Sample Publications*

Mohammad Jaber Borran

*All of the appeared or submitted publications, including four journal papers and eight conference papers are available at

http://www.ece.rice.edu/~mohammad/publicat.htm


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

Mohammad Jaber Borran, Ashutosh Sabharwal, Behnaam Aazhang, and Don H. Johnson
ECE Department, Rice University, 6100 Main St., MS-366, Houston, TX 77005-1892
Email: {mohammad, ashu, aaz, dhj}@rice.edu

We consider a non-coherent communication system with \( M \) transmit and \( N \) receive antennas in a block Rayleigh flat fading channel with coherence interval of \( T \) symbol periods. We use the following complex baseband notation

\[
X = SH + W,
\]

where \( S \) is the \( T \times M \) transmitted matrix, \( X \) is the \( T \times N \) received matrix, and \( H \) and \( W \) are the unknown (to both transmitter and receiver) \( M \times N \) and \( T \times N \) matrices of the i.i.d. fading coefficients and additive noise terms from \( \mathcal{CN}(0,1) \). The transmitted symbols are also assumed to be power constrained:

\[
\sum_{i=1}^{M} \sum_{m=1}^{T} \mathbb{E}\{|x_{im}|^2\} = P.
\]

The capacity of non-coherent systems was studied in [1], where it was shown that at high SNR or when the coherence interval is much greater than the number of transmit antennas, capacity can be achieved by using a constellation of unitary matrices. However, at low SNR, or for small values of \( T \) (e.g., \( T = 1 \)), unitary constellations are no longer optimal. The exact expression for the pairwise error probability of unitary constellations and the Chernoff upper bound are also given in [1]. However, these expressions appear to be intractable for the general case. Therefore, inspired by Stein’s lemma [3], we propose the use of Kullback-Leibler [3] distance between distributions to approximate the exponential decay rate of pairwise error probabilities. The design problem will then be to seek constellations that have the lowest minimum KL distance between the received distributions assigned to their elements (maximin code design).

The resulting constellations coincide with the unitary designs at very high SNR or very low rates. But for high rate codes or at low SNR, different signal sets are obtained which show better probability of error performance.

The KL distance between \( p_i = p(X|S_i) \) and \( p_j = p(X|S_j) \), where \( S_i \) and \( S_j \) are two different \( T \times M \) matrices from the constellation, can be calculated as

\[
\mathcal{D}(p_i||p_j) = N \cdot \ln \left\{ |I_T + S_i S_j^H| |I_T + S_j S_i^H|^{-1} \right\} - NT - N \cdot \ln \det \left\{ |I_T + S_i S_j^H| |I_T + S_j S_i^H|^{-1} \right\},
\]

(2)

Examining (2) shows that the KL distance consists of two parts. For example, for \( M = 1 \), (2) reduces to

\[
\mathcal{D}(p_i||p_j) = \frac{1}{4} \left[ \frac{1}{|S_i|^2} \ln \left( 1 + \frac{|S_i|^2}{|S_j|^2} \right)^2 \right] - 1 + \frac{1}{|S_j|^2} \left[ \frac{1}{|S_i|^2} |S_i|^2 + \frac{1}{|S_j|^2} |S_j|^2 \right],
\]

(3)

in which, the first part is due to having different magnitudes (lying on different spheres in \( \mathbb{C}^T \)), and the second part is due to the angle between the points (lying on different one-dimensional subspaces of \( \mathbb{C}^T \)). This decoupling property suggests partitioning the signal space into subsets of concentric spheres and using only intra-subset and intersubset KL distances in the maximin problem.

The unitary space-time codes of [1] have exponential encoding and decoding complexity. In [2] suboptimal unitary constellations with very low encoding and decoding complexity are proposed. In this work, we use the low complexity designs of [2] as the subsets of our multilevel constellation. The resulting 16-point real constellation for \( T = 2 \) and \( \text{SNR} = 10 \text{dB} \) is a two level constellation, with 4 points on the inner circle and 12 points on the outer circle. The error rate performance of this constellation is simulated for different values of \( N \) and compared with the corresponding constellation proposed in [2]. The results are shown in Fig. 1. As we see, by using multilevel constellations (which use contributions from both parts of the KL distance) instead of single level unitary constellations (in which the first part of the KL distance is always zero), the performance of the non-coherent systems can be significantly improved. Moreover, the new constellations have a decoding complexity similar to the low complexity designs of [2].

REFERENCES


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

Mohammad Jaber Borran, Ashutosh Sabharwal, and Behnaam Aazhang
ECE Department, MS-366, Rice University, Houston, TX 77005-1892
Email: {mohammad,ashu,aaz}@rice.edu
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 or 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 a large number of 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

Non-coherent constellations, space-time codes, multiple antenna systems, fading channels, channel coding, wireless communications.

I. INTRODUCTION

Exploiting propagation diversity by using multiple antennas at the transmitter and receiver in wireless communication systems has been recently proposed and studied using different approaches [1–8]. In [1, 2], it was 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 this non-coherent system has been studied in [5, 6], where it has been shown that at high SNR’s, 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.

Regardless of their high design and decoding complexity, the high SNR or long coherence time requirements for the optimality of the unitary constellations make them less desirable for low-power applications in high-mobility environments. For example, all of the unitary codes 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 [10], 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, whereas the non-coherent constellations are usually needed when channel changes rapidly and training is difficult or expensive (in terms of time and power spent on that).

On the other hand, the high SNR requirement usually 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. 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. This has been a motivation for us to look for a design criterion which exploits the knowledge about statistics of the fading more efficiently. Since the expressions for average error probability are usually very complicated, and do not provide much insight for the constellation design, we will use the pairwise error probability as our performance and design criterion.

Unfortunately, the exact expression or even the Chernoff upper bound for the pairwise error probability does 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. Therefore, inspired by the Stein’s lemma, we propose the use of Kullback-Leibler (KL) distance between conditional distributions as design criterion.

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. 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 a large number of receive antennas are used.

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 assigned to the transmitted signals, 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

$$X = SH + W,$$

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 $\sqrt{\frac{T}{M}}$ 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 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 $S$. 

