

On Design Criteria and Construction of Non-coherent Space-Time Constellations

Mohammad Jaber Borran, *Student Member, IEEE*, Ashutosh Sabharwal, *Member, IEEE*, and Behnaam Aazhang, *Fellow, IEEE*

Abstract— We consider the problem of digital communication in a Rayleigh flat fading environment using a multiple-antenna system, when the channel state information is available neither at the transmitter nor at the receiver. It is known that at high SNR, or when the coherence interval is much larger than the number of transmit antennas, a constellation of unitary matrices can achieve the capacity of the non-coherent system. However, at low SNR, high spectral efficiencies, or for small values of coherence interval, the unitary constellations lose their optimality and fail to provide an acceptable performance.

In this work, inspired by the Stein’s lemma, we propose to use the Kullback-Leibler distance between conditional distributions to design space-time constellations for non-coherent communication. In fast fading, i.e., when the coherence interval is equal to one symbol period and the unitary construction provides only one signal point, the new design criterion results in PAM-type constellations with unequal spacing between constellation points. We also show that in this case, the new design criterion is equivalent to design criteria based on the exact pairwise error probability and the Chernoff information.

When the coherence interval is larger than the number of transmit antennas, the resulting constellations overlap with the unitary constellations at high SNR, but at low SNR they have a multilevel structure and show significant performance improvement over unitary constellations of the same size. The performance improvement becomes especially more significant when an appropriately designed outer code or multiple receive antennas are used. This property, together with the facts that the proposed constellations eliminate the need for training sequences and are most suitable for low SNR, makes them a good candidate for uplink communication in wireless systems.

Index Terms— Fading channels, multiple antenna systems, non-coherent constellations, space-time modulation, wireless communications.

I. INTRODUCTION

EXPLOITING propagation diversity by using multiple antennas at the transmitter and the receiver in wireless communication systems has been recently proposed and studied using different approaches [1]–[8]. In [1], [2], it has been shown that in a Rayleigh flat fading environment, the capacity of a multiple antenna system increases linearly with the smaller of the number of transmit and receive antennas, provided that the fading coefficients are known at the receiver. In a slowly fading channel, where the fading coefficients remain approximately constant for many symbol intervals, the

transmitter can send training signals that allow the receiver to accurately estimate the fading coefficients; in this case, the results of [1], [2] are applicable.

In fast fading scenarios, however, fading coefficients can change into new, almost independent values before being learned by the receiver through training signals. This problem becomes even more acute when large numbers of transmit and receive antennas are being used by the system, which requires very long training sequences to estimate the fading coefficients. Even if the channel does not change very rapidly, for applications which require transmission of short control packets (such as RTS and CTS in IEEE 802.11), long training sequences have a large overhead (in terms of the amount of time and power spent on them), and significantly reduce the efficiency of the system. A non-coherent detection scheme, where receiver detects the transmitted symbols without having any information about the current realization of the channel, is more suitable for these fast fading scenarios.

The capacity of the non-coherent systems has been studied in [5], [6], where it has been shown that at high SNR, or when the coherence interval, T , is much greater than the number of transmit antennas, M , capacity can be achieved by using a constellation of unitary matrices (i.e. matrices with orthonormal columns). Optimal unitary constellations are the optimal packings in complex Grassmannian manifolds [9]. These packings are usually obtained through exhaustive or random search, and their decoding complexity is exponential in the rate of the constellation and the block length (usually assumed to be equal to the coherence interval of the channel), or linear in the number of the points in the constellation.

In [7], a systematic method for designing unitary space-time constellations has been proposed, however, the resulting constellations still have exponential decoding complexity. A group of low decoding complexity real unitary constellations has been proposed in [8]. These real constellations are optimal when the coherence interval is equal to two (symbol periods) and number of transmit antennas is equal to one. However, the proposed extension to large coherence intervals or multiple transmit antennas does not maintain their optimality.

In this paper, we consider constellations of orthogonal (rather than orthonormal) matrices, and propose a new design criterion for non-coherent space-time constellations based on the Kullback-Leibler (KL) distance [10] between conditional distributions. By imposing a multi-level structure on the constellation, we decouple the design problem into simpler optimizations and use the existing unitary designs to construct the optimal multi-level constellation. Our motivations for

M. J. Borran was with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005 USA. He is now with the Nokia Mobile Phones, Irving, TX 75039 USA (e-mail: mohammad@rice.edu).

A. Sabharwal and B. Aazhang are with the Department of Electrical and Computer Engineering (MS-366), Rice University, Houston, TX 77005 USA (e-mail: ashu@rice.edu; aaz@rice.edu).

considering non-unitary constellations and proposing the new design criterion are the following:

- The *long coherence time* requirement for the optimality of the unitary constellations makes them less desirable for high-mobility scenarios. For example, all of the unitary constellations of [6]–[8] have been designed for the case of $T > 1$. For $T = 1$, they provide only one signal point, which is obviously incapable of transmitting any information. The capacity of discrete-time fast Rayleigh fading channels ($T = 1$) has been studied in [11], and has been shown to be greater than zero. It has also been shown that the capacity achieving distribution for $T = 1$ is discrete, with a finite number of points, and one of them always located at the origin. In general, long coherence interval means a slowly fading channel (in which case a training based coherent transmission might be more desirable), whereas the non-coherent constellations are usually needed when channel changes rapidly and training is difficult or expensive (in terms of the amount of time and power spent on that).
- The *high SNR* requirement for the optimality of the unitary constellations implies low power efficiency. This is because the capacity is a logarithmic function of the average power, and thus, a linear increase in the power results in only a logarithmic increase in the capacity. Considering the power limitations in the battery operated devices, the high SNR requirement cannot be easily satisfied by mobile devices.
- It appears that the unitary designs are not completely using the information about the statistics of the fading [12]. It can be easily shown that, e.g., in the case of real constellations for $M = 1$ and $T = 2$, a non-Bayesian approach (i.e., assuming that fading is unknown, with no information about its distribution), would result in a unitary design. If the statistics of the fading process are known at the transmitter and the receiver, (e.g., if the channel is assumed to be block Rayleigh flat fading, with fading coefficients from the distribution $\mathcal{CN}(0, 1)$), then a design criterion which takes this knowledge into account more efficiently, would result in better constellations (in terms of error rate performance).
- A common performance measure for evaluating different constellations in communication systems is the average symbol error probability. However, the expressions for the average error probability are usually very complicated, and do not provide much insight into the design problem. Therefore, we initially consider the pairwise error probability as our performance and design criterion. Unfortunately, the exact expression and even the Chernoff upper bound for the pairwise error probability do not seem to be tractable for a general non-coherent constellation. Moreover, except for the special case of unitary constellations, the pairwise error probabilities are not symmetric. The KL distance is an asymmetric distance, is relatively easy to derive even for general multiple-antenna constellations, and is equal to the best achievable error exponent using hypothesis testing (Stein’s lemma [10]).

For the above reasons, we propose the use of KL distance between conditional distributions as the design criterion, and construct constellations for single- and multiple-antenna systems in fast and block fading channels which outperform the existing unitary constellations.

The main contributions of this paper are summarized in the following.

- 1) *The relation between the KL distance and the ML detector performance*: Using the fact that the KL distance is the expected value of the likelihood ratio, we show that, for a large number of observations (e.g., receive antennas in our case), the performance of the Maximum Likelihood (ML) detector is related to the KL distance between the conditional distributions (Lemma 3). We also show that in some cases ($T = 1$), for any number of receive antennas, the KL-based design criterion is equivalent to the design criteria based on the exact pairwise error probability and the Chernoff bound (Section III).
- 2) *KL-based design criterion*: We derive the KL distance between the conditional received distributions corresponding to different transmit symbols for the general case of multiple-antenna systems (Equation (31)), and propose a design criterion based on maximizing the minimum of this KL distance over the pairs of constellation points. We also derive the simplified distance for orthogonal constellations (Equation (34)).
- 3) *Multi-level structure and decoupled optimization*: By imposing a multi-level unitary structure on the constellation, we further simplify the KL distance, and decouple the design problem into simpler optimizations. As a result, we can use any existing unitary design at the levels of the multi-level constellation, and find the optimum distribution of the points among the levels, and their radiuses. In fast fading, i.e., when $T = 1$ and the unitary construction provides only one signal point, the new design criterion results in PAM-type constellations with unequal spacing between constellation points (Section V-A). When the coherence interval is larger than the number of transmit antennas, the resulting constellations overlap with the unitary constellations at high SNR, but at low SNR they have a multilevel structure and show significant performance improvement over unitary constellations of the same size (Sections V-C and V-D).

The rest of this paper is organized as follows. In Section II, we introduce the model for the system being considered throughout this paper. In Section III, we derive the exact expression and the Chernoff upper bound for the pairwise error probability in the fast fading scenario, and show that they result in the same design criterion as the one suggested by the KL distance. In Section IV, we derive the KL distance between conditional distributions of the received signal, and propose the design criterion based on that. In Section V, we present non-coherent constructions for several important cases and compare their performance with known unitary space-time constellations. We show that the new constellations can provide significant performance improvement compared to

the unitary constellations, especially at low SNR and when multiple receive antennas are used. Finally, in Section VI we bring some concluding remarks.

II. SYSTEM MODEL

We consider a communication system with M transmit and N receive antennas in a block Rayleigh flat fading channel with coherence interval of T symbol periods (i.e., we assume that the fading coefficients remain constant during blocks of T consecutive symbol intervals, and change into new, independent values at the end of each block). We use the following complex baseband notation

$$X = SH + W, \quad (1)$$

where S is the $T \times M$ matrix of transmitted signals, X is the $T \times N$ matrix of received signals, H is the $M \times N$ matrix of fading coefficients, and W is the $T \times N$ matrix of the additive received noise. Elements of H and W are assumed to be statistically independent, identically distributed circular complex Gaussian random variables from the distribution $\mathcal{CN}(0, 1)$. We intentionally avoid using the scaling factor of $\sqrt{\frac{P}{M}}$ of [7] to account for the desired signal to noise ratio (or average power constraint on the constellation). We will see in Section V, that the structure of the optimal constellation depends on the actual value of the signal to noise ratio, and constellations of the same size at different SNR's are not necessarily scaled versions of each other. Therefore we capture the SNR factor in the S matrix itself, and use the power constraint $\frac{1}{T} \sum_{t=1}^T \sum_{m=1}^M \mathbb{E} \{ |s_{tm}|^2 \} \leq P$, where s_{tm} 's are the elements of the signal matrix S .

With the above assumptions, each column of the received matrix, X , is a zero-mean circular complex Gaussian random vector with covariance matrix $I_T + SS^H$. Therefore, the conditional probability density function of X_n , the n th column of the received matrix, X , can be written as

$$\begin{aligned} p(X_n|S) &= \mathbb{E}_H \{ p(X_n|S, H) \} \\ &= \frac{\exp \left\{ -X_n^H (I_T + SS^H)^{-1} X_n \right\}}{\pi^T \det(I_T + SS^H)}. \end{aligned} \quad (2)$$

Since the columns of X are statistically independent, we have the following expression for the conditional probability density function of the whole received matrix, X ,

$$\begin{aligned} p^N(X|S) &= \prod_{n=1}^N p(X_n|S) \\ &= \frac{\exp \left\{ -\text{tr} \left[(I_T + SS^H)^{-1} X X^H \right] \right\}}{\pi^{TN} \det^N(I_T + SS^H)}. \end{aligned} \quad (3)$$

Note that the superscript N is not an exponent for p . It only specifies the column size of X and emphasizes the statistical independence of its columns. Later, when calculating the pairwise error probabilities and error exponents, we will find this notation convenient.

Assuming a signal set of size L , $\{S_l\}_{l=1}^L$, and defining $p_l^N(X) = p^N(X|S_l)$, the Maximum Likelihood (ML) detector

for this system will have the following form

$$\hat{S}_{ML} = S_{\hat{l}_{ML}}, \quad \text{where} \quad \hat{l}_{ML} = \arg \max_{l \in \{1, \dots, L\}} p_l^N(X). \quad (4)$$

If $L = 2$, then the probability of error in ML detection of S_1 (detecting S_2 given that S_1 was transmitted) is given by

$$\Pr(S_1 \rightarrow S_2) = \Pr_{p_1^N} \{ X : p_2^N(X) > p_1^N(X) \}, \quad (5)$$

where we have used the notation $\Pr_p \{ \mathcal{R} \}$ to denote the probability of set \mathcal{R} with respect to the probability density function p . Now, if we also assume that S_1 and S_2 are transmitted with equal probabilities, then the average probability of error in ML detection will be given by

$$P_e(S_1, S_2) = \frac{1}{2} \Pr(S_1 \rightarrow S_2) + \frac{1}{2} \Pr(S_2 \rightarrow S_1). \quad (6)$$

It is known [10] that this average pairwise error probability decays exponentially with the number of independent observations, and the rate of this exponential decay is given by the Chernoff information [10] between the conditional distributions.

For $L > 2$, even though (5) and (6) are no longer exact, we will still use them as an approximation for the pairwise error probability, which will, in turn, be used to derive the design criterion for space-time constellations.

For the special case of unitary transmit matrices, i.e., when $S_l^H S_l = (\frac{TP}{M}) I_M$, the exact expression and Chernoff upper bound for the pairwise error probability were calculated in [6]. However, the corresponding expressions for the general case of arbitrary matrix constellations do not seem to be easily tractable. In the next section, we will calculate these expressions for the case of a single transmit antenna in fast fading, and will show that they result in the same design criterion.

Before proceeding to the next section and deriving the expressions for the pairwise error probabilities, we notice that the conditional probability density function in (3) depends on S only through the product SS^H . Since the pairwise error probability in (5) (and also in general the average error probability of the ML detector) is determined only by the conditional probability density functions, it is clear that the performance of the non-coherent multiple-antenna constellation also depends on the constellation matrices $\{S_l\}_{l=1}^L$ only through the products $\{S_l S_l^H\}_{l=1}^L$. If $M > T$, using the Cholesky factorization $SS^H = QQ^H$ where Q is a $T \times T$ lower triangular matrix, for any constellation of $T \times M$ matrices one can find a corresponding constellation of $T \times T$ matrices with the same pairwise and average error probabilities. Also if $T \geq M$, using the singular value decomposition, we can write $S = \Phi V \Psi^H$, where Φ and Ψ are $T \times M$ and $M \times M$ unitary (orthonormal) matrices and V is an $M \times M$ real non-negative diagonal matrix. Defining $S' = \Phi V$, we have $S' S'^H = SS^H$. This means that for any constellation of $T \times M$ arbitrary matrices, one can find a constellation of $T \times M$ orthogonal matrices with the same pairwise and average error probabilities. Since the average power of the constellation also depends only on

the expected value of the product SS^H ,

$$P = \frac{1}{T} \sum_{t=1}^T \sum_{m=1}^M \mathbb{E} \{ |s_{tm}|^2 \} = \frac{1}{T} \mathbb{E} \{ \text{tr} [SS^H] \}, \quad (7)$$

the new constellation will also have the same average power as the original constellation. Therefore, (similar to the result of [5] for capacity) we have the following result.

