

Design Criteria and Construction of Multiple-Antenna Partially Coherent Constellations and Coded Modulation

Mohammad Jaber Borran, Ashutosh Sabharwal, and Behnaam Aazhang

Department of Electrical and Computer Engineering, Rice University

6100 Main Street, MS-366, Houston, TX 77005

Email: {mohammad,ashu,aaz}@rice.edu

Abstract

We consider multiple-antenna communication systems in Rayleigh fading channel, where the transmitter does not know the channel coefficients and the receiver has only an estimate of them. We further assume that the transmitter and receiver know the statistics of the estimation error. We refer to this system as *partially coherent* system, for which we derive the expressions for the optimal detector and study the code and constellation design problems. Finding the Chernoff bound intractable, and inspired by Stein's Lemma, we propose to use the Kullback-Leibler (KL) distance between conditional distributions to design space-time codes and constellations for partially coherent systems. We show that the KL distance is relatively easy to derive and work with, and furthermore, provides an efficient design criterion.

Using the KL-based design criterion, we construct constellations for multiple-antenna systems which can be decoded in the presence of channel estimation errors, and thus are suitable for fading scenarios with short coherence intervals. The proposed constellations are multi-level, with multi-dimensional spherical constellations at each level. We also propose a recursive construction for the constituent spherical subsets of the multiple-antenna partially coherent constellations. The new partially coherent constellations provide significant performance improvement over the conventional single-antenna PSK and QAM constellations and multiple-antenna techniques such as Bell Lab's Space-Time (BLAST) architecture and Orthogonal Transmit Diversity (OTD) schemes, when the estimation variance is comparable to the reciprocal of the signal-to-noise ratio. More specifically, we show that by using the proposed

constellations, the error floors due to the estimation errors can be reduced by as much as one order of magnitude.

Next, we bring coding into the picture and derive a KL-based design criterion for partially coherent coded modulation. We also propose a construction method for partially coherent coded modulation using the idea of mapping by set partitioning. We show that, in the presence of channel estimation errors, the proposed codes provide substantial performance improvement over the conventional coded modulation techniques, and the gains in this case are even larger than the gains obtained in the case of uncoded systems.

Index Terms

Partially coherent constellations, imperfect channel state information, space-time codes, multiple-antenna systems, fading channels, channel coding, wireless communications.

I. INTRODUCTION

Wireless communication systems can be categorized into three categories, in terms of the amount of channel state information available at the receiver:

- 1) Systems with complete channel state information at the receiver, which are usually referred to as *coherent systems*,
- 2) Systems with no channel state information at the receiver, which are usually referred to as *non-coherent systems*, and
- 3) Systems with estimated (and hence, erroneous) channel state information at the receiver, which in this paper are referred to as *partially coherent systems*.

The perfect channel state information assumption in the first category is not very realistic in many practical cases. However, it usually simplifies the analysis and also provides a good approximation when channel variations are very slow. In such cases, relatively long training sequences can be transmitted and very reliable channel estimates can be obtained at the receiver. In [1], it has been shown that when the length of the coherence interval of the channel tends to infinity, the capacity of the non-coherent system approaches the capacity of the coherent system, meaning that perfect channel training is possible without much loss in the actual data rate.

The system model in the second category is the most precise model for any wireless communication system, in the sense that it formulates the problems of channel estimation and data

detection in a joint problem. The capacity of the non-coherent systems has been studied in [1]–[3]. Constellation design for multiple-antenna non-coherent systems has been studied in [4]–[7]. Even though this model is the most precise model, the exponential growth of the constellation size with the length of the coherence interval makes the joint design problem very difficult. Decoupling the channel estimation and data detection problems simplifies the constellation design. However, to achieve the best performance, accurate models for both channel training and data transmission phases are required. A finite training length results in a non-zero estimation variance, and an accurate system model for the data transmission phase needs to capture this estimation error.

The capacity of the fading channels with imperfect channel state information at the receiver has been studied in [8]–[12]. In [10], for a given channel measurement variance, upper and lower bounds have been derived for the mutual information between the transmitted and received signals. These two bounds correspond to the two extreme cases when the estimation error is either completely constructive or completely destructive. In the lower bound, the variance of the additive white Gaussian noise is increased by an amount equal to the product of the estimation variance and transmission power, whereas in the upper bound, the transmission power is increased by the same amount. According to these bounds, the loss in the capacity due to the estimation errors is negligible at low Signal to Noise Ratio (SNR) (the upper and lower bounds overlap with the capacity curve with perfect channel state information). For a fixed estimation variance, as SNR increases, the two bounds start diverging, and do not provide much insight into the actual value of the channel capacity. In [11], as a rule of thumb, it has been suggested that in order to avoid degradation, the estimation error should be negligible compared to the reciprocal of the SNR.

In this paper, we consider multiple-antenna partially coherent communication systems in flat Rayleigh fading channels. First, we define the system model and derive the expressions for the optimal partially coherent detectors. Next, we formulate the constellation design problem, and using a similar approach to [7] and based on the KL distance between conditional received distributions, we derive a design criterion for partially coherent space-time constellations. We show that, for the two extreme values of the estimation variance, the proposed criterion reduces to the existing criteria for the coherent [13] and non-coherent [7] space-time constellations. We

simplify this criterion for the case of vector constellations, in which each constellation point involves only one channel use.

Among other contributions of this paper is a suboptimal, yet efficient and systematic, method to design multiple-antenna partially coherent constellations. The proposed constellations have a multi-level structure, with multi-dimensional spherical constellations at each level. We also propose a recursive construction method for the spherical constellations, in which each constellation is constructed from several lower dimensional constellations. Our multi-level spherical structure and recursive construction enable us to design relatively large constellations which outperform the conventional constellations and existing multiple-antenna techniques, and provide substantial performance gains in the presence of channel estimation errors.

Another contribution of this paper is a design criterion and construction method for partially coherent coded modulation. Here, we use the additive property of the KL distance to derive a design criterion similar to the design criterion for coded modulation in the Additive White Gaussian (AWGN) Channel. Then we use the idea of *mapping by set partitioning* [14], [15], but with the KL distance instead of the Euclidean distance, to design partially coherent trellis coded modulation. Similar techniques have been developed in [16], [17] for coherent coded modulation in fast fading scenarios. We design single- and multiple-antenna partially coherent coded modulations and evaluate their performance. Our numerical results demonstrate substantial performance gains over conventional coded modulations and trellis coded transmit diversity schemes.

In Section II, we introduce the model for the system being considered throughout this paper. In Section III, we derive the KL distance and an *expected KL distance* between distributions corresponding to the transmitted signals, and propose the design criterion based on that. In Section IV, first through a single-antenna example, we demonstrate the effect of the channel estimation error in the shape and structure of the optimal constellations. We also present some numerical results that show significant improvement in the performance by using the new constellations instead of the conventional PSK and QAM constellations. Next, we study the constellation design problem for the case of multiple transmit antennas, and simplify the design criterion for the important case of vector constellations. We also present some numerical results

that show substantial improvement in the performance by using the new constellations instead of the conventional multiple-antenna techniques such as the Bell Lab's Space-Time (BLAST) architecture and Orthogonal Transmit Diversity (OTD). In Section V, we derive a design criterion and construction method for partially coherent coded modulation in block fading channels. We also present numerical results for single- and multiple-antenna systems which show substantial gains over the conventional trellis coded modulation techniques. Finally, we bring the concluding remarks in Section VI.

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 to 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 with power constraint $\sum_{t=1}^T \sum_{m=1}^M \mathbb{E} \{|s_{tm}|^2\} = TP$, where s_{tm} 's are the elements of the signal matrix S , 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 also assume that $H = \hat{H} + \tilde{H}$, where \hat{H} is known to the receiver but \tilde{H} is not. Furthermore, we assume that \tilde{H} has i.i.d. elements from $\mathcal{CN}(0, \sigma_E^2)$, and is statistically independent from \hat{H} (this can be obtained, e.g., by using an LMMSE estimator).

With the above assumptions, the n th column X_n , of the received matrix X , will have a circular complex Gaussian distribution with mean $S\hat{H}_n$ and covariance matrix $I_T + \sigma_E^2 S S^H$,

$$p(X_n | S, \hat{H}_n) = \mathbb{E}_{\tilde{H}_n} \left\{ p(X_n | S, \hat{H}_n, \tilde{H}_n) \right\} = \frac{\exp\{-(X_n - S\hat{H}_n)^H (I_T + \sigma_E^2 S S^H)^{-1} (X_n - S\hat{H}_n)\}}{\pi^T \det(I_T + \sigma_E^2 S S^H)}, \quad (2)$$

where \hat{H}_n is the n th column of the channel estimate matrix \hat{H} , and \tilde{H}_n is the n th column of the estimation error matrix \tilde{H} . Therefore, since the columns of X are statistically independent,

the conditional probability density of the received matrix can be written as

$$p(X|S, \hat{H}) = \prod_{n=1}^N p(X_n|S, \hat{H}_n) = \frac{\exp\{-\text{tr}[(I_T + \sigma_E^2 S S^H)^{-1}(X - S\hat{H})(X - S\hat{H})^H]\}}{\pi^{TN} \det^N(I_T + \sigma_E^2 S S^H)}. \quad (3)$$

Assuming a signal set of size L , $\{S_i\}_{i=1}^L$, and defining $p_l(X) = p(X|S_l, \hat{H})$, the Maximum Likelihood (ML) detector for this system will have the following form

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

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

$$\Pr(S_1 \rightarrow S_2) = \Pr\{p_2(X) > p_1(X)|S_1\}. \quad (5)$$

For $L > 2$, even though (5) is no longer exact, we still use it as an approximation for the pairwise error probability.

III. DESIGN CRITERION

The conditional symbol error probability of the ML detector, given that a specific element of the signal set is transmitted, is obtained by summing the pairwise error probabilities corresponding to that signal point. The average error probability of the ML detector, is then obtained by averaging these conditional error probabilities over the signal set. This quantity is usually dominated by the largest term, i.e., the maximum of (5) over the signal set. Therefore, like most of the other constellation/code design techniques, we use the maximum of (5) over the signal set as the performance criterion, and try to find optimal constellations by minimizing it over all possible constellations of the given size. Unfortunately, the exact expression or even the Chernoff bound for (5) in general seems to be intractable. Therefore, much like our approach in [7], we use Stein's Lemma and propose to use the Kullback-Leibler (KL) distance between distributions (which is an upper bound on the rate of the exponential decay of the pairwise error probability), as the performance criterion. The optimal constellations are then obtained by searching for signal sets which have the largest minimum KL distance.

