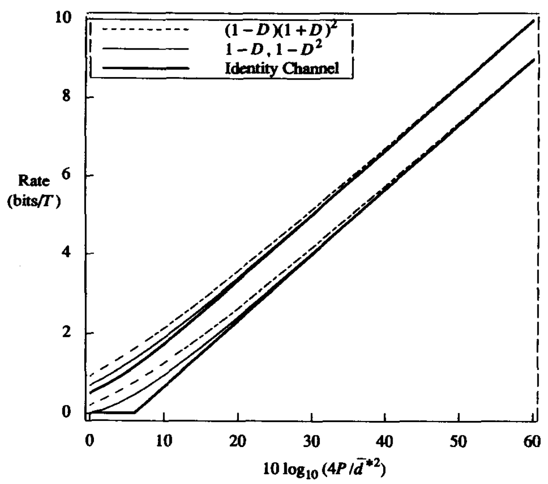

Fig. 5. Upper and lower bounds on d-rates vs. $4P/\bar{d}^{*2}$ in dB for some PR channels.

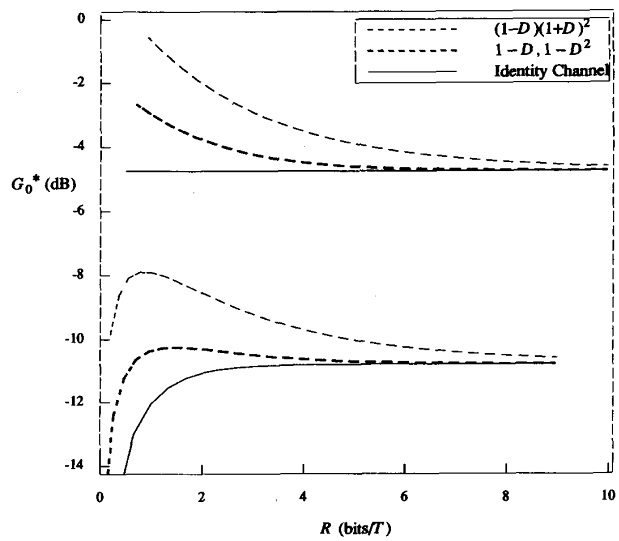

Fig. 6. Bounds on asymptotic coding gain G_0^* vs. rate for the identity, $1-D$, $1-D^2$, and $(1-D)(1+D)^2$ channels.

Fig. 5 indicate that normalized minimum distance for the identity channel is approximately 2 dB less than the normalized minimum distance for the $1-D$ channel. However, this difference decreases to zero as R increases.

To elaborate further, suppose that $R=1$ bit$/T$. Fig. 5 shows that $10\log_{10}[\bar{d}^{*2}/(4P)]$ for the identity channel is less than or equal to -4.7 in this case. Consequently, for large enough K, $10\log_{10}[\bar{d}^{*2}/(4P)]$ for any identity channel block code is upper bounded by $-4.7+10\log_{10}K$. For the $1-D$ channel, the corresponding upper bound is $-3.0+10\log_{10}K$. Equating these two quantities indicates that the input length of the identity channel signal set must be approximately 1.5 times longer than the input length for the $1-D$ channel signal set to achieve the same minimum distance. Using the lower bounds in Fig. 5 gives approximately the same result. At a rate of 2 bits$/T$, the bounds on d-rate indicate that $\bar{d}^{*2}/(4P)$ is approximately the same for both the $1-D$ and identity channels.

The bounds on d-rate can also be used to compute bounds on the coding gain of optimal signal sets relative to single-step detection as $K\to\infty$. Specifically, for fixed R and K, the coding gain of an optimal signal set is given by

$$CG = 10\log_{10}\frac{\bar{d}^2KP_{ss}}{Pd_{ss}^2} = G_0 + 10\log_{10}K, \qquad (4.5)$$

and from (2.9)

$$G_0 = 10\log_{10}\frac{\bar{d}^2P_{ss}}{Pd_{ss}^2} = 10\log_{10}\frac{(4^R-1)\bar{d}^2}{12\alpha^2P}. \qquad (4.6)$$

Let $G_0^* = \lim_{K\to\infty}G_0(K,R) = 10\log_{10}[(4^R-1)\bar{d}^{*2}/(12\alpha^2P)]$. Then upper and lower bounds on the coding gain G_0^* can be obtained from the preceding upper and lower bounds on $\bar{d}^*$, where $C(\bar{d}^*)=R$.

Upper and lower bounds on G_0^* vs. R are shown in Fig. 6 for some PR channels. (Note that these results are independent of how the channels are normalized, since a normalization factor would multiply both $\bar{d}^2$ and d_{ss}^2, leaving the coding gain unchanged.) These curves indicate that for small rates (i.e., less than 4 bits$/T$), optimal signal sets for the identity channel give less coding gain than optimal signal sets for the PR channels considered. For example, at a rate of 2 bits$/T$, the upper bounds for the identity and $1-D$ channels differ by approximately 1.5 dB. The lower bounds differ by about 1 dB.