Theorem 1: Increasing the number of transmit antennas beyond T does not improve the error probability performance of the non-coherent multiple-antenna systems. Furthermore, the optimum error probability performance can be achieved by a constellation of orthogonal matrices.

Therefore, in the rest of this paper, we will assume that $M \leq T$, and that the constellation matrices are orthogonal.

III. PAIRWISE ERROR PROBABILITY FOR FAST FADING ($T = 1$)

Throughout this section, we will assume that there are only two signal points in the constellation, i.e., $L = 2$. As mentioned in the previous section, there is no gain in using more transmit antennas than T . Therefore, we also assume that $M = 1$. With this assumption, the transmit matrix is simply a complex scalar. We denote this scalar by s . The conditional probability density of the received signal given the transmitted symbol is given by

$$p^N(X|s) = \frac{1}{\pi^N(1+|s|^2)^N} \exp\left(\frac{-\|X\|^2}{1+|s|^2}\right). \quad (8)$$

In the following, we derive the exact expression for $\Pr(s_1 \rightarrow s_2)$ in this case.

Lemma 1: For a single antenna communication system (i.e., $M = N = 1$) in a fast fading environment (i.e., $T = 1$), the pairwise error probability of non-coherent ML detector is given by

$$\Pr(s_1 \rightarrow s_2) = \begin{cases} \left(\frac{1+|s_1|^2}{1+|s_2|^2}\right)^{\frac{1+|s_2|^2}{|s_2|^2-|s_1|^2}}, & \text{for } |s_1| < |s_2|; \\ 1 - \left(\frac{1+|s_1|^2}{1+|s_2|^2}\right)^{\frac{1+|s_2|^2}{|s_2|^2-|s_1|^2}}, & \text{for } |s_1| > |s_2|. \end{cases} \quad (9)$$

Proof: Using (5) and (8) (with $N = 1$), and assuming that $|s_1| < |s_2|$, we have

$$\begin{aligned} \Pr(s_1 \rightarrow s_2) &= \Pr_{p_1} \{x : p_2(x) > p_1(x)\} \\ &= \Pr_{p_1} \left\{ x : \frac{1}{\pi(1+|s_2|^2)} \exp\left(\frac{-|x|^2}{1+|s_2|^2}\right) \right. \\ &\quad \left. > \frac{1}{\pi(1+|s_1|^2)} \exp\left(\frac{-|x|^2}{1+|s_1|^2}\right) \right\} \\ &= \Pr_{p_1} \{x : |x|^2 > A\} \\ &= \int_{\sqrt{A}}^{\infty} \int_0^{2\pi} p_1(\rho e^{i\theta}) \rho d\theta d\rho \\ &= \frac{1}{1+|s_1|^2} \int_A^{\infty} \exp\left(\frac{-\rho^2}{1+|s_1|^2}\right) d(\rho^2) \\ &= \exp\left(\frac{-A}{1+|s_1|^2}\right) \\ &= B^{\frac{1}{1-B}}, \end{aligned}$$

where

$$\begin{aligned} A &= \frac{1}{\frac{1}{1+|s_1|^2} - \frac{1}{1+|s_2|^2}} \ln\left(\frac{1+|s_2|^2}{1+|s_1|^2}\right) \\ &= \frac{(1+|s_1|^2)(1+|s_2|^2)}{|s_2|^2 - |s_1|^2} \ln\left(\frac{1+|s_2|^2}{1+|s_1|^2}\right), \end{aligned} \quad (10)$$

and

$$B = \frac{1+|s_1|^2}{1+|s_2|^2}. \quad (11)$$

Similarly, it can be shown that if $|s_1| > |s_2|$, we have

$$\Pr(s_1 \rightarrow s_2) = 1 - B^{\frac{1}{1-B}}.$$

This completes the proof. \blacksquare

Theorem 2: For a single transmit antenna communication system (i.e., $M = 1$) in a fast fading environment (i.e., $T = 1$), the pairwise error probability of non-coherent ML detector is given by

$$\Pr(s_1 \rightarrow s_2) = \begin{cases} \left[\sum_{n=0}^{N-1} \frac{1}{n!} \left(\frac{-N \ln(B)}{1-B}\right)^n \right] \exp\left(\frac{N \ln(B)}{1-B}\right), & \text{for } B < 1; \\ 1 - \left[\sum_{n=0}^{N-1} \frac{1}{n!} \left(\frac{-N \ln(B)}{1-B}\right)^n \right] \exp\left(\frac{N \ln(B)}{1-B}\right), & \text{for } B > 1; \end{cases} \quad (12)$$

where B is as in (11).

Proof: Proof is by induction, and is given in Appendix A. \blacksquare

It is interesting to notice that, in this case, the two pairwise error probabilities, i.e., $\Pr(s_1 \rightarrow s_2)$ and $\Pr(s_2 \rightarrow s_1)$ are not equal, even in the limit as $|s_1| \rightarrow |s_2|^-$ or $|s_1| \rightarrow |s_2|^+$. The following two identities can be verified easily.

$$\lim_{|s_1| \rightarrow |s_2|^-} \Pr(s_1 \rightarrow s_2) = \left[\sum_{n=0}^{N-1} \frac{N^n}{n!} \right] \exp(-N), \quad (13)$$

$$\lim_{|s_1| \rightarrow |s_2|^+} \Pr(s_1 \rightarrow s_2) = 1 - \left[\sum_{n=0}^{N-1} \frac{N^n}{n!} \right] \exp(-N). \quad (14)$$

For $N = 1$, the above equalities reduce to the following:

$$\lim_{|s_1| \rightarrow |s_2|^-} \Pr(s_1 \rightarrow s_2) = \frac{1}{e} \approx 0.3679,$$

$$\lim_{|s_1| \rightarrow |s_2|^+} \Pr(s_1 \rightarrow s_2) = 1 - \frac{1}{e} \approx 0.6321.$$

This discontinuity can be explained as follows. With the ML detector, the decision region corresponding to each constellation point is the region in which that constellation point has the largest likelihood among all of the constellation points. The likelihood functions in this case (binary signaling, single transmit antenna, fast fading, and non-coherent detection) are complex Gaussian probability density functions with zero mean and variance $1+|s_l|^2$, for $l = 1, 2$. The boundary of the two decision regions is the line along which the two likelihood functions are equal. For $N = 1$, this boundary is a circle in the complex plane centered at the origin, with radius \sqrt{A} , where A is as in (10). Fig. 1 shows a cross-section of the likelihood functions and the boundary of the decision regions for $s_1 = 0$

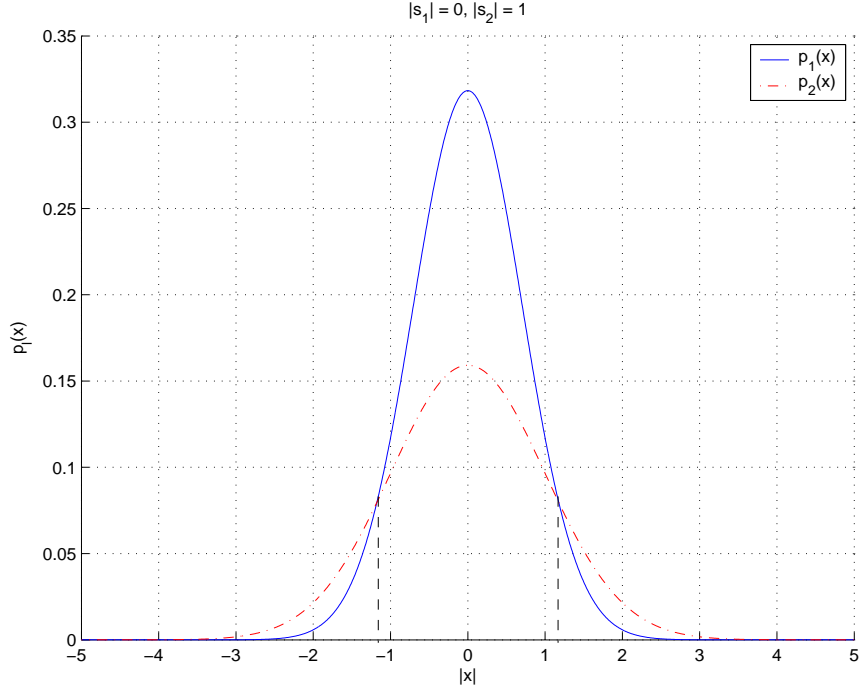


Fig. 1. The likelihood functions for $|s_1| = 0$, $|s_2| = 1$, and $N = 1$

and $s_2 = 1$. Fig. 2 shows the radius of the boundary vs. $|s_1|$, when $s_2 = 1$. As we can see in these figures, the boundary is not defined when $|s_1| = |s_2|$, because in this case the two likelihood functions are equal everywhere. However, both right and left limits of the boundary radius (as $|s_1| \rightarrow |s_2|$) exist and are, in fact, equal. It can be easily verified that the limit is equal to the standard deviation of the common distribution, or

$$\lim_{|s_1| \rightarrow |s_2|} A = 1 + |s_2|^2. \quad (15)$$

The probability of mistaking s_1 for s_2 is the volume under the conditional pdf corresponding to s_1 in the decision region of s_2 . The discontinuity of the pairwise error probability comes from the fact that, even though the radius of the boundary of the decision regions does not have a jump discontinuity at $|s_1| = |s_2|$, the decision region of s_2 itself suddenly changes from outside the boundary circle to inside the circle, as $|s_1|$ changes from a value smaller than $|s_2|$ to a value larger than $|s_2|$. Since the volumes under the Gaussian density function in the two regions of inside and outside a circle with radius equal to the standard deviation of the distribution are different, the left and right limits of the pairwise error probability are also different, resulting in a jump discontinuity. This discontinuity can be seen in Fig. 3 which shows the two pairwise error probabilities as a function of B for $N = 1$ ($B < 1$ and $B > 1$ correspond to $|s_1| < |s_2|$ and $|s_1| > |s_2|$, respectively).

Another interesting point to observe in Figs. 1 and 2 is that the signal points may not belong to their respective decision regions. As we see in Fig. 1, the radius of the boundary of the decision region is greater than one (around 1.1774), whereas both constellation points have magnitudes smaller than or equal to one. This is not the case for large values of $|s_1|$

as shown in Fig. 2, e.g., for $|s_1| = 2$, where the radius of the boundary is less than 2. Fig. 2 also compares the radius of the boundary of the decision region with the arithmetic mean of the magnitudes of the constellation points (probably the most intuitive, yet incorrect, value for the radius of the boundary). As we see, for large values of $|s_1|$, the radius of the boundary is much smaller than the arithmetic mean.

From Fig. 3, we observe that in the two disjoint regions $\{|s_1| < |s_2|\}$ and $\{|s_1| > |s_2|\}$, the pairwise error probability is a monotonic function of B as defined in (11). Therefore, assuming $|s_1| < |s_2|$, minimizing the pairwise error probability is equivalent to minimizing B .

The next proposition gives the Chernoff bound on the exponent for pairwise error probability in this case.

Proposition 1: Consider a single transmit antenna communication system (i.e., $M = 1$), in a fast fading environment (i.e., $T = 1$). The largest achievable exponent for the average probability of error (i.e., the Chernoff information [10]) for this system is given by the following expression

$$\mathcal{C}(p_1^N, p_2^N) = \begin{cases} N \left[\frac{\ln(B)}{B-1} - \ln \left(\frac{\ln(B)}{B-1} \right) - 1 \right], & B \neq 1; \\ 0, & B = 1; \end{cases} \quad (16)$$

where B is as in (11).

Proof: See Appendix B. ■

Notice that since $B > 0$, and $\ln(B)$ and $B - 1$ have the same sign, the Chernoff distance in (16) is well-defined, and since $x - \ln(x) - 1 \geq 0$ for $x > 0$, it is always greater than or equal to zero.

Fig. 4 shows the exact average error probability, and the Chernoff bound for the average error probability given by $\frac{1}{2} \exp \{-NC(p_1^N, p_2^N)\}$, for $N = 1$. As we see, the Chernoff bound is also a monotonic function of B in the two regions

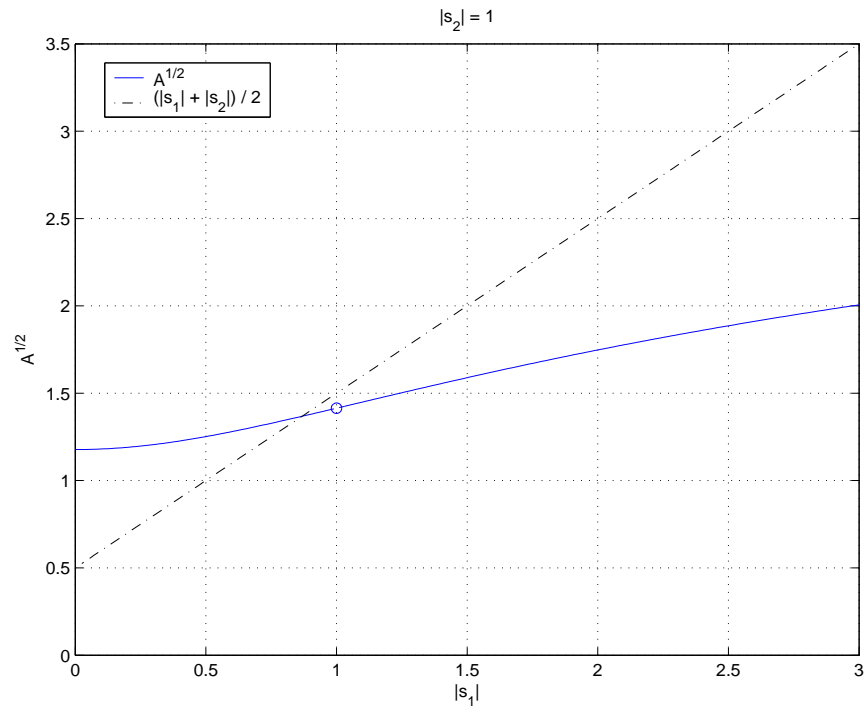


Fig. 2. The radius of the decision region vs. $|s_1|$ for $|s_2| = 1$ and $N = 1$.