Proposition 1: The KL distance between two conditional probability densities p_i and p_j of the form (3) is given by

$$\begin{aligned} \mathcal{D}(p_i||p_j) = & N \text{tr} \left\{ (I_T + \sigma_E^2 S_i S_i^H) (I_T + \sigma_E^2 S_j S_j^H)^{-1} \right\} - NT \\ & - N \ln \det \left\{ (I_T + \sigma_E^2 S_i S_i^H) (I_T + \sigma_E^2 S_j S_j^H)^{-1} \right\} \\ & + \text{tr} \left\{ (I_T + \sigma_E^2 S_j S_j^H)^{-1} (S_i - S_j) \widehat{H} \widehat{H}^H (S_i - S_j)^H \right\}. \end{aligned} \quad (6)$$

Proof: See Appendix A. ■

The above KL distance is a function of \widehat{H} , which is not known at the transmitter, and cannot be used as a design criterion for partially coherent constellations. According to Stein's Lemma [18], the KL distance gives the best achievable error exponent using hypothesis test. Therefore, asymptotically in N , the pairwise error probability of mistaking S_j for S_i , of the best hypothesis test designed to maximize the exponential decay rate of this error probability (this is not necessarily the ML detector) will be an exponential function, with $-\mathcal{D}(p_i||p_j)$ in the exponent. To obtain the *expected KL distance* between signal points S_i and S_j , we find the expected value of this exponential function with respect to the distribution of \widehat{H} , as given in the following proposition.

Proposition 2: If the best achievable conditional pairwise error probability of mistaking S_j for S_i is given by

$$\Pr_{best} (S_j \rightarrow S_i | \widehat{H}) \approx \exp(-\mathcal{D}(p_i||p_j)), \quad (7)$$

then the best achievable pairwise error probability of mistaking S_j for S_i will be given by

$$\Pr_{best} (S_j \rightarrow S_i) \approx \exp(-\overline{\mathcal{D}}(p_i||p_j)), \quad (8)$$

where

$$\begin{aligned} \overline{\mathcal{D}}(S_i||S_j) = & N \text{tr} \left\{ (I_T + \sigma_E^2 S_i S_i^H) (I_T + \sigma_E^2 S_j S_j^H)^{-1} \right\} - NT \\ & - N \ln \det \left\{ (I_T + \sigma_E^2 S_i S_i^H) (I_T + \sigma_E^2 S_j S_j^H)^{-1} \right\} \\ & + N \ln \det \left\{ I_M + (1 - \sigma_E^2) (S_i - S_j)^H (I_T + \sigma_E^2 S_j S_j^H)^{-1} (S_i - S_j) \right\}. \end{aligned} \quad (9)$$

Proof: See Appendix B. ■

We emphasize that (9) is obtained by taking the expectation of the error bound using (6) as the exponent, and not by taking the expectation of (6). It is interesting to notice that in the two extreme cases of $\sigma_E^2 = 0$ and $\sigma_E^2 = 1$, (9) reduces to the existing performance criteria for

coherent and non-coherent space-time codes. For $\sigma_E^2 = 0$ (perfect channel state information at the receiver, i.e., coherent communication), (9) reduces to

$$\overline{\mathcal{D}}(S_i||S_j) = N \ln \det \{ I_M + (S_i - S_j)^H (S_i - S_j) \}, \quad (10)$$

which is exactly the same performance criterion given in [13] for coherent space-time codes, and results in the *rank* and *determinant* design criteria. For $\sigma_E^2 = 1$ (no channel state information at the receiver, i.e., non-coherent communication), (9) reduces to

$$\begin{aligned} \overline{\mathcal{D}}(S_i||S_j) = & N \text{tr} \left\{ (I_T + S_i S_i^H) (I_T + S_j S_j^H)^{-1} \right\} - NT \\ & - N \ln \det \left\{ (I_T + S_i S_i^H) (I_T + S_j S_j^H)^{-1} \right\}, \end{aligned} \quad (11)$$

which is exactly the same performance criterion given in [7] for non-coherent space-time constellations. For the intermediate values of σ_E^2 , the performance criterion is a combination of the two extreme values, reflecting the fact that, for an optimal design, contributions from both terms have to be exploited to achieve better performance.

Adopting the KL distance as the performance criterion, the signal set design can be formulated as the following optimization problem

$$\begin{aligned} & \text{maximize} && \min && \overline{\mathcal{D}}(S_i||S_j), \\ & \frac{1}{L} \sum_{l=1}^L \|S_l\|^2 = TP && i \neq j \end{aligned} \quad (12)$$

where $\|S_l\|^2 = \sum_{t=1}^T \sum_{m=1}^M |(S_l)_{tm}|^2$ is the total power used to transmit S_l . Since the actual value of N does not affect the maximization in (12), in designing the optimal signal sets we can always assume $N = 1$.

IV. SIGNAL SET CONSTRUCTION

A. Single Transmit Antenna

In order to demonstrate the new design technique and the effect of channel estimation error in the structure of resulting constellations, we first consider the simple case of a single transmit antenna system in a fast fading environment. In this case, we have $M = 1$ and $T = 1$, so each S_l is simply a complex scalar. The expression for the *expected KL distance* in (9) reduces to

$$\overline{\mathcal{D}}_1(s_i||s_j) = \frac{1 + \sigma_E^2 |s_i|^2}{1 + \sigma_E^2 |s_j|^2} - 1 - \ln \left(\frac{1 + \sigma_E^2 |s_i|^2}{1 + \sigma_E^2 |s_j|^2} \right) + \ln \left[1 + (1 - \sigma_E^2) \frac{|s_i - s_j|^2}{1 + \sigma_E^2 |s_j|^2} \right]. \quad (13)$$

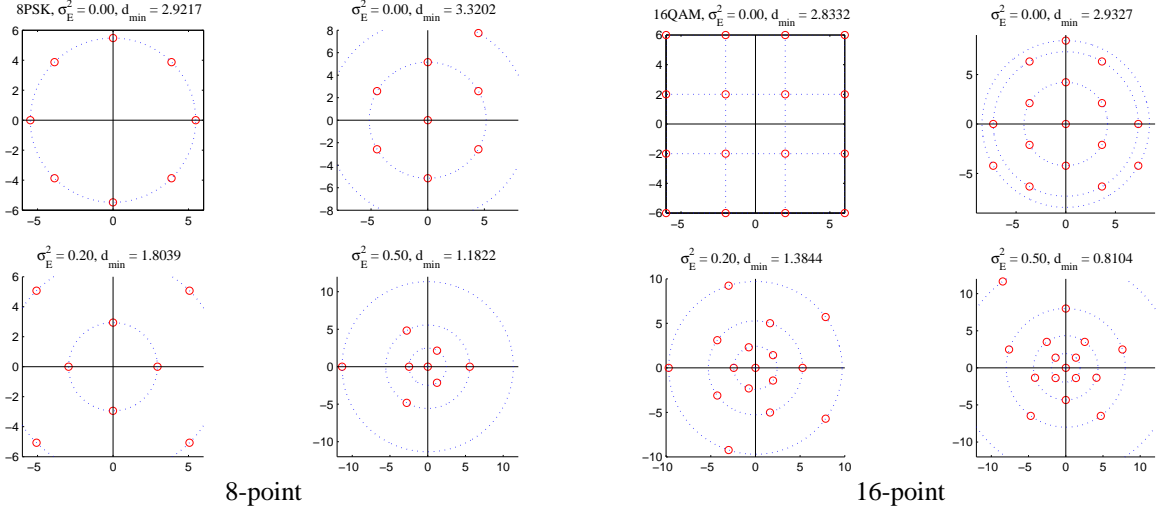


Fig. 1. Optimal constellations of size 8 and 16 for $M = 1$, $T = 1$, average energy per bit of 10, and different values of σ_E^2

Using the same idea of multilevel unitary (circular, in this case) constellations of [7], we consider constellations which consist of points on concentric circles, and solve the optimization problem to find the optimum values for the number of circles, their radii, and the number of constellation points on each circle. This approach is explained in greater detail in the next section, where we consider the more general case of multiple-antenna constellations.

The resulting 8- and 16-point constellations with average energy per bit of 10 and for different values of σ_E^2 are shown in Figure 1. Notice that in design of these constellations, we have allowed for appropriate rotations of the circular subsets to obtain a larger minimum KL distance. As we see in this figure, for larger values of estimation variance, the optimal partially coherent constellations tend to have more layers. The reason is that for larger values of the estimation variance, the Euclidean distance term inside the log function in (13) gets a smaller weight, and the first part of the KL distance which only depends on the magnitudes of the constellation points becomes more dominant. In the extreme case of $\sigma_E^2 = 1$, as mentioned earlier, the logarithmic term in (13) vanishes, and the constellation becomes a PAM-type non-coherent constellation with a number of levels equal to the number of constellation points [7].

The symbol error rate performance of the above constellations at $\sigma_E^2 = 0.5$ are simulated for different values of number of receive antennas (N) and compared with the commonly-used

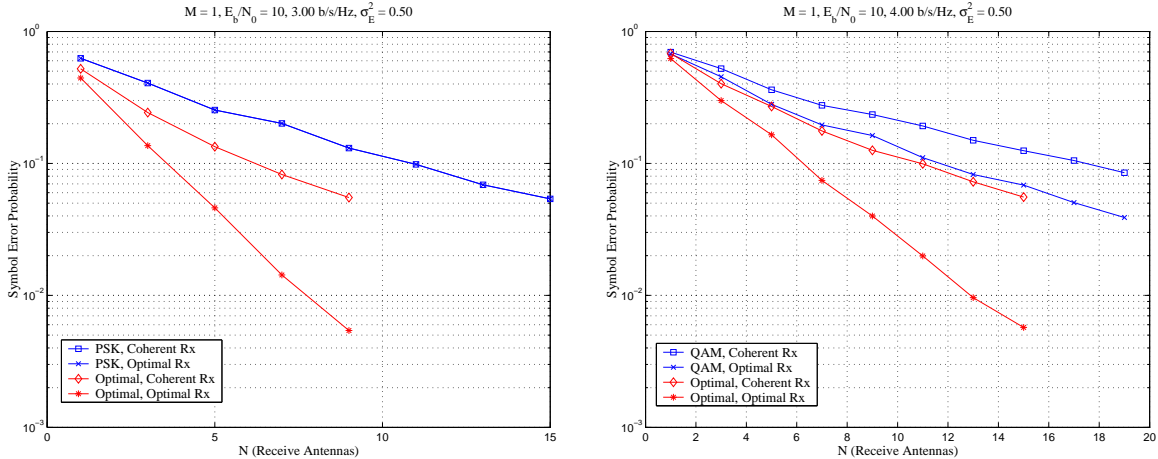


Fig. 2. Symbol error rate for constellations of size 8 and 16 for $M = 1$, $T = 1$, $E_b/N_0 = 10$ dB, and $\sigma_E^2 = 0.5$

8PSK and 16QAM constellations. The results are shown in Figure 2, where by “Coherent Rx” we refer to the coherent detector, i.e. a detector that assumes perfect channel state information, and hence performs the demodulation based on the Euclidean distance. The “Optimal Rx”, on the other hand, uses the optimal (ML) detector of (4) with the true likelihood function that incorporates the estimation variance. As expected, due to the larger minimum KL distance of the new constellations, the exponential decay of the symbol error rate versus N is much higher for the new constellations. We also observe that, even though the optimal receiver improves the performance of the QAM constellation, and the optimal constellation provides performance improvement even with the coherent receiver, in order to obtain the best performance both the optimal constellation and the optimal receiver are required, and neither one alone can provide near optimal performance. Furthermore, the optimal receiver does not improve the performance of the PSK constellation. In fact, it is easy to verify that for a constant energy constellation, the optimal receiver is equivalent to the coherent receiver.