5
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

\[
p(X_n|S) = \mathbb{E}_H \left\{ p(X_n|S, H) \right\} = \frac{\exp \left\{ -X_n^H (I_T + SS^H)^{-1} X_n \right\}}{\pi^T \det (I_T + SS^H)}.
\] (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, \( X \),

\[
p^N(X|S) = \prod_{n=1}^{N} p(X_n|S) = \frac{\exp \left\{ -\text{tr} \left( (I_T + SS^H)^{-1} XX^H \right) \right\}}{\pi^T N \det^N (I_T + SS^H)}.
\] (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 \( L \), \( \{S_i\}_{i=1}^{L} \), and defining \( p_i^N(X) = p^N(X|S_i) \), the Maximum Likelihood (ML) detector for this system will have the following form

\[
\hat{S}_{ML} = S_{\hat{t}_{ML}}, \quad \text{where} \quad \hat{t}_{ML} = \arg \max_{i \in \{1, \ldots, L\}} p_i^N(X).
\] (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

\[
\Pr(S_1 \rightarrow S_2) = \Pr_{p_1^N} \left\{ X : p_2^N(X) > p_1^N(X) \right\},
\] (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

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

It is known [11] 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 [11] 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 $S_i^H S_i = \left( \frac{E_T}{N} \right) I_M$, 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 probability density function in (3) depends on $S$ only through $SS^H$. Since the pairwise error probability in (5) is determined by the two probability density functions, it is clear that the performance of the non-coherent multiple-antenna constellation also depends on the constellation matrices $\{S_i\}_{i=1}^L$ only through $\{S_i S_i^H\}_{i=1}^L$. 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 we can find a corresponding constellation of $T \times T$ matrices with the same pairwise and average error probabilities. Since the average power of the constellation can be written as

$$P_{av} = \frac{1}{T} \sum_{t=1}^{T} \sum_{m=1}^{M} \mathbb{E}\{ |s_{tm}|^2 \} = \frac{1}{T} \mathbb{E}\{ \text{tr} [SS^H] \} ,$$

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 non-coherent multiple-antenna systems.

Therefore, in the rest of this paper, we will always assume that $M \leq T$.

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

7
signal given the transmitted symbol is given by
\[ p^N(X|s) = \frac{1}{\pi^N} \frac{1}{(1 + |s|^2)^N} \exp \left( \frac{-|X|^2}{1 + |s|^2} \right). \] (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

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

Proof: Using (5) and (8) (with \( \rho = 1 \)), we have

\[
\Pr(s_1 \rightarrow s_2) = \Pr_{p_1} \{ x : p_2(x) > p_1(x) \} = \Pr_{p_1} \{ x : \frac{1}{\pi(1 + |s_2|^2)} \exp \left( \frac{-|x|^2}{1 + |s_2|^2} \right) > \frac{1}{\pi(1 + |s_1|^2)} \exp \left( \frac{-|x|^2}{1 + |s_1|^2} \right) \} \\
= \Pr_{p_1} \{ x : |x|^2 > A \} = \int \int p_1(\rho e^{i\theta}) \rho d\theta d\rho \\
= \frac{1}{1 + |s_1|^2} \int_A^{\infty} \exp \left( \frac{-\rho^2}{1 + |s_1|^2} \right) d(\rho^2) = \exp \left( \frac{-A}{1 + |s_1|^2} \right) \\
= B^{1-\pi},
\]

where

\[
A = \frac{1}{1 + |s_1|^2} - \frac{1}{1 + |s_2|^2} \ln \left( \frac{1 + |s_2|^2}{1 + |s_1|^2} \right) = \frac{(1 + |s_1|^2)(1 + |s_2|^2)}{|s_2|^2 - |s_1|^2} \ln \left( \frac{1 + |s_2|^2}{1 + |s_1|^2} \right),
\] (10)

and

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

Similarly, it can be shown that if \( |s_1| > |s_2| \), we have

\[
\Pr(s_1 \rightarrow s_2) = 1 - B^{1-\pi}.
\]

This completes the proof.

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

\[
\Pr(s_1 \rightarrow s_2) = \begin{cases} 
\sum_{n=0}^{N-1} \frac{1}{n!} \left( \frac{-N \ln(B)}{1-B} \right)^n \exp \left( \frac{N \ln(B)}{1-B} \right) & \text{for } B < 1 \\
1 - \sum_{n=0}^{N-1} \frac{1}{n!} \left( \frac{-N \ln(B)}{1-B} \right)^n \exp \left( \frac{N \ln(B)}{1-B} \right) & \text{for } B > 1
\end{cases}, \] (12)
where $B$ is as in (11).

**Proof:** Proof is by induction, and is given in Appendix I. \[\square\]

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

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

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

(13)

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

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

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

(14)

This discontinuity can be seen in Figure 1(a) which shows the two pairwise error probabilities as a function of $B$ for $N = 1$.

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 for 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 [11]) for this system is given by the following expression

$$
C(p_1^N, p_2^N) = N \left[ \frac{\ln(B)}{B - 1} - \ln \left( \frac{\ln(B)}{B - 1} \right) - 1 \right],
$$

(15)

where $B$ is as in (11).

**Proof:** See Appendix II. \[\square\]

Figure 1(b) shows the exact average error probability, and the Chernoff bound for the average error probability given by $\frac{1}{2} \exp \left\{ -NC(p_1^N, p_2^N) \right\}$, for $N = 1$. As we see, the Chernoff bound is also a monotonic function of $B$ in the two regions of $\{B < 1\}$ and $\{B > 1\}$. As we will see
Fig. 1. (a) Pairwise error probabilities, and (b) Average error probability and the Chernoff bound, for single antenna in fast fading.

later, the Kullback-Leibler (KL) distance between the two conditional distributions corresponding to the two different transmitted symbols, also has a similar behavior as a function of $B$. Therefore, the three different criteria of maximizing the minimum of (a) the exact pairwise error probability, (b) the Chernoff bound on the pairwise error exponent, and (c) the KL distance between the corresponding distributions, are all equivalent to minimizing the maximum of $B$ (assuming $B < 1$), and will result in the same constellation.