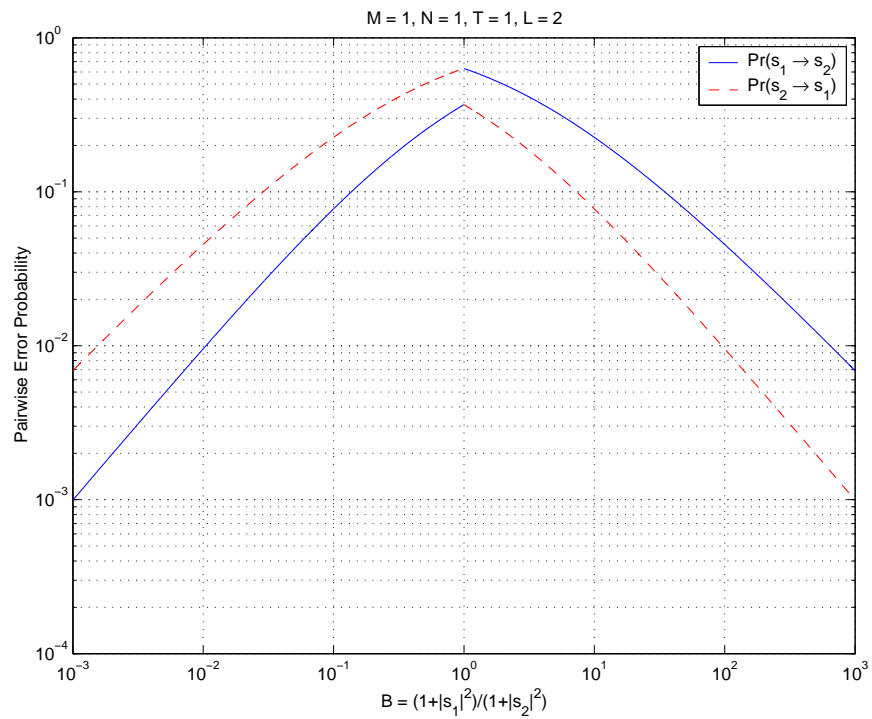


Fig. 3. Pairwise error probabilities for single antenna in fast fading.

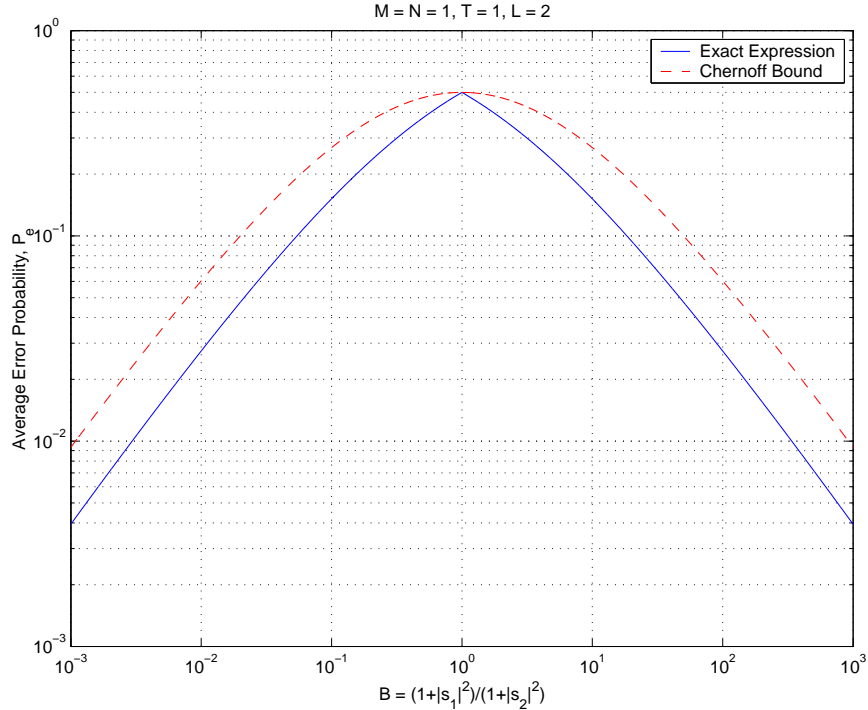


Fig. 4. Average error probability and the Chernoff bound for single antenna in fast fading.

of $\{B < 1\}$ and $\{B > 1\}$. We will see later in Section V-A, that the Kullback-Leibler (KL) distance between the two conditional distributions corresponding to the two different transmitted symbols is given by

$$\mathcal{D}_1(p_1^N \| p_2^N) = N [B - \ln(B) - 1]. \quad (17)$$

This expression is also a monotonic function of B in the two regions of $\{B < 1\}$ and $\{B > 1\}$. Therefore, the three different criteria of (a) minimizing the maximum of the exact pairwise error probability, (b) maximizing the minimum of the Chernoff distance, and (c) maximizing the minimum of the KL distance, are all equivalent to minimizing the maximum of B (assuming $B < 1$), and will result in the same constellation. In Section V-A, we design new constellations based on this design criterion and compare their performance with the performance of the conventional PAM constellations.

IV. DESIGN CRITERION

It is known [5] that the capacity achieving signal matrix for the non-coherent systems can be written as $S = \Phi V$, where Φ is a $T \times M$ isotropically distributed unitary matrix, and V is an independent $M \times M$ real, nonnegative, diagonal matrix. It is also known [6] that when the SNR is high or when $T \gg M$, the capacity can be achieved by a unitary signal matrix (i.e., a signal matrix with orthonormal columns obtained by setting V to a deterministic multiple of the identity matrix).

Based on these results, unitary space-time modulation has been studied in [6], where expressions for the exact pairwise error probability as well as the Chernoff upper bound for the pairwise error probability of unitary matrices have been derived. These expressions suggest a design criterion based on minimizing the singular values of the product matrices, $\Phi_i^H \Phi_j$,

over all pairs of the constellation points. In [8], it has been shown that this design criterion is approximately equivalent to maximizing the minimum so-called *square Euclidean distance* between subspaces spanned by columns of the constellation matrices. It can be shown (see Appendix F) that the square Euclidean distance between subspaces is equivalent to the *chordal distance*, as defined in [13]. Therefore, unitary designs can be considered as packings in the complex Grassmannian manifolds [13].

The problem with the unitary constellations is that they are optimal only at high SNR or when $T \gg M$. These requirements are rather restrictive, and cannot be met in many situations of practical importance. Operation at high SNR means low power efficiency, which is in contradiction with the low power requirements of the wireless systems. On the other hand, large coherence interval means a slowly fading channel, in which case a training based coherent signaling might be more desirable. In fact, the main motivation for non-coherent communication is to deal with fast fading scenarios where training is either impossible or very expensive (in terms of the fraction of time and energy spent on that). Even if the coherence interval is large, because of the exponential growth of the constellation size and (in most cases) decoding complexity with T , one might decide to design a constellation for a block length with is much smaller than T . At low SNR, or when the block length is not much larger than M , the unitary designs lose their optimality and fail to provide a desirable performance. For these reasons, in this work we do not assume a unitary structure on the constellation matrices, and try to design non-coherent signal sets of matrices with orthogonal (rather than orthonormal) columns.

Unlike the case of unitary constellations, the pairwise error

probability of the non-coherent ML detector, which is approximately given by (5), does not appear to be tractable in the more general case of orthogonal matrices. Even the Chernoff distance (see Appendix B for the definition and an example), which determines the exponential decay rate of the average pairwise error probability of the ML detector [10], does not seem to admit a simple closed form expression for arbitrary orthogonal multiple-antenna constellations. Therefore, inspired by the Stein's lemma [10], we will use, as our performance criterion, the upper bound on the exponential decay rate of the pairwise error probability, given by the Kullback-Leibler (KL) distance [10] between conditional distributions. If p_1 and p_2 are two probability density functions on the probability space $(\mathcal{X}, \mathcal{F})$, then the KL distance between them is defined as

$$\begin{aligned} \mathcal{D}(p_1 \| p_2) &= \mathbb{E}_{p_1} \left\{ \ln \left(\frac{p_1(x)}{p_2(x)} \right) \right\} \\ &= \int_{\mathcal{X}} p_1(x) \ln \left(\frac{p_1(x)}{p_2(x)} \right) dx, \end{aligned} \quad (18)$$

where \mathbb{E}_{p_1} denotes expectation with respect to p_1 . We also use the notation $\Pr_p \{\mathcal{R}\}$ to denote the probability of set $\mathcal{R} \in \mathcal{F}$ with respect to the probability density function p .

Stein's lemma [10] relates the KL distance with the pairwise error probabilities of hypothesis testing:

Lemma 2 (Stein's lemma): Let $X_1, X_2, \dots, X_N \in \mathcal{X}$ be drawn i.i.d. according to the probability density function q on \mathcal{X} . Consider the hypothesis test between $q = p_1$ and $q = p_2$, where p_1 and p_2 are probability density functions on \mathcal{X} , and $\mathcal{D}(p_1 \| p_2) < \infty$. Let $\mathcal{A}_N \subseteq \mathcal{X}^N$ be an acceptance region for hypothesis 1. Denote the probabilities of error by

$$\alpha_N = \Pr(p_1 \rightarrow p_2) = \Pr_{p_1} \{\mathcal{A}_N^c\}, \quad (19)$$

$$\beta_N = \Pr(p_2 \rightarrow p_1) = \Pr_{p_2} \{\mathcal{A}_N\}, \quad (20)$$

and define

$$\beta_N^\epsilon = \min_{\mathcal{A}_N \subseteq \mathcal{X}^N, \alpha_N < \epsilon} \beta_N. \quad (21)$$

Then

$$\lim_{\epsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{N} \ln \beta_N^\epsilon = -\mathcal{D}(p_1 \| p_2). \quad (22)$$

In other words, the best achievable error exponent for $\Pr(S_2 \rightarrow S_1)$ with the constraint that $\Pr(S_1 \rightarrow S_2)$ is smaller than a given value, is given by $\mathcal{D}(p(X|S_1) \| p(X|S_2))$. It turns out that this error exponent is not achieved with the maximum likelihood detector, but with a detector which is highly biased in favor of the second hypothesis. Nevertheless, it serves as an upper bound on the pairwise error exponent of the ML detector. The following lemma shows that the performance of the ML detector is, in fact, related to the KL distance between the distributions. (As we saw in Section III, at least in fast fading, i.e., when $T = 1$, the KL-based design criterion is equivalent to the design criterion based on the exact pairwise error probability and also the Chernoff bound.)

Lemma 3: Let $X_1, X_2, \dots, X_N \in \mathcal{X}$ be drawn i.i.d. according to the probability density function p_0 on \mathcal{X} . Consider two hypothesis tests, one between $q = p_0$ and $q = p_1$, and the other between $q = p_0$ and $q = p_2$, where p_1 and p_2 are probability density functions on \mathcal{X} , and $0 <$

$\mathcal{D}(p_0 \| p_2) < \mathcal{D}(p_0 \| p_1) < \infty$. Let $L_{1N} = \prod_{n=1}^N \frac{p_0(X_n)}{p_1(X_n)}$ and $L_{2N} = \prod_{n=1}^N \frac{p_0(X_n)}{p_2(X_n)}$ denote the likelihood ratios for the two tests, so that the probabilities of mistaking p_0 for p_1 and p_2 using the ML detector are given by $P_{e1N} = \Pr_{p_0} \{L_{1N} < 1\}$ and $P_{e2N} = \Pr_{p_0} \{L_{2N} < 1\}$. Also denote by $\Delta\mathcal{D}$ the difference between the two KL distances, i.e., $\Delta\mathcal{D} = \mathcal{D}(p_0 \| p_1) - \mathcal{D}(p_0 \| p_2) > 0$. With these assumptions, we will have

$$\Pr_{p_0} \left\{ L_{1N} < e^{\frac{N\Delta\mathcal{D}}{2}} L_{2N} \right\} \rightarrow 0 \text{ as } N \rightarrow \infty. \quad (23)$$

Proof: We have

$$\begin{aligned} \frac{1}{N} \ln \left(\frac{L_{1N}}{L_{2N}} \right) &= \frac{1}{N} \ln \left[\prod_{n=1}^N \frac{p_0(X_n)/p_1(X_n)}{p_0(X_n)/p_2(X_n)} \right] \\ &= \frac{1}{N} \sum_{n=1}^N \ln \left[\frac{p_2(X_n)}{p_1(X_n)} \right]. \end{aligned} \quad (24)$$

Since X_i 's are assumed to be drawn i.i.d. according to the probability density function p_0 , by the weak law of large numbers, we have

$$\frac{1}{N} \sum_{n=1}^N \ln \left[\frac{p_2(X_n)}{p_1(X_n)} \right] \rightarrow \mathbb{E}_{p_0} \left\{ \ln \left[\frac{p_2(X_1)}{p_1(X_1)} \right] \right\} \quad \text{in probability w.r.t } p_0. \quad (25)$$

Multiplying the numerator and denominator of the argument of the $\ln(\cdot)$ function by $p_0(x)$, we have

$$\begin{aligned} \mathbb{E}_{p_0} \left\{ \ln \left[\frac{p_2(X_1)}{p_1(X_1)} \right] \right\} &= \mathbb{E}_{p_0} \left\{ \ln \left[\frac{p_0(X_1)}{p_1(X_1)} \right] \right\} - \\ &\quad \mathbb{E}_{p_0} \left\{ \ln \left[\frac{p_0(X_1)}{p_2(X_1)} \right] \right\} \\ &= \mathcal{D}(p_0 \| p_1) - \mathcal{D}(p_0 \| p_2) \\ &= \Delta\mathcal{D}. \end{aligned} \quad (26)$$

From equations (24), (25), and (26) we have

$$\frac{1}{N} \ln \left(\frac{L_{1N}}{L_{2N}} \right) \rightarrow \Delta\mathcal{D} \text{ in probability w.r.t } p_0, \quad (27)$$

which means that for any $\delta > 0$,

$$\Pr_{p_0} \left\{ \left| \frac{1}{N} \ln \left(\frac{L_{1N}}{L_{2N}} \right) - \Delta\mathcal{D} \right| > \delta \right\} \rightarrow 0 \text{ as } N \rightarrow \infty. \quad (28)$$

Let $\delta = \frac{\Delta\mathcal{D}}{2}$. From (28) we will have,

$$\begin{aligned} \Pr_{p_0} \left\{ \frac{1}{N} \ln \left(\frac{L_{1N}}{L_{2N}} \right) < \frac{\Delta\mathcal{D}}{2} \right. \\ \left. \text{or } \frac{1}{N} \ln \left(\frac{L_{1N}}{L_{2N}} \right) > \frac{3\Delta\mathcal{D}}{2} \right\} \rightarrow 0 \text{ as } N \rightarrow \infty, \end{aligned} \quad (29)$$

or

$$\Pr_{p_0} \left\{ L_{1N} < e^{\frac{N\Delta\mathcal{D}}{2}} L_{2N} \right\} \rightarrow 0 \text{ as } N \rightarrow \infty. \quad (30)$$

The above lemma states that, for sufficiently large N , with high probability the likelihood ratio of the first test is greater than the likelihood ratio of the second test, and the ratio of the two likelihood ratios grows exponentially with N . Recalling that the error probability of each test is the probability that its

corresponding likelihood ratio is smaller than one, this implies that for large N , the first test will have a lower probability of error than the second test.