B. Multiple Transmit Antennas

In this section, we design partially coherent constellations for the multiple-antenna systems. For a fixed spectral efficiency, the constellation size grows exponentially with T . For example, to achieve a spectral efficiency of 4 b/s/Hz with $T = 5$, one needs to design a constellation

of $2^{20} = 1,048,576$ points. With multiple antennas at the transmitter and receiver, even larger spectral efficiencies are expected, making the constellation design over multiple symbol intervals even more difficult, and their decoding complexity prohibitively large. For these reasons, in this section we only consider the case of $T = 1$. Notice that the actual coherence interval of the channel can be larger than this value. In fact, in this section, we use the T parameter to only specify the first dimension of the constellation matrices, and not the actual coherence interval of the channel. It is also important to notice that, with $T = 1$, each transmitted matrix will have a unit rank, and thus will not be able to provide any transmit diversity gain. Therefore, if the actual coherence interval of the channel is larger than one, and in the absence of channel estimation errors, any transmit diversity scheme is expected to show a better performance at high SNR. However, as we will see later, in the presence of channel estimation errors, the performance of the proposed constellations can be significantly better than transmit diversity schemes with comparable computational complexity.

Assuming $T = 1$, each S_l will be a complex row vector. The expression for the *expected KL distance* in (9) reduces to

$$\begin{aligned} \bar{\mathcal{D}}_1(S_i||S_j) = & \frac{1+\sigma_E^2\|S_i\|^2}{1+\sigma_E^2\|S_j\|^2} - 1 - \ln\left(\frac{1+\sigma_E^2\|S_i\|^2}{1+\sigma_E^2\|S_j\|^2}\right) \\ & + \ln \det \left\{ I_2 + \frac{1-\sigma_E^2}{1+\sigma_E^2\|S_j\|^2} (S_i - S_j)^H (S_i - S_j) \right\}. \end{aligned} \quad (14)$$

Using the identity

$$\det(I_M + A_{M \times N} B_{N \times M}) = \det(I_N + B_{N \times M} A_{M \times N}), \quad (15)$$

We will have

$$\bar{\mathcal{D}}_1(S_i||S_j) = \frac{1 + \sigma_E^2\|S_i\|^2}{1 + \sigma_E^2\|S_j\|^2} - 1 - \ln\left(\frac{1 + \sigma_E^2\|S_i\|^2}{1 + \sigma_E^2\|S_j\|^2}\right) + \ln\left\{1 + \frac{1 - \sigma_E^2}{1 + \sigma_E^2\|S_j\|^2}\|S_i - S_j\|^2\right\}. \quad (16)$$

From (16), one can see that if two constellation points (vectors) have the same norm (i.e., lie on the same M -dimensional complex sphere centered at the origin), the first three terms will cancel out, and the KL distance will be a monotonic function of the Euclidean distance between them. Therefore, if one considers only constant power constellations (i.e., constellations for which all of the points lie on the same sphere centered at the origin), then the design criterion becomes maximizing the minimum Euclidean distance between constellation points, similar to the case of perfect CSI at the receiver. The design problem in this case reduces to the problem

of packing points on the surface of an M -dimensional complex sphere (or $2M$ -dimensional real sphere), which is a well-studied problem (see, e.g., [19] and references therein).

On the other hand, if two points lie on different spheres, the minimum KL distance between them will happen when they lie on a line which passes through the origin, and will be determined by the radii of the two spheres. This means that if one partitions the constellation into subsets of concentric M -dimensional complex spheres, C_1, \dots, C_K , with radii r_1, \dots, r_K , containing l_1, \dots, l_K points, respectively, and defines the intra-subset and inter-subset distances as

$$\mathcal{D}_{intra}(k) = \min_{S_i, S_j \in C_k} \ln \left\{ 1 + \frac{1 - \sigma_E^2}{1 + \sigma_E^2 r_k^2} \cdot 4r_k^2 \sin^2(\angle \overline{S}_i, \overline{S}_j) \right\}, \quad (17)$$

and

$$\mathcal{D}_{inter}(k, k') = \frac{1 + \sigma_E^2 r_k^2}{1 + \sigma_E^2 r_{k'}^2} - 1 - \ln \left(\frac{1 + \sigma_E^2 r_k^2}{1 + \sigma_E^2 r_{k'}^2} \right) + \ln \left\{ 1 + \frac{1 - \sigma_E^2}{1 + \sigma_E^2 r_{k'}^2} |r_k - r_{k'}|^2 \right\}, \quad (18)$$

then the minimum KL distance between the constellation points will be greater than or equal to the minimum of the inter-subset and intra-subset KL distances. In (17), \overline{S}_i and \overline{S}_j are real vectors constructed by concatenating the real and imaginary parts of S_i and S_j , respectively:

$$\begin{aligned} \overline{S}_i &= [\Re(S_i), \Im(S_i)]^T, \\ \overline{S}_j &= [\Re(S_j), \Im(S_j)]^T, \end{aligned} \quad (19)$$

and $\angle \overline{S}_i, \overline{S}_j$ denotes the angle between the two $2M$ -dimensional real vectors.

Therefore, instead of solving the (computationally complex) original optimization in (12), we can solve the following simplified maximin problem to find a close-to-optimal L -point multilevel constellation of $1 \times M$ vectors with average power P :

$$\begin{aligned} & \text{maximize} && \min \left\{ \min_{k=1, \dots, K} \mathcal{D}_{intra}(k), \min_{k=1, \dots, K-1} \mathcal{D}_{inter}(k, k+1) \right\}, \\ 1 \leq K \leq L, \frac{1}{L} \sum_{k=1}^K l_k r_k^2 &= P, \sum_{k=1}^K l_k &= L \\ 0 \leq r_1 < r_2 < \dots < r_K & & \end{aligned} \quad (20)$$

where, without loss of generality, we have assumed that $r_1 < r_2 < \dots < r_K$.

In (20), K and l_1, \dots, l_K are discrete variables, while r_1, \dots, r_K are continuous variables. This optimization problem can be solved using the approach explained in [7]. For any fixed value of K and l_1, \dots, l_K satisfying the specified constraints, (20) reduces to a continuous optimization over r_1, \dots, r_K , which can be solved numerically. Moreover, since the intra-subset

distance is an increasing function of r_k , the solution of (20) 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.

As mentioned above, the design problem for each subset is equivalent to a packing problem on the surface of an M -dimensional complex ($2M$ -dimensional real) sphere. This is a well-studied problem (see, e.g., [19] and references therein). However, since the design and decoding complexities of the optimal packings are usually high, in the next section we propose a recursive construction for the spherical constellations, which results in a systematic design and, in some cases, low complexity decoding algorithm.

C. A Recursive Construction for Spherical Constellations

We denote, by $S_n(L)$, the L -point recursively constructed n -dimensional real spherical constellation. We start from $n = 2$, and define

$$S_2(L) = \left\{ \left[\begin{array}{c} \cos((l-1)2\pi/L) \\ \sin((l-1)2\pi/L) \end{array} \right] \right\}_{l=1}^L. \quad (21)$$

For $n > 2$, we construct the constellation by using a number of $(n-1)$ -dimensional recursive constellations as latitudes of the n -dimensional constellation. In the following, we explain this procedure with more detail for the case of $n = 3$.

Fig. 3 shows a 32-point 3-dimensional constellation ($S_3(32)$) constructed from two $S_2(9)$ and one $S_2(14)$ constellations. The minimum angle between points in the $S_2(9)$ constellations is denoted by β . However, when this constellation is used as a subset of an S_3 constellation, the effective angle between the constellation points is no longer β . The following equation gives the effective minimum intra-subset angle, α ,

$$\sin\left(\frac{\alpha}{2}\right) = \cos(\theta) \sin\left(\frac{\beta}{2}\right), \quad (22)$$

where θ is the latitude of the subset.

Assuming a constellation of P levels (latitudes), we define the intra-subset and inter-subset distances by

$$d_{intra}(p) = \sin^2(\alpha_p), \quad \text{for } p = 1, \dots, P, \quad (23)$$

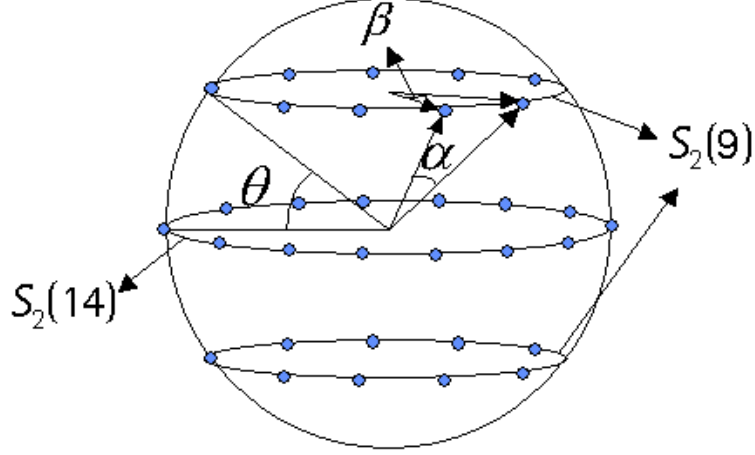


Fig. 3. A three-level three-dimensional real spherical constellation

and

$$d_{inter}(p, p') = \sin^2(\theta_p - \theta_{p'}), \quad \text{for } p, p' = 1, \dots, P, \quad (24)$$

where α_p is the effective minimum intra-subset angle of the p th subset, and θ_p and $\theta_{p'}$ are the latitudes of the subsets p and p' , respectively.