IV. DESIGN CRITERION

As mentioned in the previous section, the pairwise error probability of non-coherent ML detection is approximately given by (5), and does not appear to admit a simple closed form expression in general. Therefore, instead of (5), we will use the upper bound on the rate of its exponential decay (by number of independent observations) given by the Stein’s lemma [11], as our performance criterion. According to the Stein’s lemma, the best achievable error exponent for $\Pr(S_2 \rightarrow S_1)$ with hypothesis testing and with the constraint that $\Pr(S_1 \rightarrow S_2)$ is smaller than a given value, is given by $D(p(X|S_1)||p(X|S_2))$, the Kullback-Leibler (KL) distance [11] between $p(X|S_1)$ and $p(X|S_2)$. Even though this lemma does not give the error exponent for the maximum likelihood detector, but for a detector which is highly biased in favor of one of the hypotheses, it still provides an upper bound on the pairwise error exponent for 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 will see later, at least in the case of fast fading, the KL-based design criterion is equivalent to the design criterion based on the exact pairwise error probability or the Chernoff bound.

**Lemma 2:** Let $X_1, X_2, \cdots, X_N$ be i.i.d. $\sim q$. Consider two hypothesis tests, one between $q = p_0$ and $q = p_1$, and the other one between $q = p_0$ and $q = p_2$, where $\mathcal{D}(p_0\|p_2) < \mathcal{D}(p_0\|p_1) < \infty$. Let $L_{1N} = \sum_{n=1}^{N} \ln \left( \frac{p_0(X_n)}{p_1(X_n)} \right)$ and $L_{2N} = \sum_{n=1}^{N} \ln \left( \frac{p_0(X_n)}{p_2(X_n)} \right)$ denote the log-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 $P_{e1N} = \Pr_{p_0} \{ L_{1N} < 0 \}$ and $P_{e2N} = \Pr_{p_0} \{ L_{2N} < 0 \}$. Then

a) $\Pr_{p_0} \{ L_{1N} < L_{2N} \} \to 0$ as $N \to \infty$, and

b) $\Pr_{p_0} \{ L_{1N} < 0 | L_{2N} \geq 0 \} \to 0$ as $N \to \infty$.

The statement (b) above, implies that for large $N$, if the second test is successful, then so will be the first one, with high probability.

**Proof:**

a) By the weak law of large numbers, we have

\[
\frac{1}{N} L_{1N} = \frac{1}{N} \sum_{n=1}^{N} \ln \left( \frac{p_0(X_n)}{p_1(X_n)} \right) \to \mathbb{E}_{p_0} \left\{ \ln \left( \frac{p_0(X_1)}{p_1(X_1)} \right) \right\} = \mathcal{D}(p_0\|p_1)
\]

in probability w.r.t $p_0$.

\[
\frac{1}{N} L_{2N} = \frac{1}{N} \sum_{n=1}^{N} \ln \left( \frac{p_0(X_n)}{p_2(X_n)} \right) \to \mathbb{E}_{p_0} \left\{ \ln \left( \frac{p_0(X_1)}{p_2(X_1)} \right) \right\} = \mathcal{D}(p_0\|p_2)
\]

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

\[
\Pr_{p_0} \left\{ \left| \frac{1}{N} L_{1N} - \mathcal{D}(p_0\|p_1) \right| > \delta \right\} \to 0 \quad \text{as} \quad N \to \infty,
\]

or for any $\epsilon > 0$, there exist $N_1(\epsilon)$ and $N_2(\epsilon)$ such that

\[
N > N_1(\epsilon) \Rightarrow \Pr_{p_0} \left\{ \left| \frac{1}{N} L_{1N} - \mathcal{D}(p_0\|p_1) \right| > \delta \right\} < \epsilon,
\]

\[
N > N_2(\epsilon) \Rightarrow \Pr_{p_0} \left\{ \left| \frac{1}{N} L_{2N} - \mathcal{D}(p_0\|p_2) \right| > \delta \right\} < \epsilon.
\]

Equation (16) implies that for $N > N_0(\epsilon)$,

\[
\Pr_{p_0} \{ L_{1N} < T \} < \frac{\epsilon}{2},
\]

\[
\Pr_{p_0} \{ L_{2N} > T \} < \frac{\epsilon}{2}.
\]
Now, we have
\[
\Pr_{p_0} \{ L_{1N} < L_{2N} \} = \Pr_{p_0} \{ L_{1N} < L_{2N} \& L_{1N} < T \} + \Pr_{p_0} \{ L_{1N} < L_{2N} \& L_{1N} \geq T \} \\
\leq \Pr_{p_0} \{ L_{1N} < T \} + \Pr_{p_0} \{ L_{2N} > T \} \\
< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
= \epsilon,
\]
for any $N > N_0(\epsilon)$. Therefore $\Pr_{p_0} \{ L_{1N} < L_{2N} \} \to 0$ as $N \to \infty$.

b) In (16), fix $\delta = \mathcal{D}(p_0\|p_2)$. For any $\epsilon > 0$, there exist $N_1(\epsilon)$ and $N_2(\epsilon)$ such that
\[
N > N_1(\epsilon) \Rightarrow \Pr_{p_0} \{ L_{1N} < 0 \} < \Pr_{p_0} \{ L_{1N} < \mathcal{D}(p_0\|p_1) - \mathcal{D}(p_0\|p_2) \} < \epsilon, \\
N > N_2(\epsilon) \Rightarrow \Pr_{p_0} \{ L_{2N} < 0 \} < \epsilon.
\]

Let $N_0(\epsilon) = \max \{ N_1(\epsilon), N_2(\epsilon) \}$. For any $N > N_0(\epsilon)$, we will have
\[
\Pr_{p_0} \{ L_{1N} < 0 | L_{2N} \geq 0 \} = \frac{\Pr_{p_0} \{ L_{1N} < 0 \& L_{2N} \geq 0 \} \leq \Pr_{p_0} \{ L_{1N} < 0 \} < \frac{\epsilon}{1 - \epsilon}.
\]
Therefore, $\Pr_{p_0} \{ L_{1N} < 0 | L_{2N} \geq 0 \} \to 0$, as $N \to \infty$.