In the above lemma, N is the number of independent observations. In our case, independent observations can be obtained by using an outer code which operates over several independent fading intervals, or simply by using multiple receive antennas.

In Appendix C, we show that the KL distance between p_i^N and p_j^N (obtained by substituting S_i and S_j for S in (3)), is given by

$$\mathcal{D}(p_i^N \| p_j^N) = N \text{tr} \{ (I_T + S_i S_i^H)(I_T + S_j S_j^H)^{-1} \} - N \ln \det \{ (I_T + S_i S_i^H)(I_T + S_j S_j^H)^{-1} \} - NT. \quad (31)$$

Adopting the KL distance as performance criterion, the signal set design criterion in general will be maximization of the minimum KL distance between conditional distributions corresponding to the signal points, i.e., assuming equiprobable signal points,

$$\underset{\frac{1}{T} \sum_{i=1}^T \|S_i\|^2 \leq T P}{\text{maximize}} \quad \underset{i \neq j}{\min} \quad \mathcal{D}(p_i \| p_j), \quad (32)$$

where $\|S_i\|^2 = \sum_{t=1}^T \sum_{m=1}^M |(S_i)_{tm}|^2$ is the Frobenious norm of S_i , or the total power used to transmit S_i .

If we denote, by $\lambda_{i,j}(t)$, $t = 1, \dots, T$, the T eigenvalues of $(I_T + S_i S_i^H)(I_T + S_j S_j^H)^{-1}$, the KL distance in (31) can be written as

$$\mathcal{D}(p_i^N \| p_j^N) = N \sum_{t=1}^T \{ \lambda_{i,j}(t) - \ln(\lambda_{i,j}(t)) - 1 \}. \quad (33)$$

This expression, in spite of its notational simplicity and also its resemblance to the well-known *rank* and *determinant* criteria of coherent space-time codes [4], does not provide much insight into the design problem. Moreover, the power constraint does not appear to be easily expressible in terms of the above eigenvalues. Therefore, in the next section, we will try to approach the design problem by imposing some extra constraints on the signal set and directly simplifying the original expression in (31). Since the actual value of N does not affect the maximization in (32), in designing the signal constellations we will always assume that $N = 1$.

V. SIGNAL SET CONSTRUCTION

In Theorem 1, we showed that any error probability performance achievable by a constellation of arbitrary matrices, can also be achieved by a constellation of orthogonal matrices. Therefore, in this work we will only consider matrix constellations with orthogonal columns. In Appendix D, we show that with this assumption, the KL distance expression in (31) can be written as

$$\mathcal{D}(p_i \| p_j) = \sum_{m=1}^M \left\{ \frac{1 + \|S_{im}\|^2}{1 + \|S_{jm}\|^2} - \ln \left(\frac{1 + \|S_{im}\|^2}{1 + \|S_{jm}\|^2} \right) - 1 + \frac{\|S_{im}\|^2 \|S_{jm}\|^2 - \sum_{k=1}^M |S_{ik} \cdot S_{jk}|^2}{1 + \|S_{jm}\|^2} \right\}, \quad (34)$$

where we have used the notation S_{lm} to denote the m th column of S_l . The expressions for KL distance in (34) is not very illuminating as it is. Therefore, we study the signal set construction problem through a series of special cases. These special cases will provide an understanding of the nature of the KL distance in (31) by breaking it down into simpler components. In most cases, this results in a systematic technique for constellation design.

Notice that Sections V-B and V-D correspond to single-antenna and multiple-antenna unitary constellations, and are special sub-cases of Sections V-C and V-E, respectively. It is shown in these sections that, assuming orthonormal signal matrices, the KL-based design criterion reduces to the previously proposed design criterion for the unitary constellations [6], [8]. Therefore, unitary designs can be considered as special cases of the more general constellations designed using the KL-based criterion. Simulation results corresponding to Sections V-B and V-D are presented in Sections V-C and V-E, respectively, where the unitary designs are compared with their multi-level versions designed using the KL-based criterion.

A. Fast Fading ($T = 1$)

As stated in Theorem 1, there is no gain in using more than T transmit antennas. Therefore, in this case we consider only single transmit antenna systems, where signal matrices are complex scalars. The KL distance of (34) reduces to

$$\mathcal{D}_1(p_i \| p_j) = \frac{1 + |s_i|^2}{1 + |s_j|^2} - \ln \left(\frac{1 + |s_i|^2}{1 + |s_j|^2} \right) - 1. \quad (35)$$

It can be easily verified that, similar to the pairwise error probability (12) and the Chernoff information (16), the KL distance is also a monotonic function of $B = \frac{1 + |s_i|^2}{1 + |s_j|^2}$ in the two regions of $\{B < 1\}$ and $\{B > 1\}$. Therefore, for a single transmit antenna system in fast fading, maximizing the minimum of the KL distance is equivalent to minimizing the maximum of the exact pairwise error probability as well as the Chernoff bound. The following theorem characterizes the solution to the maximin problem in this case.

Theorem 3: The solution to the maximin problem (32) for the case of $M = 1$ and $T = 1$, is given by $|s_l|^2 = \alpha^{l-1} - 1$, where α is the largest real root of the polynomial

$$f(\alpha) = \alpha^L - L(P+1)\alpha + (LP + L - 1). \quad (36)$$

Proof: See Appendix E. ■

Notice that, since $f(1) = 0$, $f'(1) = -LP < 0$, and $\lim_{\alpha \rightarrow +\infty} f(\alpha) = +\infty$, this polynomial always has a real root larger than one.

The constellations obtained from Theorem 3 are PAM-type constellations but with unequal spacing between the signal points, and the first point is always at the origin. The interesting fact is that even the relative locations of the constellation points depend on the SNR, and two constellations of the same size designed for different SNR values are not necessarily scaled versions of each other. Fig. 5 shows the locations of the signal points for a 4-point constellation vs. average transmit power. As we see, the spacings between pairs of consecutive points are not equal. However, in terms of the

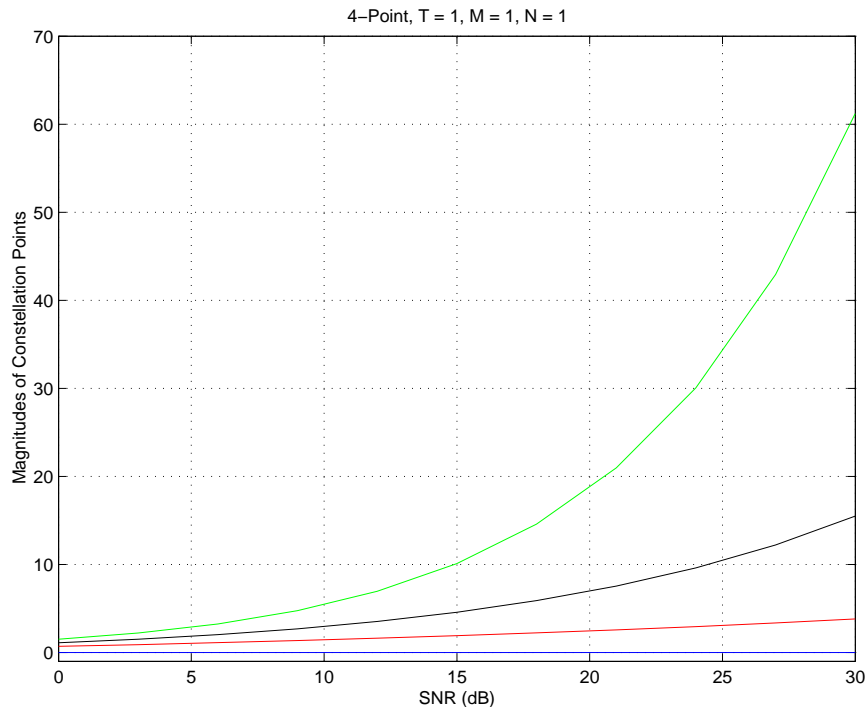


Fig. 5. Magnitudes of the optimal signal points for a 4-point constellation with $M = 1$, $T = 1$, and $N = 1$.

KL distance, these points are, by construction, equally spaced. At high SNR, the outer points have to be placed farther apart than the inner points, to maintain a constant KL distance. Therefore, for a PAM constellation with equally spaced points, the outer points have a smaller KL distance than the inner points. In fact, it can be easily shown that in (35), if the ratio of the magnitudes of two constellation points is constant, by increasing SNR the KL distance between those two points converges to a finite constant. This results in an error floor for a PAM constellation as shown in Fig. 6. However, as we see in this figure, the optimal constellation does not see any error floor.

B. $T > 1$, $M = 1$, and $\|S_l\|^2 = TP$ for $l = 1, \dots, L$

This is the case of single transmit antenna systems in block fading environment, with constellation points which are column vectors and all lie on a sphere in \mathbb{C}^T . Since all of the points have the same magnitude, these constellations can be considered as single antenna unitary constellation. The KL distance of (34) reduces to

$$\begin{aligned} \mathcal{D}_2(p_i \| p_j) &= \frac{\|S_i\|^2 \|S_j\|^2 - |S_i \cdot S_j|^2}{1 + \|S_j\|^2} \\ &= \frac{(TP)^2 \sin^2(\angle S_i, S_j)}{1 + TP}, \end{aligned} \quad (37)$$

where \cdot is the inner product operation, and $\angle S_i, S_j$ denotes the angle between signal vectors S_i and S_j . This distance depends only on the angle between the signal points.

The optimum constellation in this case, is obviously the one that is designed to maximize the minimum angle between subspaces spanned by the signal points (or equivalently, minimizes the maximum absolute inner product or correlation between

signal points). This is the same design criterion proposed for unitary constellations [7], [8] if only one transmit antenna is considered. Examples of such designs can also be found in [7] and [8].

For $T = 2$, if we confine ourselves to real constellations, the above criterion results in the signal set $\left\{ \left[\begin{array}{c} \cos((l-1)\pi/L) \\ \sin((l-1)\pi/L) \end{array} \right] \right\}_{l=1}^L$, which is the same as the signal set proposed in [8]. As also mentioned in [8], these so-called PSK constellations, have the advantage of low complexity decoding based on a single phase calculation and quantization. Therefore, we will use these constellations for a more general design explained in the next subsection. Notice that the angle between adjacent points is π/L , not $2\pi/L$. This is because this angle is actually the angle between subspaces containing the constellation points, and thus has to be considered modulo π .

C. $T \geq 1$ and $M = 1$

This is the general case for single-antenna constellations. The KL distance in (34) reduces to

$$\begin{aligned} \mathcal{D}(p_i \| p_j) &= \underbrace{\frac{1 + \|S_i\|^2}{1 + \|S_j\|^2} - \ln \left(\frac{1 + \|S_i\|^2}{1 + \|S_j\|^2} \right)}_{\mathcal{D}_1(p_i \| p_j)} - 1 \\ &\quad + \underbrace{\frac{\|S_i\|^2 \|S_j\|^2 \sin^2(\angle S_i, S_j)}{1 + \|S_j\|^2}}_{\frac{\|S_i\|^2}{\|S_j\|^2} \mathcal{D}_2(p_i \| p_j)}. \end{aligned} \quad (38)$$

As we see, the KL distance between any two points consists of two parts: $\mathcal{D}_1(p_i \| p_j)$ due to having different magnitudes (lying

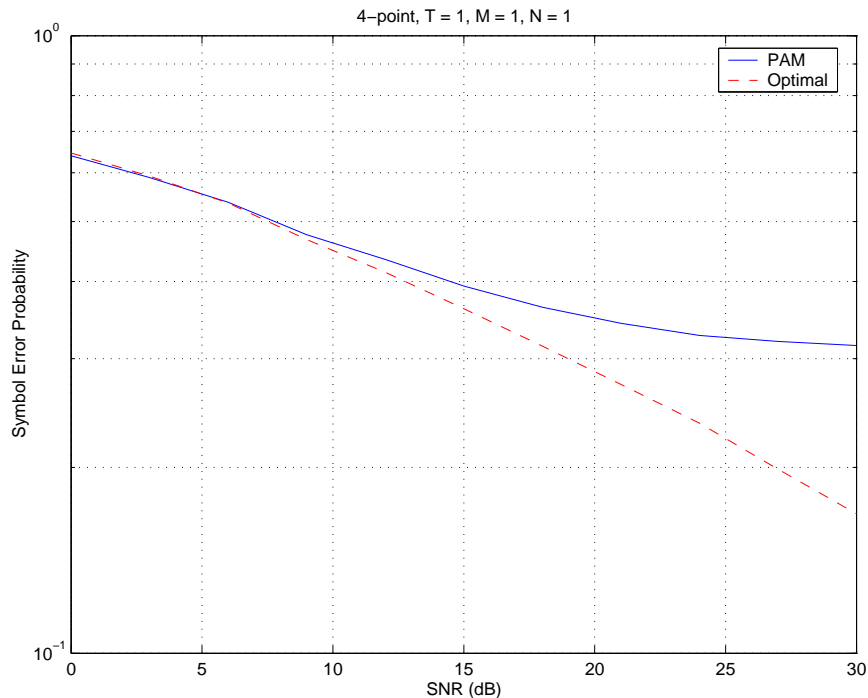


Fig. 6. Symbol error rate comparison with regular PAM for a 4-point constellation with $M = 1$, $T = 1$, and $N = 1$.

on different spheres in \mathbb{C}^T , and $\frac{\|S_i\|^2}{\|S_j\|^2} \mathcal{D}_2(p_i \| p_j)$ due to the angle between the points (lying on different one-dimensional subspaces of \mathbb{C}^T). If two points lie on the same sphere, $\mathcal{D}_1(p_i \| p_j) = 0$, and if they lie on the same complex plane (one dimensional subspace), $\mathcal{D}_2(p_i \| p_j) = 0$. In general, the overall distance is greater than or equal to either of these parts. This property of the KL distance in (38) suggests partitioning the signal space into subsets of concentric spheres C_1, \dots, C_K , of radius r_1, \dots, r_K , containing l_1, \dots, l_K points, respectively, and defining the intra-subset and inter-subset distances as

$$\mathcal{D}_{intra}(k) = \min_{S_i, S_j \in C_k} \frac{r_k^4 \sin^2(\angle S_i, S_j)}{1 + r_k^2}, \quad (39)$$

and

$$\mathcal{D}_{inter}(k, k') = \frac{1 + r_k^2}{1 + r_{k'}^2} - \ln \left(\frac{1 + r_k^2}{1 + r_{k'}^2} \right) - 1. \quad (40)$$