Next, similar to the approach in the previous section, we simplify the optimization problem by only maximizing the minimum of the intra-subset and inter-subset distances, instead of maximizing the minimum distance between all pairs of constellation points. For this, we solve the following optimization problem:

$$\begin{array}{ll} \text{maximize} & \min B \\ 1 \leq P \leq L, \sum_{p=1}^P l_p = L, -\frac{\pi}{2} \leq \theta_1 < \dots < \theta_P \leq \frac{\pi}{2} & \end{array}, \quad (25)$$

where l_p denotes the number of the points in the p th subset, and

$$B = \left\{ \{d_{intra}(p)\}_{p=1}^P, \{d_{inter}(p, p+1)\}_{p=1}^{P-1} \right\}. \quad (26)$$

Similar to the optimization in the previous section, here P and l_1, \dots, l_P are discrete variables, whereas $\theta_1, \dots, \theta_P$ are continuous variables. For a given choice of P and l_1, \dots, l_P , satisfying the specified constraints, we have a continuous optimization problem which can be solved numerically to find the optimal values for $\theta_1, \dots, \theta_P$. We do not usually need to try all of

the possible values for P . Starting from $P = 1$, and increasing the value of P by one at each time, the search can be stopped once the optimum minimum distance obtained from the above optimizations stops increasing. Also it can be shown that the optimal values for l_1, \dots, l_P will satisfy the following extra constraint

$$l_p \geq l_{p-1} \text{ or } l_p \geq l_{p+1}, \quad \text{for } p = 2, \dots, P - 1, \quad (27)$$

which can be used to further restrict the search domain.

For the case of $n > 3$, we use the same procedure as explained above, with the difference that instead of S_2 constellations, we use S_{n-1} constellations as the subsets, and construct the constellations recursively.

D. Performance Evaluation

We considered two different spectral efficiencies of 4 and 8 b/s/Hz, and designed partially coherent constellations for a 2×2 (two-transmit and two-receive antenna) system, for different values of channel estimation error (0.0, 0.01, 0.05, and 0.10). We evaluated the performance of these constellations through simulation. As our reference curves, we also simulated the case in which two independent QPSK or 16QAM constellations were used at the two transmit antennas (resulting in spectral efficiencies of 4 and 8 b/s/Hz, respectively). These cases are similar to the V-BLAST scheme of [20], with the difference that, in order to have a fair comparison, we used the optimal (ML) detector, and not a linear receiver as suggested in the V-BLAST scheme.

In order to compare the proposed constellations with a transmit diversity scheme, we considered the orthogonal transmit diversity approach of [21] (the Alamouti scheme), which has a similar decoding complexity to our scheme, and evaluated its performance for a 2×2 system with 16QAM and 256QAM constellations (4 and 8 b/s/Hz) through simulation. These orthogonal transmit diversity schemes are designed for $T = 2$. Therefore, even though our constellations are designed for $T = 1$, in our simulations, we used a channel with coherence time of two symbol intervals.

Figure 4 shows the symbol error rate curves for the case of 4 b/s/Hz and estimation variances of 0 (left) and 0.01 (right). As we see, in the absence of estimation error (left), the QPSK and the optimal two-antenna constellations have almost the same performance. We also notice that,

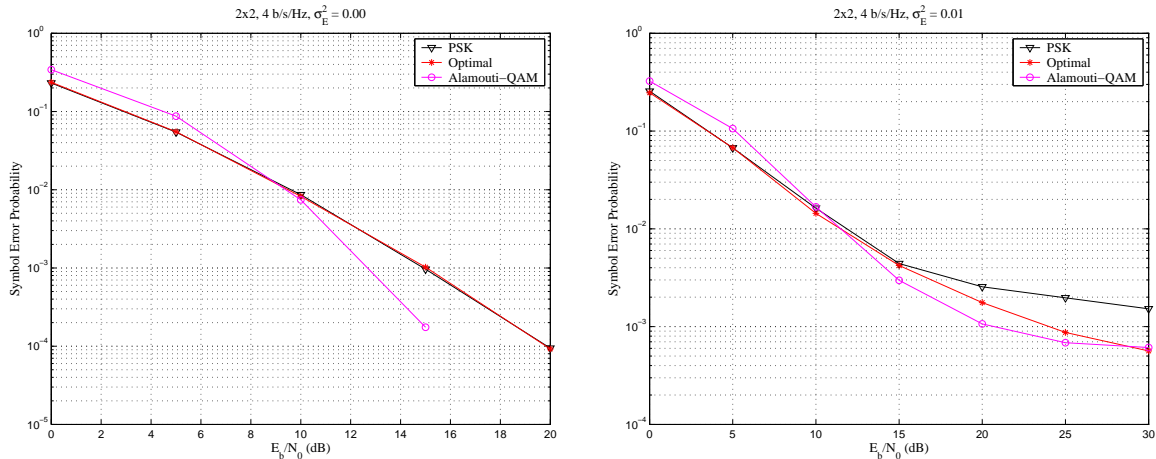


Fig. 4. Symbol error rate for 4 b/s/Hz with $M = N = 2$, and $\sigma_E^2 = 0.0$ (left) and $\sigma_E^2 = 0.01$ (right).

because of higher order of transmit diversity, at high SNR the Alamouti scheme in this case shows better performance compared to both QPSK and the optimal two-antenna constellations.

With 1% estimation error (Figure 4, right), the new constellations start showing better performance for SNR values larger than 15dB. We also notice that Alamouti scheme suffers from performance degradation at high SNR, and crosses the curve of the new constellations at around 30dB.

Figure 5 shows the symbol error rate curves for the case of 4 b/s/Hz and estimation variances of 0.05 (left) and 0.10 (right). As we see, the performance gain obtained by using the new constellations is substantial. We also notice that the Alamouti scheme suffers from a severe performance degradation because of channel estimation error, and its performance becomes even worse than the conventional QPSK constellations without any transmit diversity. The reason is that the linear combining at the receiver in the Alamouti scheme depends very much on the accuracy of the channel estimates. Since the channel coefficients are used for both interference cancellation and coherent demodulation at the receiver, the estimation errors have a two-fold effect on the performance, and result in severe increase in the error rate.

Figure 6 shows the symbol error rate curves for the case of 8 b/s/Hz and estimation variances of 0.0 (left) and 0.01 (right). As we see, in the absence of estimation error (left), the 16QAM and the optimal two-antenna constellations have almost the same performance. We also notice

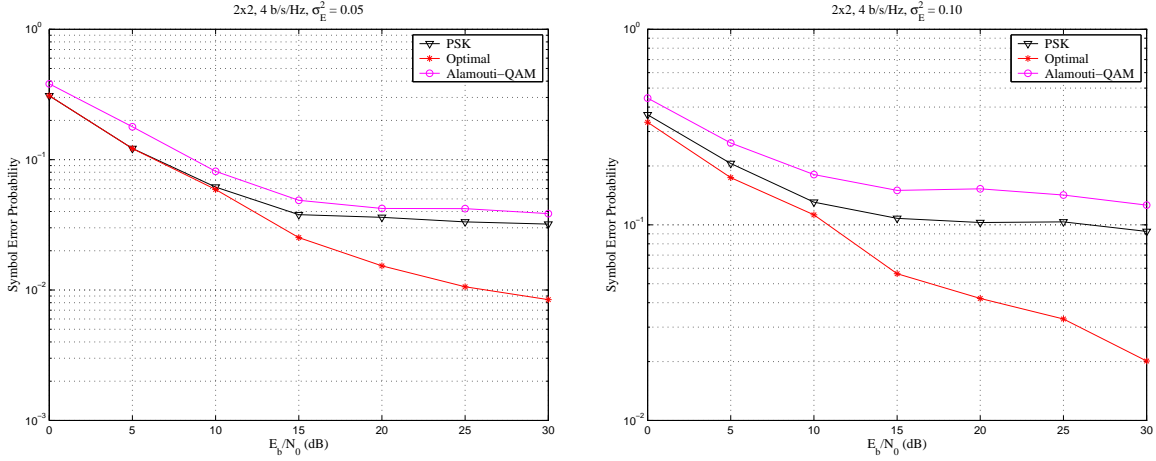


Fig. 5. Symbol error rate for 4 b/s/Hz with $M = N = 2$, and $\sigma_E^2 = 0.05$ (left) and $\sigma_E^2 = 0.10$ (right).

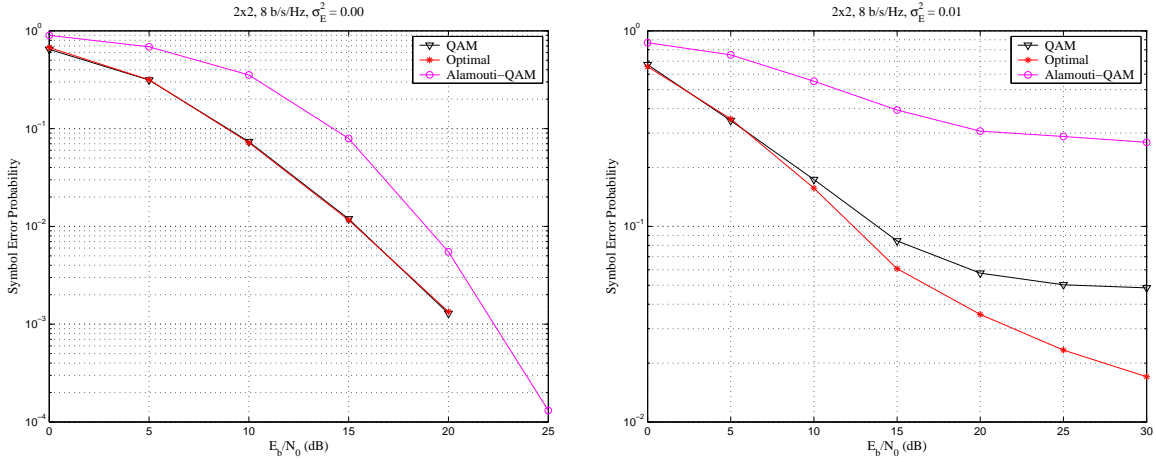


Fig. 6. Symbol error rate for 8 b/s/Hz with $M = N = 2$, and $\sigma_E^2 = 0.00$ (left) and $\sigma_E^2 = 0.01$ (right).

that, even though the Alamouti scheme has a larger transmit diversity advantage (larger slope in the figure), however because of smaller coding advantage performs worse than the conventional 16QAM constellations for SNR values of up to around 23dB.