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 III, 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
\[
\mathcal{D}(p_i^N\|p_j^N) = N \text{tr} \{ (I_T + S_iS_i^H)(I_T + S_jS_j^H)^{-1} \} - N \ln \det \{ (I_T + S_iS_i^H)(I_T + S_jS_j^H)^{-1} \} - NT.
\]

Adopting the KL distance as performance criterion, the signal set design criterion in general will be maximization of the minimum KL distance between distributions assigned to the signal points, i.e., assuming equiprobable signal points,
\[
\text{maximize } \min \mathcal{D}(p_i\|p_j), \quad \frac{1}{L} \sum_{l=1}^{L} \|S_l\|^2 \leq TP \quad i \neq j
\]
where $\|S_l\|^2 = \sum_{t=1}^{T} \sum_{m=1}^{M} |(S_l)_{tm}|^2$ is the Frobenious norm of $S_l$, or the total power used to transmit $S_l$. 

12
If we denote, by \( \lambda_{i,j}(t), t = 1, \cdots, T \), the \( T \) eigenvalues of \((I_T + S_i S_i^H)(I_T + S_j S_j^H)^{-1}\), the KL distance in (17) can be written as

\[
\mathcal{D}(p_i^N || p_j^N) = N \sum_{t=1}^{T} \{ \lambda_{i,j}(t) - \ln(\lambda_{i,j}(t)) - 1 \}. \tag{19}
\]

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 (17). Since the actual value of \( N \) does not affect the maximization in (18), in designing the signal constellations we will always assume that \( N = 1 \).

V. SIGNAL SET CONSTRUCTION

In [5], it has been shown that the capacity of non-coherent systems can be achieved by an orthogonal random signal matrix constructed as the product \( S = \Phi V \), where \( \Phi \) is an isotropically distributed \( T \times M \) matrix with orthonormal columns, and \( V \) is an independent \( M \times M \) real, non-negative, diagonal matrix. Therefore, in designing matrix constellations for multiple-antenna non-coherent systems, we will confine ourselves to constellations of orthogonal matrices (i.e., matrices with orthogonal columns). With this assumption, the KL distance expression in (17) can be written as

\[
\mathcal{D}(p_i || p_j) = \sum_{m=1}^{M} \left\{ \frac{1 + \| S_{im} \|^2}{1 + \| S_{jm} \|^2} - \ln \left( \frac{1 + \| S_{im} \|^2}{1 + \| S_{jm} \|^2} \right) - 1 + \frac{\| S_{im} \|^2 \| S_{jm} \|^2}{1 + \| S_{jm} \|^2} - \sum_{k=1}^{M} |S_{ik} \cdot S_{jm}|^2 \right\}, \tag{20}
\]

where we have used the notation \( S_{im} \) to denote the \( m \)th column of \( S_i \). The expressions for KL distance in (20) 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 (17) by breaking it down into simpler components. In most cases, this results in a systematic technique for constellation design.
A. $M = 1$ and $T = 1$

This is the case of a single transmit antenna system in fast fading channel, where signal matrices are complex scalars, and the fading coefficients are independent from one symbol to the other. The KL distance of (20) reduces to

$$D_1(p_i\|p_j) = \frac{1 + |s_i|^2}{1 + |s_j|^2} - \ln \left(\frac{1 + |s_i|^2}{1 + |s_j|^2}\right) - 1. \quad (21)$$

It can be easily verified that, similar to the pairwise error probability (12) and the Chernoff information (15), the KL distance is also a monotonic function of $\frac{1 + |s_j|^2}{1 + |s_j|^2}$ 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 maximizing the minimum of the exact pairwise error probability as well as the Chernoff information. The following theorem characterizes the solution to the maximin problem in this case.

**Theorem 3:** The solution to the maximin problem (18) for the case of $M = 1$ and $T = 1$, is given by $|s_i|^2 = \alpha^{L-1} - 1$, where $\alpha$ is the largest real root of the polynomial $\alpha^L - L(P+1)\alpha + (LP+L-1)$.

**Proof:** See Appendix V.

The resulting constellations 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. Figure 2(a) shows the locations of the signal points for a 4-point constellation vs. average transmit power. As we see, spacing between pairs of consecutive points is not constant. However, in terms of the 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 (21), 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 Figure 2(b). However, as we see in this figure, the optimal constellation does not see any error floor.
Fig. 2. (a) Magnitudes of the optimal signal points, and (b) Symbol error rate comparison with regular PAM, for a 4-point constellation with $M = 1, T = 1,$ and $N = 1.$

\[\text{(b)}\]

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

This is the case in which constellation points are column vectors and they all lie on a sphere in $\mathbb{C}^T$. The KL distance of (20) reduces to

\[
D_2(p_i\|p_j) = \frac{\|S_i\|^2\|S_j\|^2 - |S_i \cdot S_j|^2}{1 + \|S_j\|^2} = \frac{(TP)^2 \sin^2(\angle S_i, S_j)}{1 + TP},
\]

(22)

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. Since all of the points have the same magnitude, this case can be considered as a unitary constellation for $M = 1$.

The optimum constellation in this case, is obviously the one that is designed to maximize the minimum angle between subspaces assigned to the signal points (or equivalently, minimizes the maximum inner product or correlation between signal points). Such designs can 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 adja-
cent points is $\pi / L$, 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 $\pi$.