Without loss of generality, we can assume that $r_1 < r_2 < \dots < r_K$. With this assumption, and using the fact that if $r_{k''} > r_{k'} > r_k$ then $\mathcal{D}_{inter}(k, k'') > \mathcal{D}_{inter}(k, k')$, we can reformulate the design problem as the following suboptimal maximin problem over K , l_1, \dots, l_K , and r_1, \dots, r_K :

$$\begin{aligned} & \text{maximize} && \min \mathcal{A}, \\ & 1 \leq K \leq L, \frac{1}{L} \sum_{k=1}^K l_k r_k^2 \leq TP, \sum_{k=1}^K l_k = L && (41) \\ & 0 \leq r_1 < r_2 < \dots < r_K \end{aligned}$$

where

$$\mathcal{A} = \left\{ \left\{ \mathcal{D}_{intra}(k) \right\}_{k=1}^K, \left\{ \mathcal{D}_{inter}(k, k+1) \right\}_{k=1}^{K-1} \right\}. \quad (42)$$

The suboptimality of this approach comes from the fact that in the above formulation of $\mathcal{D}_{inter}(k, k')$, it is assumed that for any two subsets C_k and C_{k+1} , there are two constellation

points $S_i \in C_k$ and $S_j \in C_{k+1}$, such that $\mathcal{D}(p_i \| p_j) = \mathcal{D}_{inter}(k, k+1)$. This assumption is not necessarily true for the optimal constellation. However, since $\mathcal{D}(p_i \| p_j) \geq \mathcal{D}_1(p_i \| p_j)$, it is guaranteed that the actual minimum KL distance of the resulting constellation from the above optimization will not be smaller than the minimum in (41). Moreover, with this assumption, for a fixed K and fixed l_1, \dots, l_K such that $\sum_{l=1}^K l_k = L$, we can decouple the original optimization problem into the following simpler problems:

- 1) For each subset (each $k \in \{1, \dots, K\}$), find the best configuration of l_k points on the surface of the k th sphere, C_k , i.e., maximize the minimum intra-subset distance inside the k th subset. Notice that r_k appears in $\mathcal{D}_{intra}(k)$ only through a multiplicative factor, and does not affect this optimization. Also notice that this step is equivalent to designing a single antenna unitary constellation of l_k points (see Section V-B), and any existing unitary design (e.g., [7] or [8]) can be used at this step.
- 2) Solve the following continuous optimization to find the radiuses of the subsets:

$$\begin{aligned} & \text{maximize} && \min \mathcal{A}, \\ & 0 \leq r_1 < r_2 < \dots < r_K && (43) \\ & \frac{1}{L} \sum_{k=1}^K l_k r_k^2 \leq TP \end{aligned}$$

where \mathcal{A} is as in (42). This optimization problem can be solved numerically, e.g., using the `fminimax` function of the Matlab program (which uses a Sequential Quadratic Programming method to solve the non-linear constrained optimization problems).

The solution to the problem in (41) can then be obtained by searching over all possible values for K and l_1, \dots, l_K such

that $\sum_{l=1}^K l_k = L$. The following proposition can be used to further restrict the domain of search:

Proposition 2: The solution of (41) satisfies the following inequalities:

$$l_1 \leq l_2 \leq \dots \leq l_{K-1}. \quad (44)$$

Proof: Let $\{K, l_1, \dots, l_K, r_1, \dots, r_K\}$ be the solution of (41), with $l_K \geq 1$. Suppose, for the sake of contradiction, that $l_k > l_{k+1}$ for some $k \in \{1, \dots, K-2\}$. Now, by removing one point from subset $k+2$ and adding it to subset $k+1$ (and rearranging the points in these two subsets to maximize the minimum intra-subset distances), and specifying the parameters of the new constellation by a "''" sign (e.g., $l'_{k+1} = l_{k+1} + 1$ and $l'_{k+2} = l_{k+2} - 1$), we will have

- 1) \mathcal{D}_{intra} is an increasing function of the subset radius, and a non-increasing function of the subset size (number of points in the subset). Therefore, since $r'_{k+1} > r'_k$ and $l'_{k+1} \leq l'_k$, we have

$$\mathcal{D}'_{intra}(k+1) > \mathcal{D}'_{intra}(k) = \mathcal{D}_{intra}(k), \quad (45)$$

and since $r'_{k+2} = r_{k+2}$ and $l'_{k+2} < l_{k+2}$ we have,

$$\mathcal{D}'_{intra}(k+2) > \mathcal{D}_{intra}(k+2). \quad (46)$$

Since all other intra-subset distances and also the inter-subset distances are not affected by this change, the overall minimum KL distance of the new constellation will be greater than or equal to the minimum KL distance of the original constellation.

- 2) $\sum_{i=1}^K l'_i (r'_i)^2 - \sum_{k=1}^K l_k r_k^2 = (r'_{k+1})^2 - r_{k+2}^2 = r_{k+1}^2 - r_{k+2}^2 < 0$. Therefore, the average power of the new constellation is smaller than the average power of the original constellation.

Now, by appropriately scaling the new constellation, we can make the average powers of the two constellations equal, and obtain a new constellation which has a larger minimum KL distance. This is a contradiction with our initial assumption. Therefore, the solution of (41) should satisfy (44). ■

For a fixed K , the collection of all of the possible K -tuples (l_1, \dots, l_K) , satisfying $\sum_{l=1}^K l_k = L$ and (44), can be found recursively. The details are omitted here for brevity. The optimization in (41) can then be solved through the following steps:

- 1) Set $K = 1$.
- 2) Find the collection of all of the possible K -tuples (l_1, \dots, l_K) , satisfying $\sum_{l=1}^K l_k = L$ and (44).
- 3) For each member of the above collection, perform the following steps, and find the best achievable minimum distance:
 - a) For each subset (each $k \in \{1, \dots, K\}$), find the best configuration of l_k points on the surface of the k th sphere, C_k , i.e., maximize the minimum intra-subset distance inside the k th subset (unitary design).
 - b) Solve the continuous optimization in (43) to find the radiuses of the subsets.
- 4) Store the parameters of the constellation with the largest minimum KL distance from the previous step as the best candidate with K levels.

- 5) Increase K by one. If $K \leq L$ go to the second step.
- 6) Among all of the above L candidates, choose the constellation with the largest minimum KL distance, as the solution of (41).

Assuming that the best unitary constellations (of the type mentioned in Section V-B, and) of arbitrary size are known, this approach significantly simplifies the design problem by reducing the number of design parameters from $2LT$ (real and imaginary parts of the elements of the constellation vectors), to $2K+1$ (number of the subsets, and radius and number of the points in each subset).

Notice that, since unlike the square Euclidean distance, the KL distance does not scale with the average power of the constellation, the structure of the optimal constellation based on the above criterion depends on the actual value of the signal to noise ratio, and constellations of the same size at different SNR values are not scaled versions of each other. It is also worthwhile to notice that at high SNR (for large values of r_k or r_{k+1} or both), we have the following approximations

$$\begin{aligned} \mathcal{D}_{intra}(k) &\approx r_k^2 \sin^2(\angle S_i, S_j), \\ \mathcal{D}_{inter}(k, k+1) &\approx \begin{cases} a - \ln(a) - 1 & \text{if } \frac{r_k^2}{r_{k+1}^2} = a, \\ \ln(r_{k+1}^2) & \text{if } r_k^2 \text{ is kept fixed.} \end{cases} \end{aligned} \quad (47)$$

This means that $\mathcal{D}_{intra}(k)$ increases almost linearly with SNR, whereas $\mathcal{D}_{inter}(k, k+1)$ either approaches a constant value, or increases at most logarithmically with SNR. As a result, at high SNR, having multiple levels is not desirable, and one should only consider constellations of constant magnitude. In this case, the KL-based design criterion reduces to the design criterion for the single antenna unitary constellations (see Section V-B), confirming the high SNR optimality of the unitary constellations.

The decoding of the multi-level unitary constellations proposed in this section can be done in a similar way to that of trellis coded modulation schemes, i.e., in two steps of "point in subset decoding" and "subset decoding". If a unitary code with low decoding complexity, such as the schemes described in [8], is used inside each subset, then the point in subset decoding step can be done at a very low cost, and considering the fact that the number of subsets is usually much smaller than the size of the whole constellation, the overall decoding complexity of the code will be much lower than the regular ML decoder.

For the special case of real constellations with $T = 2$, the angle between adjacent points in the k th subset is simply π/l_k (see Section V-B), and the maximin problem in (41) is relatively easy to solve. The resulting 2, 4, 8 and 16-point constellations with average powers of 0.5 and 5 are shown in Figs. 7 and 8. Each axis in these figures actually represents a complex plane corresponding to one transmit symbol interval. The symbol error rate performance of the 8 and 16-point constellations at $P = 5$ (SNR $\simeq 7$ dB) are simulated for different values of N and compared with the corresponding constellations proposed in [8]. The results are shown in Figs. 9 and 10. As expected, due to the larger minimum KL distance of the new constellations, the exponential decay of the symbol error rate vs. N has a much higher rate for the

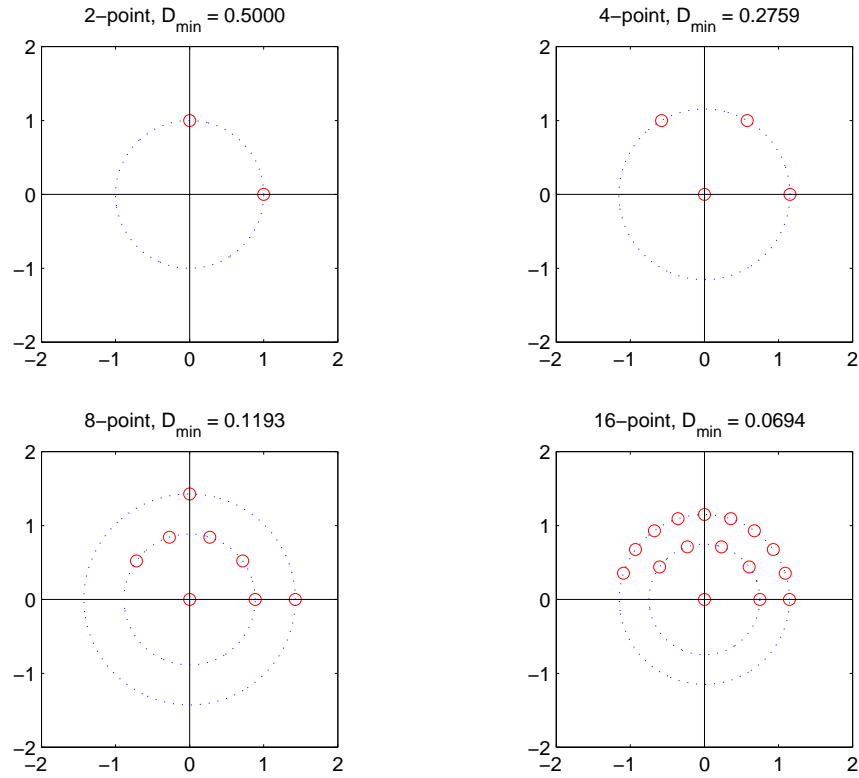


Fig. 7. KL-optimal constellations of size 2, 4, 8, and 16 for $M = 1$, $T = 2$, and $P = 0.5$.

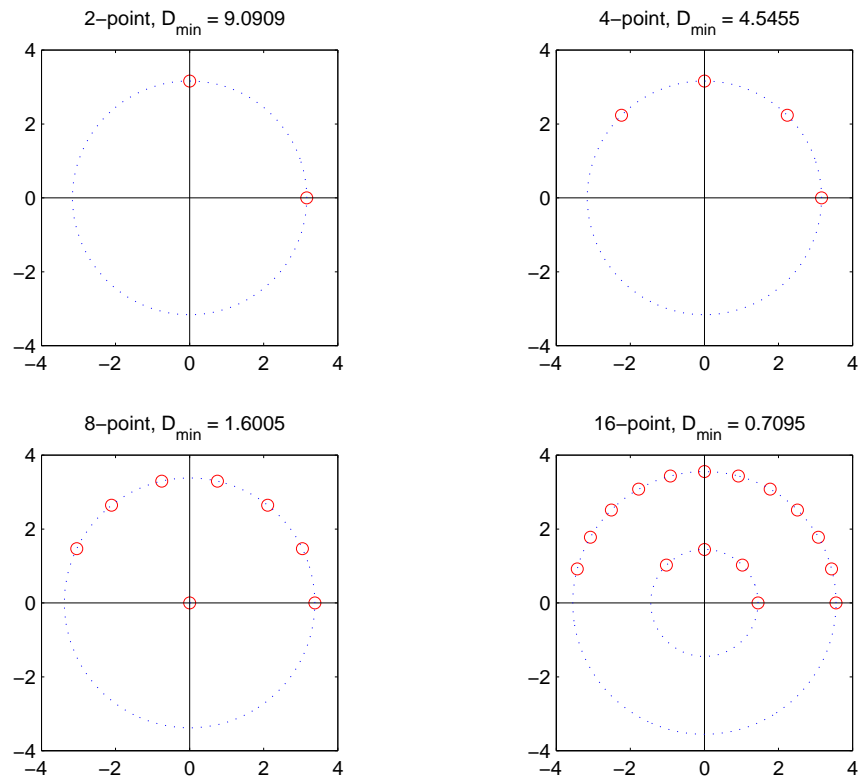


Fig. 8. KL-optimal constellations of size 2, 4, 8, and 16 for $M = 1$, $T = 2$, and $P = 5$.