Note that the bounds on coding gain for the identity channel are independent of R for large R. This observation is anticipated by results in [3] and [4], if it is assumed that for any large integer M, there is a $K>M$ such that a nearly optimal signal set of length K can be obtained by selecting a section of a dense lattice within a sphere. (Justification for this assumption is given in [18], where it is shown that lattice codes can achieve the Shannon capacity of an additive Gaussian noise channel.) Assuming this is true, then there is a sequence of dimensions $K_1, K_2, \cdots$, which increases without bound, for which lattice codes are nearly optimal. The coding gain of an optimal signal set for each of these input lengths can then be separated into the gain obtained from the dense lattice plus the shaping gain, both of which are independent of R when R is large.

As $R\to\infty$, the upper (lower) bounds on coding gain for each of the channels shown in Fig. 6 converge to the same limit. This is because the upper bound on d-rate, given by (4.3), satisfies

$$\bar{C}(\bar{d}^*) \ge \frac{1}{2}\int_0^1 \log_2\left(\frac{4P|H(f)|^2}{\bar{d}^{*2}}\right)df$$
$$= \log_2\frac{2\sqrt{P}}{\bar{d}^*} + \int_0^1 \log_2|H(f)|\,df, \qquad (4.7)$$

where the inequality becomes asymptotically tight as

$\bar{d}^* \to 0$. For the PR channels considered, and the identity channel, $\int_0^1 \log_2 |H(f)| \, df = \log_2 h[0] = 0$. Setting $\bar{C}(\bar{d}^*) = R$, solving for $\bar{d}^*/\sqrt{P}$ via (4.7), and substituting into the expression for G_0^* gives

$$\lim_{R\to\infty} G_0^* \le \lim_{R\to\infty} 10\log_{10} \frac{4(4^{-R})(4^R - 1)}{12} = 10\log_{10}\frac{1}{3} = -4.77 \text{ dB}, \tag{4.8}$$

for the channels considered. Similarly, it is easily verified that the lower bound on d-rate, given by (4.2), satisfies

$$\underline{C}(\bar{d}^*) = \log_2 \frac{\sqrt{P}}{\bar{d}^*} + \int_0^1 \log_2 |H(f)| \, df + \delta(d^*), \tag{4.9}$$

where $\delta(d^*)$ is a correction term that satisfies $\lim_{d^* \to 0} \delta(d^*) = 0$, so that

$$\lim_{R\to\infty} G_0^* \ge 10\log_{10}\frac{1}{12} = -10.8 \text{ dB}. \tag{4.10}$$

The 6 dB difference between the bounds on G_0^* as $R \to \infty$ is a direct consequence of the 6 dB difference along the x axis in Fig. 5 between the asymptotes of $\underline{C}(\bar{d}^*)$ and $\bar{C}(\bar{d}^*)$.

Observe that the difference between the upper (or lower) bound on G_0^* for a PR channel and the upper (lower) bound for the identity channel shown in Fig. 6 resembles the asymptotic coding gain of the BSC for the PR channel shown in Fig. 1. This is consistent with the decomposition of coding gain discussed in Section IIIA. Specifically, as $K \to \infty$, (4.5) implies that the difference between coding gains of optimal signal sets for two channels at a given R is the difference in $G_0^*(R)$ for each channel. Consequently, (3.3) implies that $G_0^*(R)$ for a PR channel minus $G_0^*(R)$ for the identity channel is approximately equal to the asymptotic BSC coding gain for the PR channel. Fig. 6 shows that the bounds on G_0^* given here also satisfy this relation.

V. Conclusion

This paper has studied some properties of l_2/l_2 signal design for large input lengths K. Our main result is that for large K, the coding gain of an optimal signal set for a PR channel can be decomposed into the coding gain of an optimal code for the identity channel of length K plus the coding gain of a baseline signal constellation, defined in Section III. The latter gain was explicitly evaluated as $K \to \infty$ and depends only on the information rate and the channel transfer function. For the channels considered, the coding gain of the BSC, relative to single-step detection, is positive, and approaches zero monotonically as $R \to \infty$. For large R and K, optimal signal sets for the PR channels considered therefore give approximately the same coding gain as optimal signal sets for the identity channel.

For fixed R, the coding gain of optimal signal sets as $K \to \infty$ can be evaluated by computing the asymptotic normalized distance $\bar{d}^*$ for which the d-rate of the channel, $C(\bar{d}^*) = R$. Unfortunately, $C(\bar{d}^*)$ cannot be computed exactly for channels of interest, although upper and lower bounds are available. Plots of normalized squared distance, d^2/K, vs. K were shown in Section IV, and were generated by combining the results in [1, Section IV] with results in the literature [3] on coding gains offered by dense lattices relative to the integer lattice. For the cases considered, the results are consistent with the upper and lower bounds on asymptotic minimum distance computed from bounds on the d-rate, and indicate that these asymptotic bounds are useful for $K \ge 10$.