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

This is the general case for single-antenna constellations. The KL distance in (20) reduces to

$$
\mathcal{D}(p_i || p_j) = \frac{1 + \|S_i\|^2}{1 + \|S_j\|^2} - \ln \left( \frac{1 + \|S_i\|^2}{1 + \|S_j\|^2} \right) - 1 + \frac{\|S_i\|^2 \|S_j\|^2 \sin^2(\angle S_i, S_j)}{1 + \|S_j\|^2}
$$

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 on different spheres in $C^T$), and $\frac{\|S_i\|^2 \|S_j\|^2 \sin^2(\angle S_i, S_j)}{1 + \|S_j\|^2} \mathcal{D}_2(p_i || p_j)$ due to the angle between the points (lying on different one-dimensional subspaces of $C^T$). 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 (23) suggests partitioning the signal space into subsets of concentric spheres $C_1, \ldots, C_K$, of radii $r_1, \ldots, r_K$, containing $l_1, \ldots, l_K$ points, respectively, and defining the intrasubset and intersubset distances as

$$
\mathcal{D}_{\text{intra}}(k) = \min_{S_i, S_j \in C_k} \frac{r_i^2 \sin^2(\angle S_i, S_j)}{1 + r_i^2} \quad \text{and} \quad \mathcal{D}_{\text{inter}}(k, k') = \frac{1 + r_k^2}{1 + r_{k'}^2} - \ln \left( \frac{1 + r_k^2}{1 + r_{k'}^2} \right) - 1.
$$

Without loss of generality, we can assume that $r_1 < r_2 < \cdots < r_K$. With this assumption, we can reformulate the design problem as the following suboptimal maximin problem

$$
\frac{1}{L} \sum_{k=1}^{K} l_k r_k^2 \leq TP, \quad \sum_{k=1}^{K} l_k = L
$$

with $r_1 < r_2 < \cdots < r_K$.

The suboptimality of this approach comes from the fact that in the above formulation of $\mathcal{D}_{\text{inter}}(k, k')$, 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 $\mathcal{D}(p_i || p_j) = \mathcal{D}_{\text{inter}}(k, k + 1)$. 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 be larger.
than the minimum in (25). Moreover, assuming that the best unitary constellations (of type in Section V-B) 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).

The above multilevel structure for the non-coherent constellations is somewhat similar to the rectangular lattice structure of the QAM coherent constellations. However, since unlike the square Euclidean distance, the KL distance does not scale with the average power of the constellation, the design problem in this case is more difficult than the case of QAM constellations.

For the special case of real constellations with $T = 2$, the angle between adjacent points in the $k$th subset is simply $\pi/l_k$, and the maximin problem in (25) is fairly easy to solve. The resulting 2, 4, 8 and 16-point constellations with average powers of 0.5 and 5 are shown in Figure 3. 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 $P_{av} = 5$ (SNR $\approx 7$dB) are simulated for different values of $N$ and compared with the corresponding constellations proposed in [8]. The results are shown in Figure 4. 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 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
Fig. 4. Symbol error rate for constellations of size 8 and 16 for $M = 1$, $T = 2$, and SNR = 10 dB.

constellations of [8] have minimum KL distances of 1.3313 and 0.3460, respectively.

The decoding can be done in a similar way to that of trellis coded modulation schemes, i.e., in two phases 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 phase 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.

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

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 (20) reduces to

$$D(p_i || p_j) = \sum_{m=1}^{M} \frac{(TP/M)\|S_{im}\|^2 - \sum_{k=1}^{M} |S_{ik} : S_{jm}|^2}{1 + TP/M} = \frac{TP}{M+TP} \sum_{m=1}^{M} \left\{ \|S_{jm}\|^2 - \sum_{k=1}^{M} \frac{|S_{im} : S_{ik}|^2}{\|S_{ik}\|^2} \right\}$$

$$= \frac{TP}{M+TP} \sum_{m=1}^{M} d_E^2(S_{jm}, W_{S_i}) = \frac{(TP)^2}{M(M+TP)} d_E^2(W_{S_i}, W_{S_j}) ,$$

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 VI, we show that the Euclidean distance defined in [8] and the chordal distance defined in [12] 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 capacity of the non-coherent multiple-antenna channel can also be viewed as sphere packing in the product space of Grassmannian manifolds.

E. $M \geq 1$, $T \geq 1$, and $S_{ik} \cdot S_{jm} = 0$ for all $i$, $j$, and $k \neq m$

In this case, each column of any constellation matrix is orthogonal to all of the other columns of all of the constellation points. If we define the $m$th column space of the constellation as the linear space spanned by the $m$th column of all of the constellation matrices, this constraint means that any two different column spaces of the constellation are orthogonal. With this assumption, the KL distance expression in (20) reduces to

$$D(p_i \| p_j) = \sum_{m=1}^{M} \left\{ \frac{1 + \|S_{im}\|^2}{1 + \|S_{jm}\|^2} - \ln \left( \frac{1 + \|S_{im}\|^2}{1 + \|S_{jm}\|^2} \right) - 1 + \frac{\|S_{im}\|^2 \|S_{jm}\|^2 \sin^2 (\angle S_{im}, S_{jm})}{1 + \|S_{jm}\|^2} \right\}. \quad (27)$$

As we see, the KL distance in this case breaks into summation of $M$ KL distances of the form (23), and each term depends only on one column from each matrix. This means that we can partition the $T$-dimensional space as direct summation of $M$ orthogonal subspaces (which will serve as column spaces of the constellation), and use the method of Section V-C to independently construct single-antenna constellations inside each subspace. The multiple-antenna constellation then can be obtained by forming the Cartesian product of all of these single-antenna constellations.

An example of this approach, is the multiple-antenna extension of the PSK-type constellations of [8]. Assuming that $T = 2M$, in [8] the $T$-dimensional space is suggested to be partitioned into $M$ 2-dimensional subspaces, each one generated by a different pair of elements of the standard basis. Inside each subspace (or equivalently, for each transmit antenna over two symbol periods), a single-antenna PSK-type constellation is used. In [8], instead of considering the Cartesian product of the single-antenna constellations, the same constellation point is repeated on different antennas to achieve the maximum available diversity. However, considering the fact that in this case, $T$ is
times larger than the single-antenna case, this results in a decrease in rate by a factor of $M$. Using the Cartesian product maintains the same rate as the single-antenna case, however it does not provide any performance improvement. The minimum KL distance of the resulting multiple-antenna constellation is the same as the minimum KL distance of the single-antenna constituent constellation. In fact, this approach is equivalent to designing a single-antenna constellation for $T = 2$, and forming uncoded blocks of length $T = 2M$ and transmitting them over the same antenna.

\[ F. \quad M \geq 1, T \geq 1, \text{ and } S_i^H S_l = d_l I_M \text{ for } l = 1, \cdots, L \]