With 1% estimation error (Figure 6, right), the new constellations show significant performance improvement compared to the 16QAM constellations. We also notice that Alamouti scheme suffers from a severe performance degradation and reaches an error floor of about 3×10^{-1} due to estimation errors.

V. PARTIALLY COHERENT CODED MODULATION

A. System Model

The system model that we consider in this section is similar to the one considered in Section II, with the difference that we consider B blocks of data, each of length T symbol intervals (where T is the coherence time of the channel). For this, we stack B signal matrices of size $T \times M$ (where M is the number of transmit antennas, to form the transmitted matrix, S . Similarly, we collect the received signals in a $T \times BN$ matrix, X , where N is the number of receive antennas. With these assumptions, the received matrix can be expressed in terms of the transmitted matrix, channel matrix, and the additive noise, using the following expression:

$$X = SH + W, \quad (28)$$

where

$$\begin{aligned} S &= \begin{bmatrix} S^1 & \cdots & S^B \end{bmatrix}, \\ X &= \begin{bmatrix} X^1 & \cdots & X^B \end{bmatrix}, \\ W &= \begin{bmatrix} W^1 & \cdots & W^B \end{bmatrix}, \end{aligned} \quad H = \begin{bmatrix} H^1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & H^B \end{bmatrix}, \quad (29)$$

and

$$\begin{aligned} S^b &= \begin{bmatrix} S_{11}^b & \cdots & S_{1M}^b \\ \vdots & \ddots & \vdots \\ S_{T1}^b & \cdots & S_{TM}^b \end{bmatrix}, & X^b &= \begin{bmatrix} X_{11}^b & \cdots & X_{1N}^b \\ \vdots & \ddots & \vdots \\ X_{T1}^b & \cdots & X_{TN}^b \end{bmatrix}, \\ H^b &= \begin{bmatrix} H_{11}^b & \cdots & H_{1N}^b \\ \vdots & \ddots & \vdots \\ H_{M1}^b & \cdots & H_{MN}^b \end{bmatrix}, & W^b &= \begin{bmatrix} W_{11}^b & \cdots & W_{1N}^b \\ \vdots & \ddots & \vdots \\ W_{T1}^b & \cdots & W_{TN}^b \end{bmatrix}, \end{aligned} \quad (30)$$

for $b = 1, \dots, B$. The entries of W are assumed to be independent circular complex Gaussian random variables from the distribution $\mathcal{CN}(0, 1)$. Also, with the block fading assumption on the channel with coherence interval of T , the non-zero entries of H are also independent circular complex Gaussian random variables from the distribution $\mathcal{CN}(0, 1)$. With these independence assumptions, we will have

$$p(X|S, H) = \prod_{b=1}^B p(X^b|S^b, H^b). \quad (31)$$

If we similarly define the $TM \times TN$ block diagonal matrices of channel estimates, \hat{H} , and the estimation error, \tilde{H} , at the receiver, as

$$\hat{H} = \begin{bmatrix} \hat{H}^1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \hat{H}^B \end{bmatrix}, \quad \tilde{H} = \begin{bmatrix} \tilde{H}^1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \tilde{H}^B \end{bmatrix}, \quad (32)$$

where

$$\hat{H}^b = \begin{bmatrix} \hat{H}_{11}^b & \cdots & \hat{H}_{1N}^b \\ \vdots & \ddots & \vdots \\ \hat{H}_{M1}^b & \cdots & \hat{H}_{MN}^b \end{bmatrix}, \quad \tilde{H}^b = \begin{bmatrix} \tilde{H}_{11}^b & \cdots & \tilde{H}_{1N}^b \\ \vdots & \ddots & \vdots \\ \tilde{H}_{M1}^b & \cdots & \tilde{H}_{MN}^b \end{bmatrix}, \quad (33)$$

for $b = 1, \dots, B$, so that $H = \hat{H} + \tilde{H}$, we will have

$$X = S(\hat{H} + \tilde{H}) + W. \quad (34)$$

We assume that the non-zero entries of \hat{H} and \tilde{H} are independent zero-mean circular complex Gaussian random variables. We also assume an estimation variance of σ_E^2 per channel coefficient, resulting in a $\mathcal{CN}(0, \sigma_E^2)$ distribution for the non-zero entries of \tilde{H} and a $\mathcal{CN}(0, (1 - \sigma_E^2))$ distribution for the entries of \hat{H} . By setting σ_E^2 equal to zero or one, this model reduces to the coherent and non-coherent system models, respectively.

Using the above distributions for the matrices \hat{H} and \tilde{H} , we will have

$$\begin{aligned} p(X|S, \hat{H}) &= \mathbb{E}_{\tilde{H}} \left\{ p(X|S, \hat{H}, \tilde{H}) \right\} \\ &= \prod_{b=1}^B \frac{\exp\left\{-\text{tr}\left[(I_T + \sigma_E^2 S^b S^{bH})^{-1} (X^b - S^b \hat{H}^b) (X^b - S^b \hat{H}^b)^H\right]\right\}}{\pi^{TN} \det^N(I_T + \sigma_E^2 S^b S^{bH})}. \end{aligned} \quad (35)$$

The ML decoder will find the signal matrix which maximizes the above expression for the given received matrix and channel estimate. Taking log of (35) and ignoring the common terms, we obtain the log-likelihood function as

$$\begin{aligned} L(X|S, \hat{H}) &= - \sum_{b=1}^B \text{tr} \left[(I_T + \sigma_E^2 S^b S^{bH})^{-1} (X^b - S^b \hat{H}^b) (X^b - S^b \hat{H}^b)^H \right] \\ &\quad - N \ln [\det(I_T + \sigma_E^2 S^b S^{bH})]. \end{aligned} \quad (36)$$

Since each term in the sum depends on the transmitted matrix only in one coherence interval, the decoder can use a Viterbi algorithm.

B. Code Design Criterion

Similar to Section III, we use the Kullback-Leibler (KL) distance as our performance criterion. Using (35) and the fact that the KL distance between two product distributions is the sum of the KL distances between the individual distributions, the KL distance between the two conditional distributions $p_i(X) = p(X|S_i, \hat{H})$ and $p_j(X) = p(X|S_j, \hat{H})$ will be given by:

$$\mathcal{D}(p_i||p_j) = \sum_{b=1}^B \mathcal{D}^b(\hat{H}^b), \quad (37)$$

where (see Equation (6))

$$\begin{aligned} \mathcal{D}^b(\hat{H}^b) = & N \text{tr} \left\{ (I_T + \sigma_E^2 S_i^b S_i^{bH}) (I_T + \sigma_E^2 S_j^b S_j^{bH})^{-1} \right\} - NT \\ & - N \ln \det \left\{ (I_T + \sigma_E^2 S_i^b S_i^{bH}) (I_T + \sigma_E^2 S_j^b S_j^{bH})^{-1} \right\} \\ & + \text{tr} \left\{ (I_T + \sigma_E^2 S_j^b S_j^{bH})^{-1} (S_i^b - S_j^b) \hat{H}^b \hat{H}^{bH} (S_i^b - S_j^b)^H \right\}. \end{aligned} \quad (38)$$

(For simplicity of the notation, we do not include the signal matrices S_i and S_j in the arguments of the function \mathcal{D}^b .)

As we see, these KL distances depend on \hat{H} , and cannot be directly used as a design metric. We would like to find an *expected KL distance* similar to Section III, to be able to derive the design criterion. According to Stein's Lemma [18], the KL distance gives the best achievable error exponent using hypothesis test. Therefore, asymptotically in N , the pairwise error probability of mistaking S_j for S_i , of the best hypothesis test designed to maximize the exponential decay rate of this error probability (this is not necessarily the ML detector) will be approximately given by

$$\text{Pr}_{\text{best}}(S_j \rightarrow S_i | \hat{H}) \approx \exp(-\mathcal{D}(p_i||p_j)) = \exp\left(-\sum_{b=1}^B \mathcal{D}^b(\hat{H}^b)\right). \quad (39)$$

To obtain the *expected KL distance*, we find the expected value of (39) with respect to the distribution of \hat{H} , which is a product distribution:

$$p(\hat{H}) = p(\hat{H}^1, \dots, \hat{H}^B) = \prod_{b=1}^B p(\hat{H}^b). \quad (40)$$

We will have

$$\Pr_{best}(S_j \rightarrow S_i) = \mathbb{E}_{\hat{H}} \left\{ \Pr_{best}(S_j \rightarrow S_i | \hat{H}) \right\} \quad (41)$$

$$\approx \mathbb{E}_{\hat{H}} \left\{ \exp \left(- \sum_{b=1}^B \mathcal{D}^b(\hat{H}^b) \right) \right\} \quad (42)$$

$$= \mathbb{E}_{\hat{H}^1, \dots, \hat{H}^B} \left\{ \prod_{b=1}^B \exp \left(- \mathcal{D}^b(\hat{H}^b) \right) \right\} \quad (43)$$

$$= \prod_{b=1}^B \mathbb{E}_{\hat{H}^b} \left\{ \exp \left(- \mathcal{D}^b(\hat{H}^b) \right) \right\} \quad (44)$$

$$= \prod_{b=1}^B \exp \left(- \bar{\mathcal{D}}^b(p_i \| p_j) \right) \quad (45)$$

$$= \exp \left(- \sum_{b=1}^B \bar{\mathcal{D}}^b(p_i \| p_j) \right), \quad (46)$$

where (see Equation (9))

$$\begin{aligned} \bar{\mathcal{D}}^b(p_i \| p_j) = & N \text{tr} \left\{ (I_T + \sigma_E^2 S_i^b S_i^{bH}) (I_T + \sigma_E^2 S_j^b S_j^{bH})^{-1} \right\} - NT \\ & - N \ln \det \left\{ (I_T + \sigma_E^2 S_i^b S_i^{bH}) (I_T + \sigma_E^2 S_j^b S_j^{bH})^{-1} \right\} \\ & + N \ln \det \left\{ I_M + (1 - \sigma_E^2) (S_i^b - S_j^b)^H (I_T + \sigma_E^2 S_j^b S_j^{bH})^{-1} (S_i^b - S_j^b) \right\}. \end{aligned} \quad (47)$$

From (46), we obtain the following expression for the overall *expected KL distance*:

$$\bar{\mathcal{D}}(p_i \| p_j) = \sum_{b=1}^B \bar{\mathcal{D}}^b(p_i \| p_j). \quad (48)$$

The code design criterion, therefore, will be to maximize the minimum of the sum KL distance (sum of the individual KL distances corresponding to different signal matrices in the code words). This is an interesting result, because it is very similar to the code design criterion in AWGN channels, where the design criterion is to maximize the minimum of the sum Euclidean distance [15]. As a result, we can use techniques similar to the ones used in designing coded modulation schemes for the AWGN channel, to design good outer codes for non-coherent and partially coherent systems. In the next section, we will use an approach similar to trellis coded modulation [15] to design such outer codes.