The plots of d^2/K vs. K also illustrate the decomposition of coding gain. Specifically, the maximum coding gain of a length K code for the $1-D$ channel at $R = 2$ bits$/T$ for $K \ge 5$ was estimated by adding the coding gain obtained from using a dense lattice with optimal shaping, assuming the identity channel, to the coding gain of the BSC. The resulting normalized distance for large K was found to lie between the upper and lower bounds on asymptotic normalized distance computed from the d-rate of the channel. The decomposition of coding gain is again illustrated in Fig. 6, which shows upper and lower bounds on asymptotic coding gain vs. rate for some different PR channels.

Comparative plots of d-rates for some PR channels show that at low rates optimal signal sets with longer lengths are needed for the identity channel, as compared with the PR channels considered, to obtain the same minimum distance between outputs. This is due to the greater impulse response energy for the PR channels considered relative to the identity channel. If, however, the PR channels are normalized so that $\int_0^1 |H(f)|^2 \, df = 1$, then for fixed $\bar{d}^*$ the bounds on d-rates for the PR channels considered are *less than* the corresponding bounds for the identity channel. That is, subject to an L_2 constraint on $H(f)$, the identity channel is the best among those considered in the sense that this channel maximizes the asymptotic minimum distance corresponding to optimal signal sets.

As in Part I of this paper [1], we make the final remark that a similar approach to that used here to analyze asymptotic coding gain might be applicable to other signal design problems, i.e., the l_q/l_p problem defined in [1]. Specifically, the d-rate is easily generalized so that it corresponds to this larger class of problems, and, if it can be computed, could similarly be used to study asymptotic minimum distance vs. rate. Of course, even upper and lower bounds on this generalized d-rate, or "maximum channel throughput," may be quite difficult to obtain for specific p and q.

Acknowledgment

The author thanks K. Steiglitz and S. Verdu for helpful discussions concerning this work, and the reviewers for their comments which helped improve the original manuscript. The author is especially grateful to one of the anonymous reviewers for pointing out a numerical error in one of the original plots.

References

[1] M. L. Honig, K. Steiglitz, and S. Norman, "Optimization of signal sets for partial response channels—Part I: Numerical techniques," *IEEE Trans. Inform. Theory*, vol. 37, no. 5, pp. 1327–1341, Sept. 1991.

[2] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices, and Groups*. New York: Springer-Verlag, 1988.

[3] G. D. Forney *et al.*, "Efficient modulation for band-limited channels," *IEEE J. Select. Areas Commun.*, vol. SAC-2, no. 5, pp. 632–647, Sept. 1984.

[4] G. D. Forney, Jr., "Coset Codes—Part I: Introduction and geometrical classification," *IEEE Trans. Inform. Theory*, vol. 34, no. 5, pp. 1123–1151, Sept. 1988.

[5] G. D. Forney and L.-F. Wei, "Multidimensional constellations—Part I: Introduction, figures of merit, and generalized cross constellations," *IEEE J. Select. Areas Commun.*, vol. 7, no. 6, pp. 877–892, Aug. 1989.

[6] G. D. Forney, Jr., "Trellis shaping," submitted to *IEEE Trans. Inform. Theory*.

[7] W. L. Root, "Estimates of ϵ-capacity for certain linear communication channels," *IEEE Trans. Inform. Theory*, vol. IT-14, no. 3, pp. 361–369, May 1968.

[8] R. T. Prosser and W. L. Root, "The ϵ-entropy and ϵ-capacity of certain time-invariant channels," *J. Math. Anal. Applica.*, vol. 21, pp. 233–241, 1968.

[9] L. J. Forys and P. P. Varaiya, "The ϵ-capacity of classes of unknown channels," *Inform. Contr.*, vol. 14, pp. 376–406, 1969.

[10] M. L. Honig and P. Narayan, "A new upper bound on ϵ-rate," presented at the *IEEE Symp. Inform. Theory*, San Diego, CA, manuscript in preparation.

[11] S. Kasturia, J. T. Aslanis, and J. M. Cioffi, "Vector coding for partial-response channels," *IEEE Trans. Inform. Theory*, vol. 36, no. 4, pp. 741–762, July 1990.

[12] K. Sistanizadeh, "Block coding capacity of digital subscriber lines by the structured channel signaling technique," in *Proc. 1989 IEEE Globecom Conf.*, Dallas, TX, Nov. 1989.

[13] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 6, pp. 725–730, Nov. 1972.

[14] A. Papoulis, "Maximum entropy spectral estimation: A review," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-29, no. 6, pp. 1176–1186, Dec. 1981.

[15] M. V. Eyuboglu and G. D. Forney, Jr., "Trellis precoding: Combined coding, precoding and shaping for intersymbol interference channels," submitted to *IEEE Trans. Inform. Theory*.

[16] N. A. Zervos and I. Kalet, "Optimized decision feedback equalization versus optimized orthogonal frequency division multiplexing for high-speed data transmission over the local cable network," *Proc. ICC '89*, Paper 35.2, Boston, MA, June 1989.

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

[18] R. de Buda, "Some optimal codes have structure," *IEEE J. Select. Areas Commun.*, vol. 7, no. 6, pp. 893–899, Aug. 1989.