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 (17) reduces to

\[
\mathcal{D}(p_i \| p_j) = M \left[ \frac{1 + d_i}{1 + d_j} - \ln \left( \frac{1 + d_i}{1 + d_j} \right) - 1 \right] + \frac{d_i d_j}{1 + d_j} d_E^2(W_{S_i}, W_{S_j}),
\]

where $d_E^2(W_{S_i}, W_{S_j})$ is the square Euclidean distance [8] or chordal distance [12] between the two subspaces $W_{S_i}$ and $W_{S_j}$ spanned by columns of $S_i$ and $S_j$, defined as

\[
d_E^2(W_{S_i}, W_{S_j}) = \sum_{m=1}^{M} d_E^2 \left( \frac{S_{im}}{\sqrt{d_i}}, W_{S_j} \right) = \sum_{m=1}^{M} \left\{ \frac{\|S_{im}\|^2}{d_i} - \sum_{k=1}^{M} \frac{|S_{im} \cdot S_{jk}|^2}{d_i d_j} \right\},
\]

$\mathcal{D}_1$ 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, $C_1, \cdots, C_K$, with columns of square norm $\rho_1, \cdots, \rho_K$, containing $l_1, \cdots, l_K$ points, respectively.

With the above partitioning, we define the intra-subset and inter-subset distances as

\[
\mathcal{D}_{\text{ intra}}(k) = \min_{S_{i}, S_{j} \in C_k} \frac{\rho_k^2}{1 + \rho_k} d_E^2(S_i, S_j),
\]

and

\[
\mathcal{D}_{\text{ inter}}(k, k') = M \left[ \frac{1 + \rho_k}{1 + \rho_{k'}} - \ln \left( \frac{1 + \rho_k}{1 + \rho_{k'}} \right) - 1 \right].
\]
Without loss of generality, we can assume that $\rho_1 < \rho_2 < \cdots < \rho_K$, and solve the simplified maximin problem
\[
\text{maximize} \quad \min \left\{ \min_{k=1,\ldots,K} D_{\text{intra}}(k), \min_{k=1,\ldots,K-1} D_{\text{inter}}(k, k + 1) \right\}
\]
\[1 \leq K \leq L, \sum_{k=1}^{K} l_k \rho_k = TP, \sum_{k=1}^{K} l_k = L \]
\[0 \leq \rho_1 < \rho_2 < \cdots < \rho_K\]
\[(32)\]
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 $D_{\text{intra}}(k)$ in (30), the best achievable KL distance with that construction and with size $l_k$.

In (32), $K$ and $l_1, \ldots, l_K$ are discrete variables, while $\rho_1, \ldots, \rho_K$ are continuous variables. For any fixed value of $K$ and $l_1, \ldots, l_K$ satisfying the specified constraints, (32) reduces to a continuous optimization over $\rho_1, \ldots, \rho_K$, which can be solved numerically. Even though in (32) we have allowed $K$ to range from 1 to $L$, in practice we do not need to try all of these possible values for $K$. Starting from $K = 1$, and increasing the value of $K$ by one each time, we can stop the search once the optimum minimum distance of the solution stops increasing. Moreover, since the intra-subset distance is an increasing function of $\rho_k$, it can be shown that the optimum constellation also satisfies the extra constraint $l_1 \leq l_2 \leq \cdots \leq l_{K-1}$, which can be used to further restrict the domain of our search.

From (28), we also observe that at high SNR, $D_1$ becomes a constant, or its minimum grows at most logarithmically with SNR (when $d_i = 0$), whereas $D_2$ grows linearly with SNR for non-zero constellation points from different subspaces. As a result, at high SNR, $D_2$ becomes the dominant term, and the KL-based design criterion reduces to the Euclidean-based design criterion of the unitary constellations, confirming the high SNR optimality of the unitary constellations.

Figures 5(a) and 5(b) 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. The comparisons are between these systematic constellations and their multilevel versions. Figure 5(a) 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
Fig. 5. Performance comparison of one and two transmit antenna systematic constellations of [7] and their multilevel versions vs. (a) number of receive antennas, and (b) SNR.

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.

Figure 5(b) 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. 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 distributions assigned to the transmitted symbols. 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, and through simulations, showed that the new constellations 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.

Appendix I

Exact Pairwise Error Probability for the Single Antenna Case

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

\[ X^N = [x_1 \cdots x_N]. \]  

(I.1)

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

\[
\Pr(s_1 \to s_2) = \Pr_{p^N_1} \{ X^N : p_2^N(X^N) > p_1^N(X^N) \} \\
= \Pr_{p^N_1} \{ X^N : \frac{1}{\pi^N (1+|s_2|^2)^N} \exp \left( -\frac{\|X^N\|^2}{1+|s_2|^2} \right) > \frac{1}{\pi^N (1+|s_1|^2)^N} \exp \left( -\frac{\|X^N\|^2}{1+|s_1|^2} \right) \} \\
= \Pr_{p^N_1} \{ X^N : \|X^N\|^2 > NA \},
\]

(I.2)

where \(A\) is as in (10).

Similarly, for \(|s_1| > |s_2|\) we have

\[
\Pr(s_1 \to s_2) = \Pr_{p^N_1} \{ X^N : \|X^N\|^2 < NA \} \\
= 1 - \Pr_{p^N_1} \{ X^N : \|X^N\|^2 > NA \}
\]

(I.3)

(since \(p^N_1\) 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 3: For any \(C \geq 0\), we have

\[
\Pr_{p^N_1} \{ X^N : \|X^N\|^2 > C \} = \sum_{n=0}^{N-1} \frac{1}{n!} \left( \frac{C}{1+|s_1|^2} \right)^n \exp \left( -\frac{C}{1+|s_1|^2} \right). \]  

(I.4)

Proof: The proof is by induction, as follows.

For \(N = 1\), we have

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

(I.5)
which is true, and proven in Proposition 1.