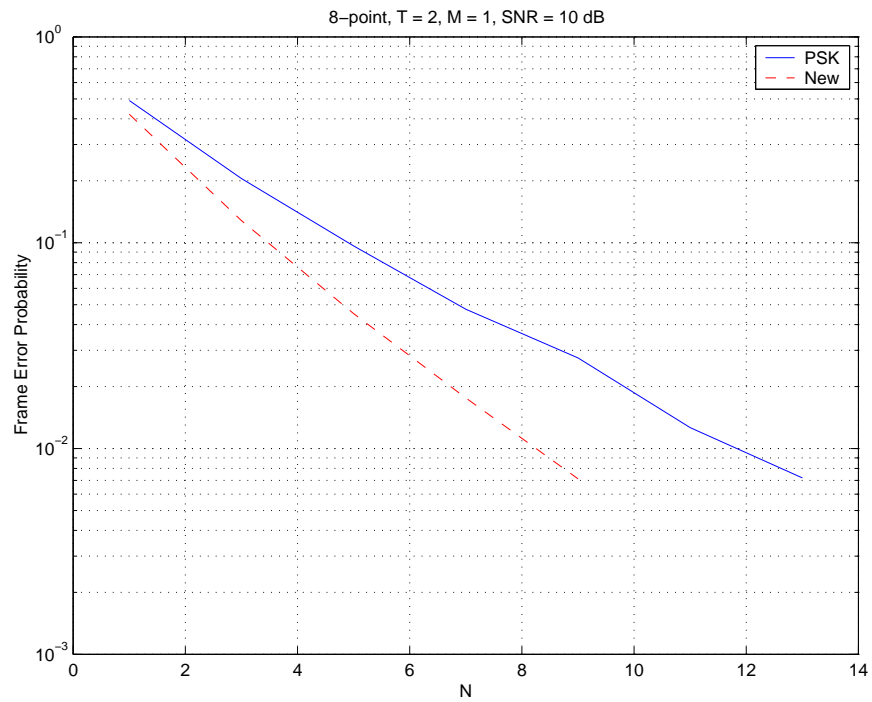


Fig. 9. Symbol error rate for an 8-point constellation for $M = 1$, $T = 2$, and SNR = 7 dB.

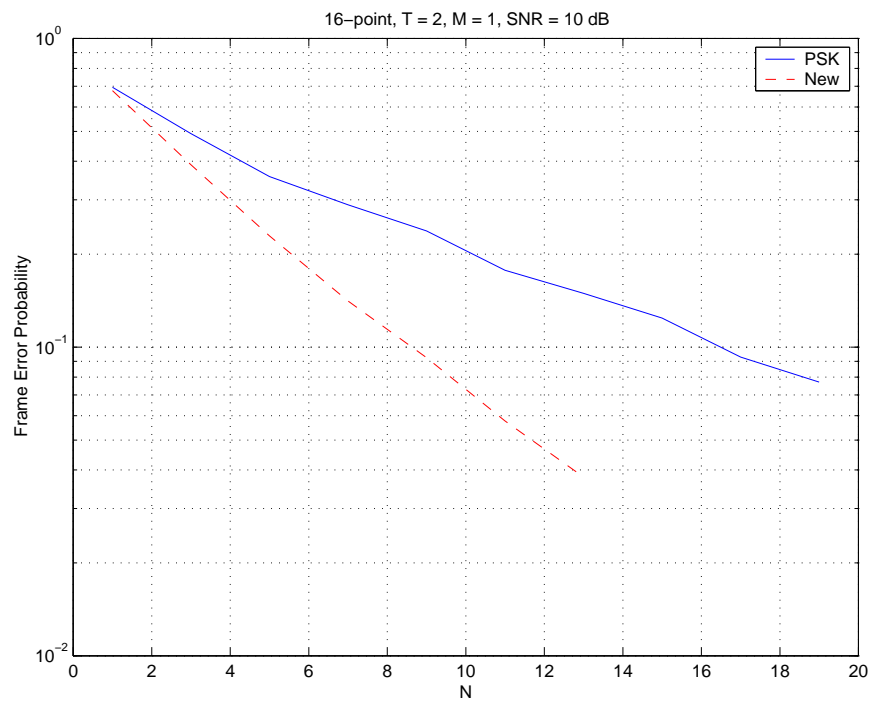


Fig. 10. Symbol error rate for a 16-point constellations for $M = 1$, $T = 2$, and SNR = 7 dB.

new constellations. The minimum KL distances of the new constellations are 1.6005 and 0.7095 for 8-point and 16-point constellations, respectively, whereas the corresponding PSK constellations of [8] have minimum KL distances of 1.3313 and 0.3460, respectively.

D. $T \geq 1$, $M \geq 1$, and $S_l^H S_l = \frac{TP}{M} I_M$ for $l = 1, \dots, L$

This is the case of unitary constellations. Since all of the columns of S_i and S_j have the same square magnitude, $\frac{TP}{M}$, the KL distance in (34) reduces to

$$\begin{aligned} \mathcal{D}(p_i \| p_j) &= \sum_{m=1}^M \frac{(TP/M) \|S_{jm}\|^2 - \sum_{k=1}^M |S_{ik} \cdot S_{jm}|^2}{1 + TP/M} \\ &= \frac{TP}{M + TP} \sum_{m=1}^M \left\{ \|S_{jm}\|^2 - \sum_{k=1}^M \frac{|S_{jm} \cdot S_{ik}|^2}{\|S_{ik}\|^2} \right\} \\ &= \frac{TP}{M + TP} \sum_{m=1}^M d_E^2(S_{jm}, W_{S_i}) \\ &= \frac{(TP)^2}{M(M + TP)} d_E^2(W_{S_i}, W_{S_j}), \end{aligned} \quad (48)$$

where W_{S_i} and W_{S_j} denote the subspaces of \mathbb{C}^T spanned by columns of S_i and S_j , respectively, $d_E(S_{jm}, W_{S_i})$ is the Euclidean distance of vector S_{jm} from subspace W_{S_i} , and $d_E(W_{S_i}, W_{S_j})$ is the *Euclidean distance* of subspaces W_{S_i} and W_{S_j} , as defined in [8]. As we see, for unitary constellations, the KL-based design criterion reduces to the Euclidean-based design criterion, and therefore, the new non-coherent space-time constellations include the existing unitary constellations as a special case.

In Appendix F, we show that the *Euclidean distance* defined in [8] and the *chordal distance* defined in [13] are equivalent. Therefore, the unitary constellations are, in fact, packings in complex Grassmannian manifolds. In [9], it has been shown that, at high SNR, the calculation of capacity of the non-coherent multiple-antenna channel can also be viewed as sphere packing in the product space of Grassmannian manifolds.

E. $T \geq 1$, $M \geq 1$, and $S_l^H S_l = d_l I_M$ for $l = 1, \dots, L$

The assumption in this case is that each signal matrix is a scalar multiple of a unitary matrix. With this assumption, the KL distance in (31) reduces to

$$\begin{aligned} \mathcal{D}(p_i \| p_j) &= M \underbrace{\left[\frac{1 + d_i}{1 + d_j} - \ln \left(\frac{1 + d_i}{1 + d_j} \right) - 1 \right]}_{\mathcal{D}_1(p_i \| p_j)} \\ &\quad + \underbrace{\frac{d_i d_j}{1 + d_j} d_E^2(W_{S_i}, W_{S_j})}_{\frac{d_i}{d_j} \mathcal{D}_2(p_i \| p_j)}. \end{aligned} \quad (49)$$

where $d_E^2(W_{S_i}, W_{S_j})$ is the square *Euclidean distance* [8] or *chordal distance* [13] between the two subspaces W_{S_i} and

W_{S_j} spanned by columns of S_i and S_j , defined as

$$\begin{aligned} d_E^2(W_{S_i}, W_{S_j}) &= \sum_{m=1}^M d_E^2\left(\frac{S_{im}}{\sqrt{d_i}}, W_{S_j}\right) \\ &= \sum_{m=1}^M \left\{ \frac{\|S_{im}\|^2}{d_i} - \sum_{k=1}^M \frac{|S_{im} \cdot S_{jk}|^2}{d_i d_j} \right\}, \end{aligned} \quad (50)$$

\mathcal{D}_1 denotes the distance between two constellation points which represent the same M -dimensional subspace of the T -dimensional space, and \mathcal{D}_2 denotes the distance between two constellation points with the same power which represent two different M -dimensional subspaces. In general, the overall distance is greater than or equal to either of these parts. Recalling that the unitary constellations are designed to maximize the *Euclidean distance* between subspaces, the above partitioning of the KL distance suggests partitioning the signal space into subsets of unitary constellations, C_1, \dots, C_K , with columns of square norm ρ_1, \dots, ρ_K , containing l_1, \dots, l_K points, respectively. Similar to the approach of Section V-C, we define the intra-subset and inter-subset distances as

$$\mathcal{D}_{intra}(k) = \min_{S_i, S_j \in C_k} \frac{\rho_k^2}{1 + \rho_k} d_E^2(S_i, S_j), \quad (51)$$

and

$$\mathcal{D}_{inter}(k, k') = M \left[\frac{1 + \rho_k}{1 + \rho_{k'}} - \ln \left(\frac{1 + \rho_k}{1 + \rho_{k'}} \right) - 1 \right]. \quad (52)$$

Without loss of generality, we can assume that $\rho_1 < \rho_2 < \dots < \rho_K$, and solve the simplified maximin problem

$$\begin{aligned} &\text{maximize} && \min \mathcal{B} \\ &1 \leq K \leq L, \frac{M}{L} \sum_{k=1}^K l_k \rho_k = TP, \sum_{k=1}^K l_k = L && (53) \\ &0 \leq \rho_1 < \rho_2 < \dots < \rho_K \end{aligned}$$

where

$$\mathcal{B} = \left\{ \left\{ \mathcal{D}_{intra}(k) \right\}_{k=1}^K, \left\{ \mathcal{D}_{inter}(k, k+1) \right\}_{k=1}^{K-1} \right\}. \quad (54)$$

to find the L -point multilevel unitary constellation of $T \times M$ matrices with average power P . At each level, we can use any existing unitary construction and substitute, for $\mathcal{D}_{intra}(k)$ in (51), the best achievable KL distance with that construction and with size l_k .

As explained in Section V-C for the case of single transmit antenna, in (53), K and l_1, \dots, l_K are discrete variables, while ρ_1, \dots, ρ_K are continuous variables. For any fixed value of K and l_1, \dots, l_K satisfying the specified constraints, (53) reduces to a continuous optimization over ρ_1, \dots, ρ_K , which can be solved numerically. Moreover, as shown in Proposition 2, the solution of (53) also satisfies the extra constraint $l_1 \leq l_2 \leq \dots \leq l_{K-1}$, which can be used to further restrict the domain of search.

Similar to the case of single transmit antenna (Section V-C), from (49) we also observe that at high SNR, \mathcal{D}_1 becomes a constant, or its minimum grows at most logarithmically with SNR (when d_i is kept fixed), whereas \mathcal{D}_2 grows linearly with SNR for non-zero constellation points from different subspaces. As a result, at high SNR, \mathcal{D}_2 becomes the dominant term, and the KL-based design criterion reduces to the

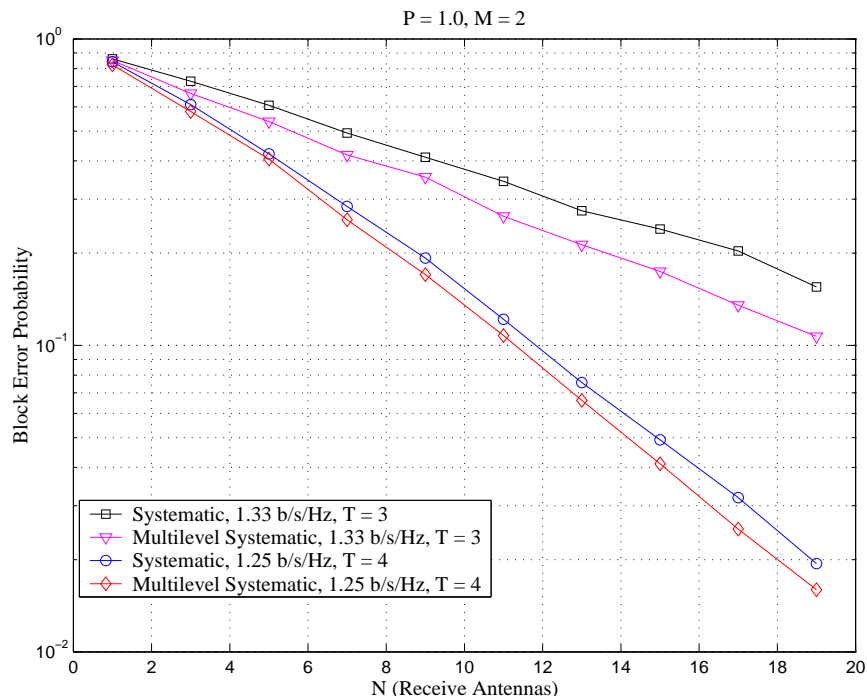


Fig. 11. Performance comparison of one and two transmit antenna systematic constellations of [7] and their multilevel versions vs. number of receive antennas.

Euclidean-based design criterion of the unitary constellations, confirming the high SNR optimality of the unitary constellations.

Figs. 11 and 12 show the error rate performance comparison of the proposed constellations with their unitary counterparts. In these examples, we have used the systematic unitary designs of [7] as the constituent subsets of the multilevel constellations. Fig. 11 shows the block error rate performance of the 16 and 32-point two-antenna constellations with $T = 3$ and $T = 4$ (resulting in spectral efficiencies of 1.33 and 1.25 b/s/Hz), respectively. The horizontal axis is the number of receive antennas, with SNR kept fixed at 0 dB. As we see, the multilevel constellations can save up to 4 receive antennas at SNR's as low as 0 dB.

Fig. 12 compares the performances of 16-point, one and two-antenna constellations for $T = 2$ and $T = 3$ (resulting in spectral efficiencies of 2 and 1.33 b/s/Hz), respectively. The horizontal axis is SNR, and the receiver is assumed to have 10 receive antennas. Even if multiple receive antennas are not available, similar gains can be obtained by encoding across several fading blocks using an outer code. For each point in the curves corresponding to the multilevel constellations, a separate optimization problem with appropriate power constraint has been solved and the resulting constellation has been used to evaluate the performance. We observe that the multilevel unitary constellation can provide up to 3 dB gain over its corresponding one-level unitary constellation at low SNR. We also notice that as SNR increases, the two curves become closer, which is expected, recalling the optimality of the unitary constellations at high SNR.

VI. CONCLUSIONS

We considered the problem of non-coherent communication in a Rayleigh flat fading environment using a multiple antenna system. We derived the design criterion for space-time constellations in this scenario based on the Kullback-Leibler distance between the distributions of the received signal conditioned on different transmitted values. We showed that close-to-optimal constellations according to the proposed criterion can be obtained by partitioning the signal space into appropriate subsets and using unitary designs inside each subset. We designed new non-coherent constellations based on the proposed criterion, and through simulations, showed that they can provide a substantial improvement in the performance over known unitary space-time constellations, especially at low SNR and when multiple receive antennas are used. We showed that unitary designs can be considered as special cases of the proposed constellations when the signal to noise ratio is high.