C. Code Construction

In the previous section, we showed that the *expected KL distance* is additive, i.e., the distance corresponding to two code words is the sum of the distances corresponding to the individual signal matrices in the code words. This property of the *expected KL distance* is similar to the additive property of the Euclidean distance, which is used by Trellis Coded Modulation (TCM) schemes [15] to design bandwidth-efficient trellis codes for AWGN channels. The main idea in coded modulation is to treat the modulation as an integral part of the encoding, and to design it in conjunction with the code to increase the minimum Euclidean distance between pairs of code words. The key to this joint coding and modulation is using an effective mapping method which is usually referred to as *mapping by set partitioning* [14].

In the method of *mapping by set partitioning*, the signal set is partitioned into several subsets of relatively large minimum intra-subset square Euclidean distance, while the minimum inter-subset distance is the same as the minimum distance of the original signal set. For example, a constellation of $L = 2^{n+k_2}$ signal points may be partitioned into 2^n subsets, each subset containing 2^{k_2} points. Each block of $m = k_1 + k_2$ information bits is also partitioned into two groups of k_1 and k_2 bits. The first group is encoded into n bits while the second group is left uncoded. Then, the n bits from the encoder are used to select one of the 2^n possible subsets, while the k_2 uncoded bits are used to select one of the 2^{k_2} points in the selected subset. In principle, block codes or convolutional codes can be used in the structure of coded modulation schemes. However, because of the simpler implementation of the soft-decision decoding of the convolutional codes and more generally trellis codes (due to the availability of the Viterbi algorithm), most of the coded modulation schemes use a trellis code as a subset encoder. In this case, the overall code (including encoded and uncoded bits) can be represented by a trellis with parallel transitions. These parallel transitions correspond to the same encoded input bits but different uncoded bits, so that the resulting outputs are from the same subset.

For a trellis coded modulation scheme, the minimum Euclidean distance between code words will be equal to the minimum of the following two quantities:

- a) minimum intra-subset Euclidean distance (due to the parallel transitions),
- b) the minimum distance in the trellis of the constituent code, usually referred to as the *free*

Euclidean distance of the code.

The set partitioning is performed with the goal of maximizing the first quantity, whereas the trellis of the constituent code is designed to maximize the second quantity. For a more detailed explanation and some heuristic rules for trellis construction refer to [15]. With an appropriate set partitioning and trellis design, the overall minimum distance of the code will be large enough to overcome the loss from the constellation expansion (due to the redundancy in the code), and provide a significant coding gain.

Similar techniques have been developed in [16], [17] for coded modulation in fast fading scenarios, when receiver is assumed to have perfect channel state information. The design criteria in this case are maximizing the symbol Hamming distance and the minimum product Euclidean distance between pairs of codewords. Therefore, the set partitioning and trellis design is performed to maximize the length of the shortest error event path and the product of the Euclidean distances along this path.

In this work, we use an idea similar to trellis coded modulation in the AWGN channels, to design good outer codes for the partially coherent constellations. The main difference between these codes and the conventional TCM schemes is that the design criterion in our case is maximizing the minimum KL distance corresponding to the pairs of code words, rather than the Euclidean distance as in TCM. Therefore, we have to perform the set partitioning and also the trellis design, based on the KL distance instead of the Euclidean distance between constellation points. As an example, we consider the four-level 16-point constellation designed for a channel estimation variance of 0.01. This constellation is shown in Figure 7 for SNR of 20dB per bit. This constellation is designed using our multilevel approach as explained in Section IV-A. We partition this constellation into eight subsets, each one containing two points. These subsets are shown in Figure 7 with different colors and markers. The partitioning is performed with the goal of maximizing the minimum intra-subset KL distance. At SNR of 20dB per bit, the original minimum KL distance of the constellation is around 3.8971, whereas the minimum intra-subset distance in the partitioned constellation is 5.5361. Even though the increase in the minimum intra-subset KL distance from this partitioning is not as significant as the increase in the intra-subset Euclidean distance of the conventional TCM schemes, as we will see in the simulation

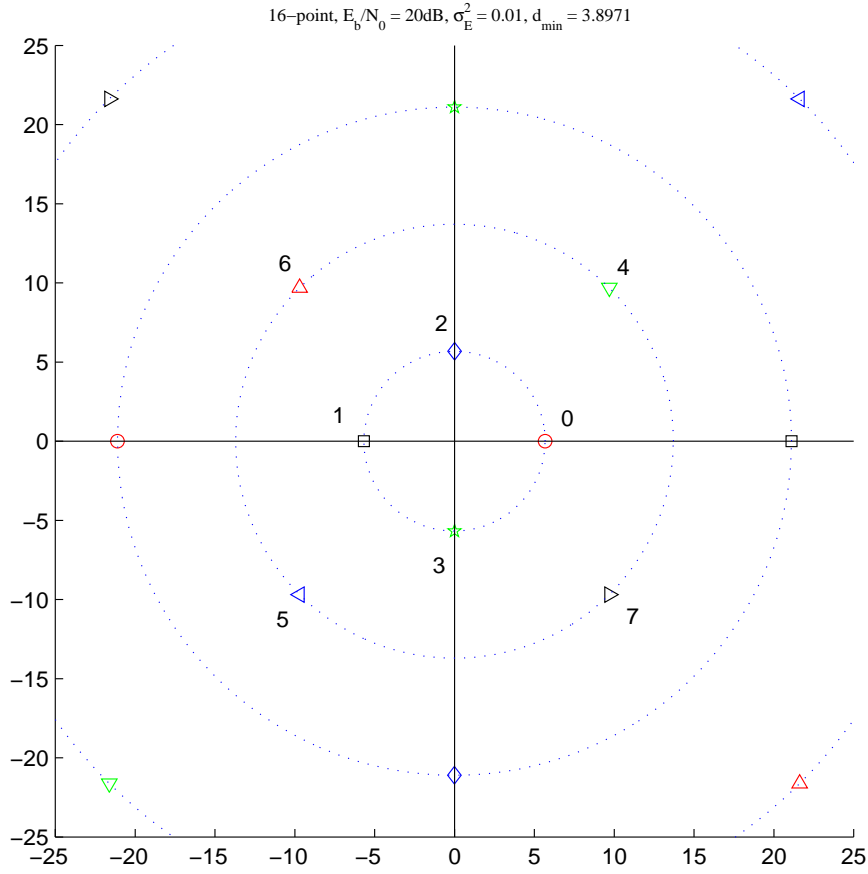


Fig. 7. The 16-point partially coherent constellation designed for estimation variance of 0.01 at SNR of 20dB per bit, and partitioned to maximize the minimum intra-subset KL distance.

results, it can still provide a substantial coding gain.

From each block of three information bits, we encode two bits using a 16-state rate $2/3$ convolutional code with octal transfer function of $[1\ 4\ 2; 4\ 3\ 0]$, to produce three encoded bits. We use these three bits to select one of the eight subsets of the constellation, and then we use the remaining uncoded bit to select one point from the selected subset. In Figure 7, the numbers next to the eight innermost signal points represent the labels of their corresponding subsets in the partitioned constellation.

Figure 8(a) compares the performances of the uncoded 8-point and 16-point partially coherent constellations with the 3 b/s/Hz trellis coded partially coherent modulation. The gap between

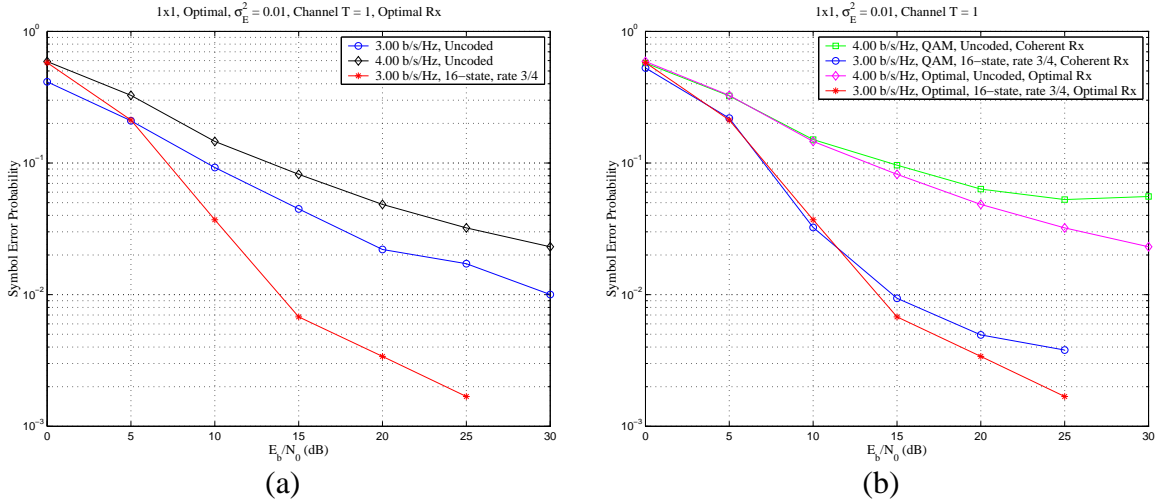


Fig. 8. (a) Coding gain, and (b) performance improvement over conventional trellis coded 16QAM, obtained by using the KL-based trellis coded partially coherent modulation at $\sigma_E^2 = 0.01$.

the performances of the 8-point and 16-point constellations is the loss due to the constellation expansion. As we see, the performance improvement from the coding scheme is so much that it not only overcomes this loss, but also provides a substantial gain over the uncoded scheme with the same spectral efficiency.

Figure 8(b) compares the performances of the uncoded and trellis coded 16QAM with the partially coherent constellations at $\sigma_E^2 = 0.01$. The coded modulation scheme for the 16QAM constellation is designed with the assumption of perfect channel state information at the receiver. The trellis coded partially coherent scheme is designed based on the KL distance. As we see, the partially coherent designs show similar performance improvements for the coded and uncoded systems in this case.

As a second example, we considered the case when channel estimation variance is 0.05. The coding gain and performance comparisons with the conventional approaches are given in Figures 9(a) and (b). As we see, the performance improvement over the uncoded and trellis coded 16QAM in this case is even more significant than the previous case. In general, as the channel estimation variance increases, larger performance gains can be obtained by using the partially coherent designs.