Now assume that (I.4) is true for \( N = K \). We prove that it will also be true for \( N = K + 1 \). Using (8) and the notation defined in (I.1), we can write

\[
p_{i}^{K+1}(X^{K+1}) = p_{i}^{K}(X^{K})p_{i}(x_{K+1}). \tag{I.6}
\]

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

\[
\mathcal{R} = \{X^{K+1} : \|X^{K+1}\|^2 > C\},
\]

\[
\mathcal{R}_1 = \{X^{K+1} : \|X^K\|^2 > C\}, \text{ and }
\]

\[
\mathcal{R}_2 = \{X^{K+1} : \|X^K\|^2 \leq C \quad \& \quad |x_{K+1}|^2 > C - \|X^K\|^2\},
\]

we have

\[
\mathcal{R}_1 \cup \mathcal{R}_2 = \mathcal{R},
\]

\[
\mathcal{R}_1 \cap \mathcal{R}_2 = \emptyset,
\]

and

\[
\text{Pr}_{p_{i}^{K+1}} \{X^{K+1} : \|X^{K+1}\|^2 < C\} = \text{Pr}_{p_{i}^{K+1}}\{\mathcal{R}\} = \text{Pr}_{p_{i}^{K+1}}\{\mathcal{R}_1\} + \text{Pr}_{p_{i}^{K+1}}\{\mathcal{R}_2\}. \tag{I.7}
\]

The first term in (I.7) can be calculated as

\[
\text{Pr}_{p_{i}^{K+1}}\{\mathcal{R}_1\} = \int_{\mathcal{R}_1} p_{i}^{K+1}(X^{K+1})dX^{K+1}
\]

\[
= \int_{\|X^K\|^2 > C} p_{i}^{K}(X^K) \left[ \int_{\mathcal{C}} p_{i}(x_{K+1})dx_{K+1} \right] dX^K \tag{I.8}
\]

\[
= \text{Pr}_{p_{i}^{K}} \{X^K : \|X^K\|^2 > C\}
\]

\[
= \left[ \sum_{n=0}^{K-1} \frac{1}{n!} \left( \frac{C}{1+|s_1|^2} \right)^n \right] \exp \left( \frac{-C}{1+|s_1|^2} \right),
\]

where the last equality follows from the fact that we have assumed (I.4) is true for \( N = K \).

The second term in (I.7) can be calculated as

\[
\text{Pr}_{p_{i}^{K+1}}\{\mathcal{R}_2\} = \int_{\mathcal{R}_2} p_{i}^{K+1}(X^{K+1})dX^{K+1}
\]

\[
= \int_{\|X^K\|^2 \leq C} p_{i}^{K}(X^K) \left[ \int_{|x_{K+1}|^2 > C - \|X^K\|^2} p_{i}(x_{K+1})dx_{K+1} \right] dX^K \tag{I.9}
\]

\[
= \frac{1}{\pi^K (1+|s_1|^2)^K} \left( \frac{C}{1+|s_1|^2} \right)^K \int_{\|X^K\|^2 \leq C} dX^K
\]

\[
= \frac{1}{K!} \left( \frac{C}{1+|s_1|^2} \right)^K \exp \left( \frac{-C}{1+|s_1|^2} \right),
\]

24
where the third equality follows from (I.5) and (8), and the last equality follows from the formula of the volume of a 2K-dimensional sphere with radius $R$,

$$V_{2K}(R) = \frac{\pi^K}{K!} R^{2K}. \quad (I.10)$$

Substituting (I.8) and (I.9) in (I.7) shows that (I.4) is true for $N = K + 1$. This completes the proof.

**APPENDIX II**

**CHERNOFF BOUND FOR THE SINGLE ANTENNA CASE**

It is easy to show that

$$C(p_1^N, p_2^N) = NC(p_1, p_2). \quad (II.1)$$

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

$$C(p_1, p_2) = - \min_{0 \leq \lambda \leq 1} \ln \left[ \mathbb{E}_{p_1} \left\{ \left( \frac{p_2(x)}{p_1(x)} \right)^\lambda \right\} \right]. \quad (II.2)$$

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

$$C(p_1, p_2) = - \min_{0 \leq \lambda \leq 1} \ln \left[ \mathbb{E}_{p_1} \left\{ \left( \frac{\lambda |x|^2}{1 + |x|^2} \right)^\lambda \exp \left( \frac{\lambda |x|^2}{1 + |x|^2} - \frac{\lambda |x|^2}{1 + |s_2|^2} \right) \right\} \right] \quad (II.3)$$

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

$$\mathbb{E}_{p_1} \{ \exp (a|x|^2) \} = \int_\mathbb{C} \frac{1}{\pi(1+|s_1|^2)} \exp \left\{ -\frac{|x|^2}{1+|s_1|^2} + a|x|^2 \right\} \, dx$$

$$= \frac{1}{1-a(1+|s_1|^2)} = \left[ 1 - \lambda + \lambda \left( \frac{1+|s_1|^2}{1+|s_2|^2} \right) \right]^{-1}. \quad (II.4)$$

Substituting (II.4) in (II.3) and using (11), we have

$$C(p_1, p_2) = - \min_{0 \leq \lambda \leq 1} \left\{ \lambda \ln(B) - \ln(1 - \lambda + \lambda B) \right\}. \quad (II.5)$$

Now, taking derivative with respect to $\lambda$ and setting it to zero, we will get

$$\lambda = \frac{1}{\ln(B)} - \frac{1}{B - 1}. \quad (II.6)$$

Substituting this value of $\lambda$ as the minimizer in (II.5) together with (II.1) results in (15).
Appendix III

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 \( l = 1, \cdots, L \), we have

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

(III.1)

since \( X_n \)'s are independent and identically distributed.