APPENDIX A

EXACT PAIRWISE ERROR PROBABILITY FOR FAST FADING

In this appendix, we prove that the expression for the exact pairwise error probability of the single transmit antenna system in fast fading is given by (12). For convenience, we use the following notation for the received vector:

$$X^N = [x_1 \cdots x_N]. \quad (55)$$

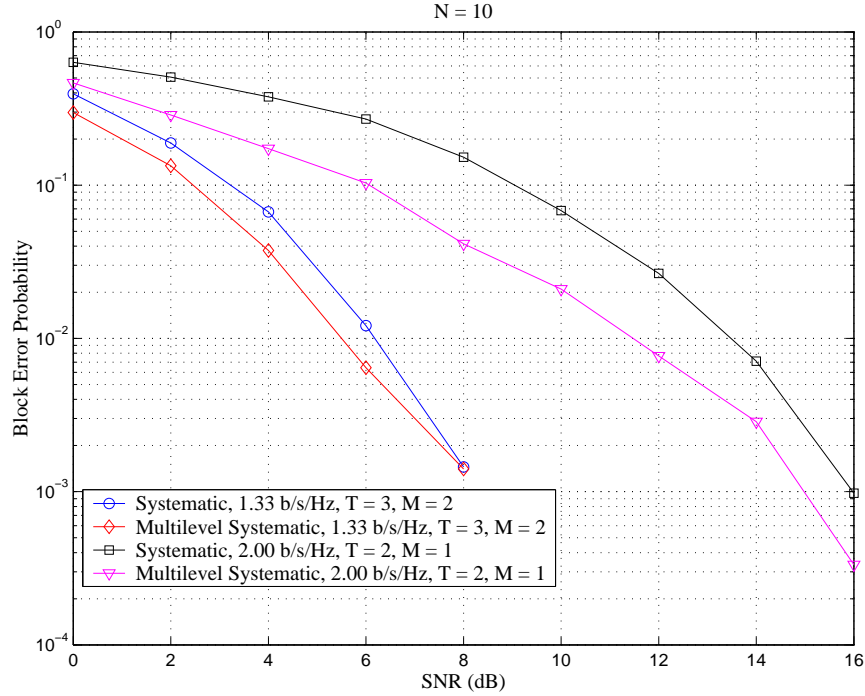


Fig. 12. Performance comparison of one and two transmit antenna systematic constellations of [7] and their multilevel versions vs. SNR.

Using (5) and (8), and assuming that $|s_1| < |s_2|$, we have

$$\begin{aligned} \Pr(s_1 \rightarrow s_2) &= \Pr_{p_1^N} \{X^N : p_2^N(X^N) > p_1^N(X^N)\} \\ &= \Pr_{p_1^N} \left\{ X^N : \frac{\exp\left(\frac{-\|X^N\|^2}{1+|s_2|^2}\right)}{\pi^N(1+|s_2|^2)^N} > \frac{\exp\left(\frac{-\|X^N\|^2}{1+|s_1|^2}\right)}{\pi^N(1+|s_1|^2)^N} \right\} \\ &= \Pr_{p_1^N} \{X^N : \|X^N\|^2 > NA\}, \end{aligned} \quad (56)$$

where A is as in (10).

Similarly, for $|s_1| > |s_2|$ we have

$$\begin{aligned} \Pr(s_1 \rightarrow s_2) &= \Pr_{p_1^N} \{X^N : \|X^N\|^2 < NA\} \\ &= 1 - \Pr_{p_1^N} \{X^N : \|X^N\|^2 > NA\} \end{aligned} \quad (57)$$

(since p_1^N does not have any mass accumulation point).

Equation (12) then follows by applying the following lemma with $C = NA$ and using (10) and (11).

Lemma 4: For any $C \geq 0$, we have

$$\Pr_{p_1^N} \{X^N : \|X^N\|^2 > C\} = \sum_{n=0}^{N-1} \frac{1}{n!} \left(\frac{C}{1+|s_1|^2} \right)^n \exp\left(\frac{-C}{1+|s_1|^2}\right). \quad (58)$$

Proof: The proof is by induction, as follows.

For $N = 1$, we have

$$\Pr_{p_1} \{\|X_1\|^2 > C\} = \exp\left(\frac{-C}{1+|s_1|^2}\right), \quad (59)$$

which is true, and proven in Proposition 1.

Now assume that (58) is true for $N = K$. We prove that it will also be true for $N = K + 1$. Using (8) and the notation defined in (55), we can write

$$p_i^{K+1}(X^{K+1}) = p_i^K(X^K)p_i(x_{K+1}). \quad (60)$$

Defining the regions \mathcal{R} , \mathcal{R}_1 , and \mathcal{R}_2 as

$$\begin{aligned} \mathcal{R} &= \{X^{K+1} : \|X^{K+1}\|^2 > C\}, \\ \mathcal{R}_1 &= \{X^{K+1} : \|X^K\|^2 > C\}, \text{ and} \\ \mathcal{R}_2 &= \{X^{K+1} : \|X^K\|^2 \leq C \ \& \ |x_{K+1}|^2 > C - \|X^K\|^2\}, \end{aligned}$$

we have

$$\begin{aligned} \mathcal{R}_1 \cup \mathcal{R}_2 &= \mathcal{R}, \\ \mathcal{R}_1 \cap \mathcal{R}_2 &= \phi, \end{aligned}$$

and

$$\begin{aligned} \Pr_{p_1^{K+1}} \{X^{K+1} : \|X^{K+1}\|^2 < C\} &= \Pr_{p_1^{K+1}} \{\mathcal{R}\} \\ &= \Pr_{p_1^{K+1}} \{\mathcal{R}_1\} + \Pr_{p_1^{K+1}} \{\mathcal{R}_2\}. \end{aligned} \quad (61)$$

The first term in (61) can be calculated as

$$\begin{aligned}
\Pr_{p_1^{K+1}}\{\mathcal{R}_1\} &= \int_{\mathcal{R}_1} p_1^{K+1}(X^{K+1}) dX^{K+1} \\
&= \int_{\|X^K\|^2 > C} p_1^K(X^K) \times \\
&\quad \left[\int_{\mathcal{C}} p_1(x_{K+1}) dx_{K+1} \right] dX^K \\
&= \Pr_{p_1^K}\{X^K : \|X^K\|^2 > C\} \\
&= \left[\sum_{n=0}^{K-1} \frac{1}{n!} \left(\frac{C}{1+|s_1|^2} \right)^n \right] \times \\
&\quad \exp\left(\frac{-C}{1+|s_1|^2} \right), \quad (62)
\end{aligned}$$

where the last equality follows from the fact that we have assumed (58) is true for $N = K$.

The second term in (61) can be calculated as

$$\begin{aligned}
\Pr_{p_1^{K+1}}\{\mathcal{R}_2\} &= \int_{\mathcal{R}_2} p_1^{K+1}(X^{K+1}) dX^{K+1} \\
&= \int_{\|X^K\|^2 \leq C} p_1^K(X^K) \times \\
&\quad \left[\int_{|x_{K+1}|^2 > C - \|X^K\|^2} p_1(x_{K+1}) dx_{K+1} \right] dX^K \\
&= \frac{\exp\left(\frac{-C}{1+|s_1|^2} \right)}{\pi^K (1+|s_1|^2)^K} \int_{\|X^K\|^2 \leq C} dX^K \\
&= \frac{1}{K!} \left(\frac{C}{1+|s_1|^2} \right)^K \exp\left(\frac{-C}{1+|s_1|^2} \right) \quad (63)
\end{aligned}$$

where the third equality follows from (59) and (8), and the last equality follows from the formula of the volume of a $2K$ -dimensional sphere with radius R ,

$$V_{2K}(R) = \frac{\pi^K}{K!} R^{2K}. \quad (64)$$

Substituting (62) and (63) in (61) shows that (58) is true for $N = K + 1$. This completes the proof. ■

APPENDIX B

CHERNOFF BOUND FOR THE SINGLE ANTENNA CASE

It can be easily shown that

$$\mathcal{C}(p_1^N, p_2^N) = N\mathcal{C}(p_1, p_2). \quad (65)$$

Therefore, in the following we only derive the expression for $\mathcal{C}(p_1, p_2)$. By definition, the Chernoff information (distance) between two probability densities p_1 and p_2 is given by

$$\mathcal{C}(p_1, p_2) = - \min_{0 \leq \lambda \leq 1} \ln \left[\mathbb{E}_{p_1} \left\{ \left(\frac{p_2(x)}{p_1(x)} \right)^\lambda \right\} \right]. \quad (66)$$

Using (8) for the conditional probability densities, we will have

$$\begin{aligned}
\mathcal{C}(p_1, p_2) &= - \min_{0 \leq \lambda \leq 1} \ln \left[\mathbb{E}_{p_1} \left\{ \left(\frac{1+|s_1|^2}{1+|s_2|^2} \right)^\lambda \times \right. \right. \\
&\quad \left. \left. \exp \left(\frac{\lambda|x|^2}{1+|s_1|^2} - \frac{\lambda|x|^2}{1+|s_2|^2} \right) \right\} \right] \\
&= - \min_{0 \leq \lambda \leq 1} \ln \left[\left(\frac{1+|s_1|^2}{1+|s_2|^2} \right)^\lambda \times \right. \\
&\quad \left. \mathbb{E}_{p_1} \left\{ \exp(a|x|^2) \right\} \right], \quad (67)
\end{aligned}$$

where

$$a = \frac{\lambda}{1+|s_1|^2} - \frac{\lambda}{1+|s_2|^2}.$$

Using (8) again with $s = s_1$ for p_1 , we have

$$\begin{aligned}
\mathbb{E}_{p_1} \left\{ \exp(a|x|^2) \right\} &= \int_{\mathcal{C}} \frac{\exp \left\{ -\frac{|x|^2}{1+|s_1|^2} + a|x|^2 \right\}}{\pi(1+|s_1|^2)} dx \\
&= \frac{1}{1-a(1+|s_1|^2)} \\
&= \left[1 - \lambda + \lambda \left(\frac{1+|s_1|^2}{1+|s_2|^2} \right) \right]^{-1} \quad (68)
\end{aligned}$$

Substituting (68) in (67) and using (11), we have

$$\mathcal{C}(p_1, p_2) = - \min_{0 \leq \lambda \leq 1} \{ \lambda \ln(B) - \ln(1 - \lambda + \lambda B) \}. \quad (69)$$

Now, since $\lambda \ln(B) - \ln(1 - \lambda + \lambda B)$ is a strictly convex function of λ for $B \neq 1$, we can find the minimum by taking derivative with respect to λ and setting it to zero, which results in

$$\lambda = \frac{1}{\ln(B)} - \frac{1}{B-1} \quad \text{for } B \neq 1. \quad (70)$$

Substituting this value of λ as the minimizer in (69) together with (65) results in (16).

APPENDIX C

THE KL DISTANCE

In this appendix, we derive the expression for the KL distance between two distributions of form (3). By definition,

$$\mathcal{D}(p_i^N \| p_j^N) = \mathbb{E}_{p_i^N} \left\{ \ln \left[\frac{p_i^N(X)}{p_j^N(X)} \right] \right\}.$$

Using (3) and defining $p_l(X_n) = p(X_n | S_l)$ for $l = 1, \dots, L$, we have

$$\begin{aligned}
\mathcal{D}(p_i^N \| p_j^N) &= \mathbb{E}_{p_i^N} \left\{ \ln \left[\frac{\prod_{n=1}^N p_i(X_n)}{\prod_{n=1}^N p_j(X_n)} \right] \right\} \\
&= \sum_{n=1}^N \mathbb{E}_{p_i^N} \left\{ \ln \left[\frac{p_i(X_n)}{p_j(X_n)} \right] \right\} \\
&= \sum_{n=1}^N \mathbb{E}_{p_i} \left\{ \ln \left[\frac{p_i(X_n)}{p_j(X_n)} \right] \right\} \\
&= N \mathbb{E}_{p_i} \left\{ \ln \left[\frac{p_i(X_n)}{p_j(X_n)} \right] \right\} \\
&= N \mathcal{D}(p_i \| p_j), \quad (71)
\end{aligned}$$

since X_n 's are independent and identically distributed.

Substituting (2) for p_i and p_j , we will have

$$\mathcal{D}(p_i||p_j) = \ln \left[\frac{\det(I_T + S_j S_j^H)}{\det(I_T + S_i S_i^H)} \right] - \mathbb{E}_{p_i} \{X_n^H K X_n\}, \quad (72)$$

where

$$K = (I_T + S_i S_i^H)^{-1} - (I_T + S_j S_j^H)^{-1}.$$

Again, using (2) for p_i , we have

$$\begin{aligned} \mathbb{E}_{p_i} \{X_n^H K X_n\} &= \text{tr} \{K (I_T + S_i S_i^H)\} \\ &= \text{tr} \left\{ I_T - (I_T + S_j S_j^H)^{-1} (I_T + S_i S_i^H) \right\} \\ &= T - \text{tr} \left\{ (I_T + S_i S_i^H) (I_T + S_j S_j^H)^{-1} \right\}. \end{aligned} \quad (73)$$

Substituting (73) in (72), we will have

$$\begin{aligned} \mathcal{D}(p_i||p_j) &= \text{tr} \left\{ (I_T + S_i S_i^H) (I_T + S_j S_j^H)^{-1} \right\} - \\ &\quad \ln \det \left\{ (I_T + S_i S_i^H) (I_T + S_j S_j^H)^{-1} \right\} - T. \end{aligned} \quad (74)$$

Equations (74) and (71) result in (31).