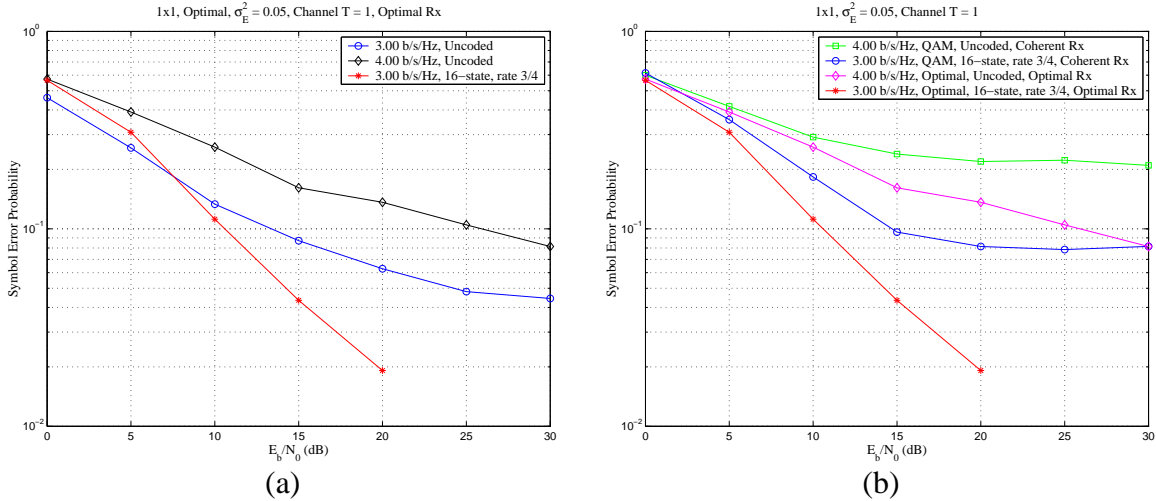


Fig. 9. (a) Coding gain, and (b) performance improvement over conventional trellis coded 16QAM, obtained by using the KL-based trellis coded partially coherent modulation at $\sigma_E^2 = 0.05$.

In the third example, we consider a system with two transmit and two receive antennas and a transmission rate of 7 b/s/Hz. We use the rate 2/3 code of the previous example to encode two bits out of every seven input bits, and leave the remaining five bits uncoded. We also use the 256-point constellations designed in the previous section for the estimation variance of 0.01, and perform an eight-way set partitioning on them to obtain eight subsets of 32 points each. The three encoded bits are used to select one of the eight possible subsets. Then the five uncoded bits are used to choose one of the 32 points in that subset.

Figure 10(a) compares the performances of the uncoded 7 b/s/Hz and 8 b/s/Hz 2×2 partially coherent systems with the trellis coded 7 b/s/Hz 2×2 partially coherent system. The gap between the performances of the uncoded 7 b/s/Hz and 8 b/s/Hz systems is the loss due to the constellation expansion. As we see in this figure, the performance improvement from the coding scheme is so much that it not only overcomes this loss, but also provides a substantial gain over the uncoded scheme with the same spectral efficiency.

We compare the performance of the above system with a system that uses a conventional 256QAM modulation in a trellis coded Alamouti scheme. The same 16-state rate 2/3 code is used to construct the outer trellis code for the coded Alamouti scheme. Figure 10(b) compares

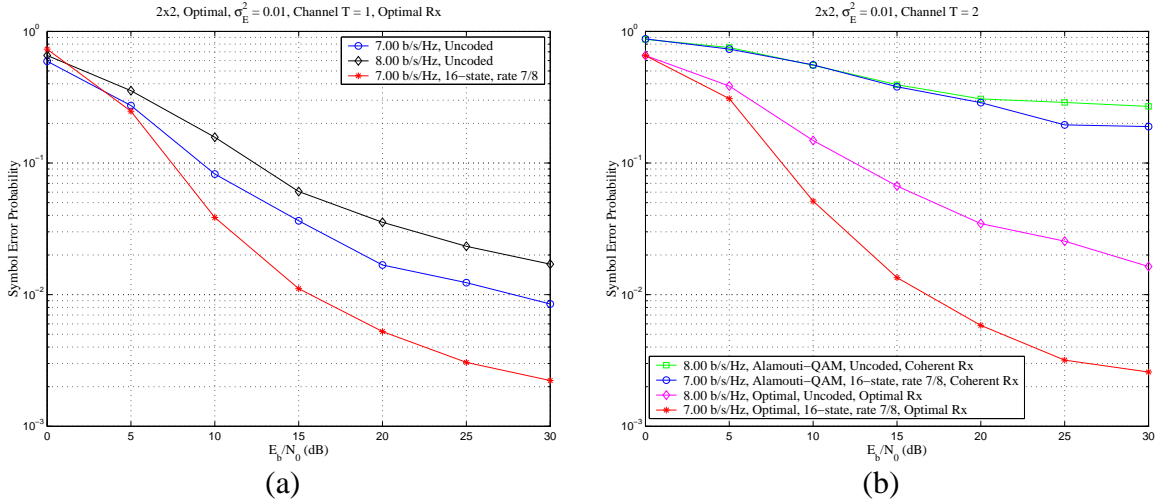


Fig. 10. (a) Coding gain, and (b) performance improvement over trellis coded Alamouti scheme with 256QAM, obtained by using the KL-based trellis coded partially coherent modulation for a 2×2 system at $\sigma_E^2 = 0.01$.

the performances of the trellis coded 2×2 partially coherent constellations with the trellis coded Alamouti scheme at the same rates. For comparison, the error rate curves of the corresponding uncoded systems are also given in this figure. As we see, the performance gain obtained by using the partially coherent designs in the coded systems is even larger than the uncoded case.

In the above example, the coherence interval of the channel is equal to two symbol intervals, which is the minimum required by the Alamouti scheme. The partially coherent coded modulation scheme, however, is designed for a coherence interval of one symbol interval. We compare the performances of the above partially coherent constellations and coded modulation schemes with different values of the channel coherence interval in Figure 11. As expected, the performance of the uncoded system is not affected by larger values of the channel coherence interval. On the other hand, since the coded system assumes independent fading coefficients from one symbol interval to the next one, larger values of channel coherence interval result in performance degradation. We see, however, that the performance degradation is not significant, especially compared to the gains provided by the proposed schemes.

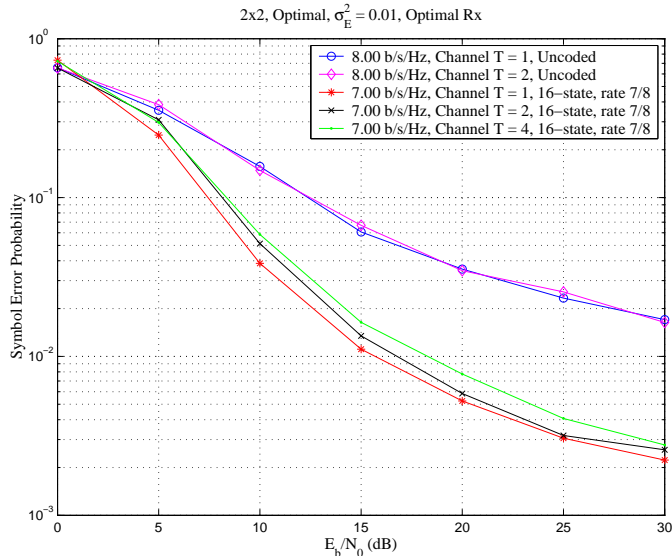


Fig. 11. Error rate performances of the partially coherent constellations and coded modulation schemes designed for a 2×2 system at $\sigma_E^2 = 0.01$, evaluated for different values of the channel coherence interval.

VI. CONCLUSIONS

We considered the problem of digital communication in a Rayleigh flat fading environment using a multiple antenna system, when only partial (imperfect) channel state information is available at the receiver. We derived the design criterion for space-time constellations in this scenario based on the Kullback-Leibler distance between conditional received distributions. Through a single-antenna example, we demonstrated the effect of the channel estimation error in the shape and structure of the optimal constellations. Using a novel recursive construction for spherical constellations, we also designed multiple-antenna vector constellations for the partially coherent systems. Through simulation, we showed that in the presence of channel estimation errors, the new constellations can provide significant improvement in the performance as compared to the conventional PSK and QAM constellations, and multiple-antenna techniques such as BLAST and OTD.

We also showed that the *expected KL distance* in block fading channels has an additive property similar to the additive property of the Euclidean distance in AWGN channels. Based on this property, we proposed to use the idea of *mapping by set partitioning*, to design partially coherent

trellis coded modulations for the block fading channels. We demonstrated the effectiveness of the proposed criterion and technique by several examples, and showed that the designed coded schemes can provide substantial coding gains. We also compared the performances of the partially coherent designs with the conventional trellis coded 16QAM constellations and trellis coded Alamouti schemes, and showed that, in the presence of channel estimation errors, significant performance gains can be obtained by using the new coded modulation schemes.