APPENDIX D THE SIMPLIFIED KL DISTANCE FOR ORTHOGONAL MATRICES

In this appendix we derive a simplified version of the KL distance in (31) for a constellation of orthogonal matrices, $\{S_l\}_{l=1}^L$ with $S_l^H S_l = D_l$ for $l = 1, \dots, L$. Here D_l is a diagonal matrix with its m th diagonal element, d_{lm} , equal to the magnitude square of the m th column of S_l , $\|S_{lm}\|^2$. Using the matrix inversion lemma [14]

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}, \quad (75)$$

we can write

$$(I_T + S_l S_l^H)^{-1} = I_T - S_l (I_M + D_l)^{-1} S_l^H. \quad (76)$$

Therefore, we will have

$$\begin{aligned} (I_T + S_i S_i^H) (I_T + S_j S_j^H)^{-1} &= I_T + S_i S_i^H - \\ &\quad S_j (I_M + D_j)^{-1} S_j^H - S_i S_i^H S_j (I_M + D_j)^{-1} S_j^H. \end{aligned} \quad (77)$$

We need to calculate the trace and determinant of this matrix, and substitute for them in (31). To find the trace of (77), we calculate the trace of each term separately. We have

$$\begin{aligned} \text{tr} \left\{ S_j (I_M + D_j)^{-1} S_j^H \right\} &= \text{tr} \left\{ S_j^H S_j (I_M + D_j)^{-1} \right\} \\ &= \text{tr} \left\{ D_j (I_M + D_j)^{-1} \right\} \\ &= \sum_{m=1}^M \frac{d_{jm}}{1 + d_{jm}} \\ &= \sum_{m=1}^M \frac{\|S_{jm}\|^2}{1 + \|S_{jm}\|^2}, \end{aligned} \quad (78)$$

and

$$\begin{aligned} \text{tr} \{S_i S_i^H\} &= \text{tr} \{S_i^H S_i\} = \text{tr} \{D_i\} \\ &= \sum_{m=1}^M d_{im} = \sum_{m=1}^M \|S_{im}\|^2. \end{aligned} \quad (79)$$

To find the trace of the last term in (77), we use the following identity which can be easily verified

$$\text{tr} \{AD\} = \text{tr} \{\text{diag}(A)D\}, \quad (80)$$

where A is an arbitrary square matrix, D is a diagonal matrix of the same size as A , and $\text{diag}(A)$ denotes a diagonal matrix constructed from the diagonal elements of A in the same order. Defining $A = S_j^H S_i S_i^H S_j$, we have

$$a_{mm} = \sum_{k=1}^M S_{jm}^H S_{ik} S_{ik}^H S_{jm} = \sum_{k=1}^M |S_{ik} \cdot S_{jm}|^2, \quad (81)$$

where a_{mm} is the m th diagonal element of A , and \cdot is the inner product operation. Using (80) and (81), we will have

$$\begin{aligned} \text{tr} \left\{ S_i S_i^H S_j (I_M + D_j)^{-1} S_j^H \right\} &= \text{tr} \left\{ S_j^H S_i S_i^H S_j (I_M + D_j)^{-1} \right\} \\ &= \text{tr} \left\{ A (I_M + D_j)^{-1} \right\} \\ &= \text{tr} \left\{ \text{diag}(A) (I_M + D_j)^{-1} \right\} \\ &= \sum_{m=1}^M \frac{\sum_{k=1}^M |S_{ik} \cdot S_{jm}|^2}{1 + \|S_{jm}\|^2}. \end{aligned} \quad (82)$$

Equations (77), (78), (79), and (82) result in

$$\begin{aligned} \text{tr} \left\{ (I_T + S_i S_i^H) (I_T + S_j S_j^H)^{-1} \right\} &= T + \sum_{m=1}^M \|S_{im}\|^2 - \\ &\quad \sum_{m=1}^M \frac{\|S_{jm}\|^2}{1 + \|S_{jm}\|^2} - \sum_{m=1}^M \frac{\sum_{k=1}^M |S_{ik} \cdot S_{jm}|^2}{1 + \|S_{jm}\|^2}. \end{aligned} \quad (83)$$

Now we calculate the determinant of $(I_T + S_i S_i^H) (I_T + S_j S_j^H)^{-1}$. For this, we use the identity [14]

$$\det(I + AB) = \det(I + BA), \quad (84)$$

to write

$$\begin{aligned} \det(I_T + S_l S_l^H) &= \det(I_M + S_l^H S_l) \\ &= \det(I_M + D_l) \\ &= \prod_{m=1}^M (1 + d_{lm}) \\ &= \prod_{m=1}^M (1 + \|S_{lm}\|^2). \end{aligned} \quad (85)$$

Using (85), we will have

$$\begin{aligned} \det \left\{ (I_T + S_i S_i^H) (I_T + S_j S_j^H)^{-1} \right\} &= \\ &= \frac{\prod_{m=1}^M (1 + \|S_{im}\|^2)}{\prod_{m=1}^M (1 + \|S_{jm}\|^2)}. \end{aligned} \quad (86)$$

Substituting (83) and (86) in (31), results in (34).

APPENDIX E
THE OPTIMAL SINGLE ANTENNA CONSTELLATION FOR
FAST FADING

Without loss of generality, let's assume that $0 \leq |s_1| \leq |s_2| \leq \dots \leq |s_L|$. Since $f(x) = x - \ln(x) - 1$ is monotonically decreasing for $x \in (0, 1)$ and monotonically increasing for $x \in (1, \infty)$, with $f(x) < f(\frac{1}{x})$ for $x \in (0, 1)$, it is clear that the minimum KL distance will occur between a pair of consecutive symbols from the above order, in the same order. Moreover, in order to solve (32), it is sufficient to solve the following minimax problem

$$\underset{\frac{1}{L} \sum_{l=1}^L |s_l|^2 \leq P}{\text{minimize}} \quad \max_{l=1, \dots, L-1} \quad \frac{1 + |s_l|^2}{1 + |s_{l+1}|^2}. \quad (87)$$

Defining

$$\alpha = \min_{l=1, \dots, L-1} \frac{1 + |s_{l+1}|^2}{1 + |s_l|^2},$$

we will have,

$$1 + |s_{l+1}|^2 \geq \alpha(1 + |s_l|^2) \Rightarrow 1 + |s_l|^2 \geq \alpha^{l-1}(1 + |s_1|^2) \quad (88)$$

for $l = 1, \dots, L$, or

$$L + \sum_{l=1}^L |s_l|^2 \geq \left(\sum_{l=1}^L \alpha^{l-1} \right) (1 + |s_1|^2),$$

which implies

$$\frac{1 - \alpha^L}{1 - \alpha} \leq \frac{L + \sum_{l=1}^L |s_l|^2}{1 + |s_1|^2} \leq \frac{L + LP}{1 + |s_1|^2},$$

(using the average power constraint). Now, since $\frac{1 - \alpha^L}{1 - \alpha}$ is a monotonically increasing function of α , it is clear that the maximum of α is obtained if and only if $s_1 = 0$, and $\frac{1 - \alpha^L}{1 - \alpha} = L(1 + P)$. This requires that all of the inequalities in (88) hold with equality. Therefore, the optimum signal set can be obtained by setting

$$|s_l|^2 = \alpha^{l-1} - 1,$$

where α is the largest real number satisfying $\frac{1 - \alpha^L}{1 - \alpha} = L(1 + P)$, or

$$\alpha^L - L(P + 1)\alpha + (LP + L - 1) = 0.$$

APPENDIX F
EQUIVALENCE OF THE EUCLIDEAN AND CHORDAL
DISTANCES BETWEEN SUBSPACES

The *chordal distance* between two M -dimensional subspaces, W_i and W_j , of \mathbb{C}^T is defined [13] as

$$d_c^2(W_i, W_j) = \sum_{m=0}^M \sin^2(\angle S_{im}, S_{jm}), \quad (89)$$

where $\{S_{im}\}_{m=1}^M$ and $\{S_{jm}\}_{m=1}^M$ are the *principal vectors* corresponding to W_i and W_j , respectively, and are recursively defined as

$$(S_{im}, S_{jm}) = \arg \max_{\substack{(u,v) \in W_i \times W_j, \|u\| = \|v\| = 1 \\ u \cdot S_{ik} = v \cdot S_{jk} = 0 \text{ for } k < m}} u \cdot v \quad (90)$$

for $m = 1, \dots, M$. We will use the following lemma to prove the equivalence of the *Euclidean* and *chordal* distances.

Lemma 5: If W_i and W_j are two M -dimensional subspaces of \mathbb{C}^T , and $\{S_{im}\}_{m=1}^M$ and $\{S_{jm}\}_{m=1}^M$ are the *principal vectors* corresponding to W_i and W_j , respectively, then

- a) $\{S_{im}\}_{m=1}^M$ and $\{S_{jm}\}_{m=1}^M$ form orthonormal bases for W_i and W_j , respectively, and
- b) $S_{im} \cdot S_{jk} = 0$ for $m \neq k$.

Proof:

- a) By definition, each principal vector has unit norm, and we have $S_{im} \cdot S_{ik} = 0$ for $k < m$. By exchanging the role of m and k , we also have $S_{ik} \cdot S_{im} = 0$ for $m < k$. Therefore, we have $S_{im} \cdot S_{ik} = 0$ for $m \neq k$.
- b) For any given m , let's define $W_j^m = \text{span}(S_{jm}, \dots, S_{jM})$. By definition,

$$\text{Proj}_{W_j^m}(S_{im}) = (S_{im} \cdot S_{jm})S_{jm}, \quad (91)$$

where $\text{Proj}_{W_j^m}(S_{im})$ is the projection of S_{im} on W_j^m . Therefore, we have

$$(S_{im} - (S_{im} \cdot S_{jm})S_{jm}) \cdot S_{jk} = 0 \quad (92)$$

for $k \geq m$, which implies

$$S_{im} \cdot S_{jk} = (S_{im} \cdot S_{jm})(S_{jm} \cdot S_{jk}) = 0 \quad (93)$$

for $k > m$. Similarly,

$$\text{Proj}_{W_i^k}(S_{jk}) = (S_{jk} \cdot S_{ik})S_{ik}, \quad (94)$$

where $W_i^k = \text{span}(S_{ik}, \dots, S_{iM})$. Therefore, we have

$$(S_{jk} - (S_{jk} \cdot S_{ik})S_{ik}) \cdot S_{im} = 0 \quad (95)$$

for $m \geq k$, which implies

$$S_{jk} \cdot S_{im} = (S_{jk} \cdot S_{ik})(S_{ik} \cdot S_{im}) = 0 \quad (96)$$

for $m > k$. Therefore, we have $S_{im} \cdot S_{jk} = 0$ for $m \neq k$. ■

Now, using $\{S_{im}\}_{m=1}^M$ and $\{S_{jm}\}_{m=1}^M$ as bases for W_i and W_j , by definition of the *Euclidean distance* between subspaces, we have

$$\begin{aligned} d_E^2(W_i, W_j) &= \sum_{m=1}^M d_E^2(S_{im}, W_j) \\ &= \sum_{m=1}^M \left\{ \|S_{im}\|^2 - \sum_{k=1}^M \frac{|S_{im} \cdot S_{jk}|^2}{\|S_{jk}\|^2} \right\} \\ &= \sum_{m=1}^M \{1 - |S_{im} \cdot S_{jm}|^2\} \\ &= \sum_{m=1}^M \{1 - \cos^2(\angle S_{im}, S_{jm})\} \\ &= \sum_{m=1}^M \sin^2(\angle S_{im}, S_{jm}) \\ &= d_c^2(W_i, W_j). \end{aligned} \quad (97)$$

REFERENCES

- [1] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
 - [2] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *AT&T-Bell Laboratories Internal Tech. Memo.*, June 1995.
 - [3] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.
 - [4] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, March 1998.
 - [5] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.
 - [6] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 543–564, March 2000.
 - [7] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 1962–1973, Sept. 2000.
 - [8] V. Tarokh, "On the design of the first unitary space-time codes that have simple encoding/decoding algorithm," in *Proceedings of the 39th Annual Allerton Conference on Communications, Control and Computing*, Monticello, IL, Oct. 2001.
 - [9] L. Zheng and D. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.
 - [10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley Interscience, 1991.
 - [11] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.
 - [12] S. Galliou, I. Kammoun, and J. Belfiore, "Space-time codes for the GLRT noncoherent detector," in *Proceedings of the IEEE International Symposium on Information Theory*, 2002, pp. 77–77.
 - [13] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc: packings in Grassmannian space," *Experimental Math*, vol. 5, pp. 139–159, 1996.
 - [14] S. M. Kay, *Fundamentals of Signal Processing: Estimation Theory*. Prentice Hall International Editions, 1993.
- Mohammad Jaber Borran** received the B.S. degree in Electrical and Electronics Engineering and the M.S. degree in Electrical and Communications Engineering from Sharif University of Technology, Tehran, Iran, in 1993 and 1996, respectively, and the Ph.D. degree in Electrical Engineering from Rice University, Houston, Texas, in 2003. He won the silver medal in the International Mathematics Olympiad, 1989, Germany, and was a recipient of the Nokia Mobile Phones R&D Intern Scholarship. He is currently a Senior Research Engineer at Nokia Mobile Phones, Irving, Texas. His research interests include wireless communications and information theory.
- Ashutosh Sabharwal** received his B.Tech. degree in Electrical Engineering from Indian Institute of Technology, New Delhi in 1993. He received his M.S. and Ph.D. degrees in Electrical Engineering in 1995 and 1999, respectively, from the Ohio State University, Columbus, Ohio. Since 2001, he has been a Faculty Fellow at Rice University, where he had joined as a Postdoctoral Research Associate in 1999. His current research interests include wireless networking and information theory. He was the recipient of the 1999 Presidential Dissertation Fellowship.
- Behnaam Aazhang** received his B.S. (with highest honors), M.S., and Ph.D. degrees in Electrical and Computer Engineering from University of Illinois at Urbana-Champaign in 1981, 1983, and 1986, respectively. From 1981 to 1985, he was a Research Assistant in the Coordinated Science Laboratory, University of Illinois. In August 1985, he joined the faculty of Rice University, Houston, Texas, where he is now the J.S. Abercrombie Professor in the Department of Electrical and Computer Engineering and the Director of Center for Multimedia Communications. He has been a Visiting Professor at IBM Federal Systems Company, Houston, Texas, the Laboratory for Communication Technology at Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, the Telecommunications Laboratory at University of Oulu, Oulu, Finland, U.S. Air Force Phillips Laboratory, Albuquerque, New Mexico, Helsinki University of Technology, Finland, and Nokia Mobile Phone, Irving, Texas. His research interests are in the areas of communication theory, information theory, and their applications with emphasis on multiple access communications, cellular mobile radio communications, and optical communication networks.
- Dr. Aazhang is a Fellow of IEEE, a recipient of the Alcoa Foundation Award 1993, the NSF Engineering Initiation Award 1987-1989, and the IBM Graduate Fellowship 1984-1985, and is a member of Tau Beta Pi and Eta Kappa Nu. He is currently serving on Houston Mayor's Commission on Cellular Towers. He has served as the Editor for Spread Spectrum Networks of IEEE Transactions on Communications 1993-1998, as the Treasurer of IEEE Information Theory Society 1995-1998, the Technical Area Chair of 1997 Asilomar Conference, Monterey, California, the Secretary of the Information Theory Society 1990-1993, the Publications Chairman of the 1993 IEEE International Symposium on Information Theory, San Antonio, Texas, and as the co-chair of the Technical Program Committee of 2001 Multi-Dimensional and Mobile Communication (MDMC) Conference in Pori, Finland.