APPENDIX A

DERIVATION OF THE KL DISTANCE

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

$$\mathcal{D}(p_i \| p_j) = \mathbb{E}_{p_i} \left\{ \ln \left[\frac{p_i(X)}{p_j(X)} \right] \right\}. \quad (49)$$

Substituting (3) for p_i and p_j , we will have

$$\mathcal{D}(p_i \| p_j) = N \ln \det [(I_T + S_i S_i^H)^{-1} (I_T + S_j S_j^H)] - T_2 + T_3, \quad (50)$$

where

$$T_2 = \mathbb{E}_{p_i} \left\{ \text{tr} \left[(I_T + \sigma_E^2 S_i S_i^H)^{-1} (X - S_i \hat{H})(X - S_i \hat{H})^H \right] \right\} \quad (51)$$

and

$$T_3 = \mathbb{E}_{p_i} \left\{ \text{tr} \left[(I_T + \sigma_E^2 S_j S_j^H)^{-1} (X - S_j \hat{H})(X - S_j \hat{H})^H \right] \right\}. \quad (52)$$

Again, using (3) for p_i , we have

$$\begin{aligned} T_2 &= \sum_{n=1}^N \mathbb{E}_{p_i} \left\{ (X_n - S_i \hat{H}_n)^H (I_T + \sigma_E^2 S_i S_i^H)^{-1} (X_n - S_i \hat{H}_n) \right\} \\ &= \sum_{n=1}^N \text{tr} \left[(I_T + \sigma_E^2 S_i S_i^H)^{-1} (I_T + \sigma_E^2 S_i S_i^H) \right] = NT. \end{aligned} \quad (53)$$

To calculate T_3 , we write

$$\begin{aligned} (X - S_j \hat{H})(X - S_j \hat{H})^H &= (X - S_i \hat{H})(X - S_i \hat{H})^H + X \hat{H}^H (S_i - S_j)^H + \\ &S_i \hat{H} X^H - S_j \hat{H} X^H - S_i \hat{H} \hat{H}^H S_i^H + S_j \hat{H} \hat{H}^H S_j^H. \end{aligned} \quad (54)$$

Therefore, T_3 consists of two terms,

$$T_3 = T_4 + T_5, \quad (55)$$

where

$$\begin{aligned}
T_4 &= \mathbb{E}_{p_i} \left\{ \text{tr} \left[(I_T + \sigma_E^2 S_j S_j^H)^{-1} (X - S_i \hat{H}) (X - S_i \hat{H})^H \right] \right\} \\
&= \sum_{n=1}^N \mathbb{E}_{p_i} \left\{ (X_n - S_i \hat{H}_n)^H (I_T + \sigma_E^2 S_j S_j^H)^{-1} (X_n - S_i \hat{H}_n) \right\} \\
&= \sum_{n=1}^N \text{tr} \left[(I_T + \sigma_E^2 S_j S_j^H)^{-1} (I_T + \sigma_E^2 S_i S_i^H) \right] \\
&= N \text{tr} \left[(I_T + \sigma_E^2 S_i S_i^H) (I_T + \sigma_E^2 S_j S_j^H)^{-1} \right]
\end{aligned} \tag{56}$$

and

$$\begin{aligned}
T_5 &= \mathbb{E}_{p_i} \left\{ \text{tr} \left[(I_T + \sigma_E^2 S_j S_j^H)^{-1} \left(X \hat{H}^H (S_i - S_j)^H + \right. \right. \right. \\
&\quad \left. \left. \left. S_i \hat{H} X^H - S_j \hat{H} X^H - S_i \hat{H} \hat{H}^H S_i^H + S_j \hat{H} \hat{H}^H S_j^H \right) \right] \right\} \\
&= \text{tr} \left[(I_T + \sigma_E^2 S_j S_j^H)^{-1} \left(S_i \hat{H} \hat{H}^H (S_i - S_j)^H + \right. \right. \\
&\quad \left. \left. S_i \hat{H} \hat{H}^H S_i^H - S_j \hat{H} \hat{H}^H S_i^H - S_i \hat{H} \hat{H}^H S_i^H + S_j \hat{H} \hat{H}^H S_j^H \right) \right] \\
&= \text{tr} \left[(I_T + \sigma_E^2 S_j S_j^H)^{-1} \left(S_i \hat{H} \hat{H}^H (S_i - S_j)^H - S_j \hat{H} \hat{H}^H (S_i - S_j)^H \right) \right] \\
&= \text{tr} \left[(I_T + \sigma_E^2 S_j S_j^H)^{-1} (S_i - S_j) \hat{H} \hat{H}^H (S_i - S_j)^H \right].
\end{aligned} \tag{57}$$

Substituting (53), (55), (56), and (57) in (50) we will have

$$\begin{aligned}
\mathcal{D}(p_i \| p_j) &= N \text{tr} \left[(I_T + \sigma_E^2 S_i S_i^H) (I_T + \sigma_E^2 S_j S_j^H)^{-1} \right] - NT \\
&\quad - N \ln \det \left[(I_T + \sigma_E^2 S_i S_i^H) (I_T + \sigma_E^2 S_j S_j^H)^{-1} \right] \\
&\quad + \text{tr} \left[(I_T + \sigma_E^2 S_j S_j^H)^{-1} (S_i - S_j) \hat{H} \hat{H}^H (S_i - S_j)^H \right].
\end{aligned} \tag{58}$$

APPENDIX B

DERIVATION OF THE *expected KL Distance*

In this appendix, we derive the expression for the *expected KL distance* between two constellation matrices S_i and S_j . It is assumed that the best achievable conditional pairwise error probability is given by

$$\text{Pr}_{\text{best}} \left(S_j \rightarrow S_i | \hat{H} \right) \approx \exp \left(-\mathcal{D}(p_i \| p_j) \right), \tag{59}$$

where $\mathcal{D}(p_i \| p_j)$ is as given by 6. To find the *expected KL distance*, we first take the expectation of the above equation with respect to the distribution of \hat{H} , given by

$$p(\hat{H}) = \frac{1}{\pi^{MN} (1 - \sigma_E^2)^{MN}} \exp \left\{ -\text{tr} \left[(1 - \sigma_E^2)^{-1} \hat{H} \hat{H}^H \right] \right\}. \tag{60}$$

For this, we break the KL distance in 6 into the sum of two terms, $\mathcal{D}(p_i||p_j) = \mathcal{D}_1 + \mathcal{D}_2$, where

$$\begin{aligned} \mathcal{D}_1 = & N \text{tr} \left[(I_T + \sigma_E^2 S_i S_i^H) (I_T + \sigma_E^2 S_j S_j^H)^{-1} \right] - NT \\ & - N \ln \det \left[(I_T + \sigma_E^2 S_i S_i^H) (I_T + \sigma_E^2 S_j S_j^H)^{-1} \right] \end{aligned} \quad (61)$$

and

$$\mathcal{D}_2 = \text{tr} \left[(I_T + \sigma_E^2 S_j S_j^H)^{-1} (S_i - S_j) \widehat{H} \widehat{H}^H (S_i - S_j)^H \right]. \quad (62)$$

Now we have

$$\Pr_{best}(S_j \rightarrow S_i) = \mathbb{E}_{\widehat{H}} \left\{ \Pr_{best}(S_j \rightarrow S_i | \widehat{H}) \right\} \quad (63)$$

$$\approx \mathbb{E}_{\widehat{H}} \left\{ \exp(-\mathcal{D}(p_i||p_j)) \right\} \quad (64)$$

$$= \mathbb{E}_{\widehat{H}} \left\{ \exp(-\mathcal{D}_1 - \mathcal{D}_2) \right\} \quad (65)$$

$$= \mathbb{E}_{\widehat{H}} \left\{ \exp(-\mathcal{D}_1) \exp(-\mathcal{D}_2) \right\} \quad (66)$$

$$= \exp(-\mathcal{D}_1) \mathbb{E}_{\widehat{H}} \left\{ \exp(-\mathcal{D}_2) \right\} \quad (67)$$

since \mathcal{D}_1 does not depend on \widehat{H} . Using the pdf of \widehat{H} given in (60), we can find the expectation in the second term of the above equation:

$$\mathbb{E}_{\widehat{H}} \left\{ \exp(-\mathcal{D}_2) \right\} = \frac{\int_{\mathbb{C}^{MN}} \exp \left\{ -\mathcal{D}_2 - \text{tr} \left[(1 - \sigma_E^2)^{-1} \widehat{H} \widehat{H}^H \right] \right\} d\widehat{H}}{\pi^{MN} (1 - \sigma_E^2)^{MN}} \quad (68)$$

$$= \frac{\int_{\mathbb{C}^{MN}} \exp \left\{ -\text{tr} \left[(1 - \sigma_E^2)^{-1} \widehat{H} \widehat{H}^H + (S_i - S_j)^H (I_T + \sigma_E^2 S_j S_j^H)^{-1} (S_i - S_j) \widehat{H} \widehat{H}^H \right] \right\} d\widehat{H}}{\pi^{MN} (1 - \sigma_E^2)^{MN}} \quad (69)$$

$$= \frac{\det^N \left[(1 - \sigma_E^2)^{-1} I_M + (S_i - S_j)^H (I_T + \sigma_E^2 S_j S_j^H)^{-1} (S_i - S_j) \right]^{-1}}{(1 - \sigma_E^2)^{MN}} \quad (70)$$

$$= \det^{-N} \left[I_M + (1 - \sigma_E^2) (S_i - S_j)^H (I_T + \sigma_E^2 S_j S_j^H)^{-1} (S_i - S_j) \right]. \quad (71)$$

Substituting (71) in (67), we will have

$$\Pr_{best}(S_j \rightarrow S_i) \approx \exp(-\overline{\mathcal{D}}(p_i||p_j)), \quad (72)$$

where

$$\begin{aligned} \overline{\mathcal{D}}(S_i||S_j) = & N \text{tr} \left[(I_T + \sigma_E^2 S_i S_i^H) (I_T + \sigma_E^2 S_j S_j^H)^{-1} \right] - NT \\ & - N \ln \det \left[(I_T + \sigma_E^2 S_i S_i^H) (I_T + \sigma_E^2 S_j S_j^H)^{-1} \right] \\ & + N \ln \det \left[I_M + (1 - \sigma_E^2) (S_i - S_j)^H (I_T + \sigma_E^2 S_j S_j^H)^{-1} (S_i - S_j) \right]. \end{aligned} \quad (73)$$

REFERENCES

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] M. J. Borran, A. Sabharwal, and B. Aazhang, "On design criteria and construction of non-coherent space-time constellations," *IEEE Transactions on Information Theory, Special Issue on Space-Time Transmission, Reception, Coding, and Signal Design*, vol. 49, no. 10, pp. 2332–2351, Oct. 2003.
- [8] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 2007–2019, Sept. 1999.
- [9] M. Medard and A. J. Goldsmith, "Capacity of time-varying channels with channel side information at the sender and receiver," in *Communication Theory Mini-Conference*, June 1999, pp. 16–20.
- [10] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [11] A. Lapidoth and S. Shamai, "Fading channels: How perfect need "perfect side information" be?" *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.
- [12] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, April 2003.
- [13] 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.
- [14] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Transactions on Information Theory*, vol. IT-28, pp. 55–67, Jan. 1982.
- [15] J. G. Proakis, *Digital Communications*. McGraw Hill, 1995.
- [16] D. Divsalar and M. K. Simon, "The design of trellis coded MPSK for fading channels: Performance criteria," *IEEE Transactions on Communications*, vol. 36, no. 9, pp. 1004–1012, Sept. 1988.
- [17] —, "The design of trellis coded MPSK for fading channels: Set partitioning for optimum code design," *IEEE Transactions on Communications*, vol. 36, no. 9, pp. 1013–1021, Sept. 1988.
- [18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley Interscience, 1991.

- [19] N. J. A. Sloane, "Tables of sphere packings and spherical codes," *IEEE Transactions on Information Theory*, vol. 27, no. 3, pp. 327–338, May 1981.
- [20] 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.
- [21] 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.