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

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

(III.2)

where

\[ K = \left( I_T + S_i S_i^H \right)^{-1} - \left( I_T + S_j S_j^H \right)^{-1}. \]

Again, using (2) for \( p_i \), we have

\[
\mathbb{E}_{p_i} \left\{ X_n^H K X_n \right\} = \text{tr} \left\{ K \left( I_T + S_i S_i^H \right) \right\} = \text{tr} \left\{ I_T - \left( I_T + S_j S_j^H \right)^{-1} \left( I_T + S_i S_i^H \right) \right\} = T - \text{tr} \left\{ \left( I_T + S_i S_i^H \right) \left( I_T + S_j S_j^H \right)^{-1} \right\}. \]

(III.3)

Substituting (III.3) in (III.2), we will have

\[
\mathcal{D}(p_i \| p_j) = \text{tr} \left\{ \left( I_T + S_i S_i^H \right) \left( I_T + S_j S_j^H \right)^{-1} \right\} - \ln \det \left\{ \left( I_T + S_i S_i^H \right) \left( I_T + S_j S_j^H \right)^{-1} \right\} - T.
\]

(III.4)

Equations (III.4) and (III.1) result in (17).
APPENDIX IV

THE SIMPLIFIED KL DISTANCE FOR ORTHOGONAL MATRICES

In this appendix we derive a simplified version of the KL distance in (17) for a constellation of orthogonal matrices, \( \{S_l\}_{l=1}^{L} \) with \( S_l^H S_l = D_l \) for \( l = 1, \ldots, L \). 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 [13]

\[
(A + BCD)^{-1} = A^{-1} - A^{-1}B \left( C^{-1} + DA^{-1}B \right)^{-1} DA^{-1}, \tag{IV.1}
\]

we can write

\[
(I_T + S_l S_l^H) = I_T - S_l (I_M + D_l)^{-1} S_l^H. \tag{IV.2}
\]

Therefore, we will have

\[
(I_T + S_l S_l^H)(I_T + S_j S_j^H)^{-1} = I_T - S_j (I_M + D_j)^{-1} S_j^H + S_i S_i^H S_j (I_M + D_j)^{-1} S_j^H. \tag{IV.3}
\]

We need to calculate the trace and determinant of this matrix, and substitute for them in (17). To find the trace of (IV.3), we calculate the trace of each term separately. We have

\[
\text{tr} \left\{ S_j (I_M + D_j)^{-1} S_j^H \right\} = \text{tr} \left\{ S_j^H S_j (I_M + D_j)^{-1} \right\} = \text{tr} \left\{ D_j (I_M + D_j)^{-1} \right\} = \sum_{m=1}^{M} d_{jm} = \sum_{m=1}^{M} \frac{\|S_{lm}\|^2}{1 + \|S_{jm}\|^2}, \tag{IV.4}
\]

and

\[
\text{tr} \left\{ S_i S_i^H \right\} = \text{tr} \left\{ S_i^H S_i \right\} = \text{tr} \left\{ D_i \right\} = \sum_{m=1}^{M} d_{im} = \sum_{m=1}^{M} \|S_{im}\|^2. \tag{IV.5}
\]

To find the trace of the last term in (IV.3), we use the following identity which can be easily verified

\[
\text{tr} \left\{ AD \right\} = \text{tr} \left\{ \text{diag}(A)D \right\}, \tag{IV.6}
\]

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 which 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

\[
a_{mm} = \sum_{k=1}^{M} S_{jm}^H S_{ik} S_{ik}^H S_{jm} = \sum_{k=1}^{M} \|S_{ik} \cdot S_{jm}\|^2, \tag{IV.7}
\]
where \( a_{mm} \) is the \( m \)th diagonal element of \( A \), and \( \cdot \) is the inner product operation. Using (IV.6) and (IV.7), we will have
\[
\text{tr} \left\{ S_i S_i^H S_j (I_M + D_j)^{-1} S_j^H \right\} = \text{tr} \left\{ S_j S_i S_i^H S_j (I_M + D_j)^{-1} \right\} = \text{tr} \left\{ A (I_M + D_j)^{-1} \right\} = \text{tr} \left\{ \text{diag}(A) (I_M + D_j)^{-1} \right\} = \sum_{m=1}^{M} \frac{\sum_{k=1}^{M} |S_{j_k} S_{j m}|^2}{1 + \|S_{jm}\|^2}.
\]

(Equation IV.8)

Equations (IV.3), (IV.4), (IV.5), and (IV.8) result in
\[
\text{tr} \left\{ (I_T + S_i S_i^H)(I_T + S_j S_j^H)^{-1} \right\} = T - \sum_{m=1}^{M} \left\| S_{j m} \right\|^2 + \sum_{m=1}^{M} \left\| S_{i m} \right\|^2 - \sum_{m=1}^{M} \frac{\sum_{k=1}^{M} |S_{j_k} S_{j m}|^2}{1 + \|S_{jm}\|^2}.
\]

(IV.9)

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 [13]
\[
\det(I + AB) = \det(I + BA),
\]

(IV.10)

to write
\[
\det(I_T + S_i S_i^H) = \det(I_M + S_i^H S_i) = \det(I_M + D_i) = \prod_{m=1}^{M} \left( 1 + d_{im} \right) = \prod_{m=1}^{M} \left( 1 + \|S_{im}\|^2 \right).
\]

(IV.11)

Using (IV.11), we will have
\[
\det\left((I_T + S_i S_i^H)(I_T + S_j S_j^H)^{-1}\right) = \frac{\prod_{m=1}^{M} \left( 1 + \|S_{im}\|^2 \right)}{\prod_{m=1}^{M} \left( 1 + \|S_{jm}\|^2 \right)}.
\]

(IV.12)

Substituting (IV.9) and (IV.12) in (17), results in (20).

APPENDIX V
THE OPTIMAL SINGLE ANTENNA CONSTELLATION FOR FAST FADING

Without loss of generality, let’s assume that \( 0 \leq |s_1| \leq |s_2| \leq \cdots \leq |s_L| \). Since \( f(x) = x - \ln(x) - 1 \) is monotonically decreasing for \( x \in (0, 1) \) and monotonically increasing for \( x \in (1, \infty) \), with \( f(x) < f\left(\frac{1}{x}\right) \) 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 (18), it is sufficient to solve the following minimax problem
\[
\begin{align*}
\text{minimize} & \quad \frac{1}{L} \sum_{l=1}^{L} |s_l|^2 \\
\text{max} & \quad P \quad l = 1, \cdots, L - 1 \quad \frac{1 + |s_l|^2}{1 + |s_{l+1}|^2}.
\end{align*}
\]

(V.1)
Defining
\[
\alpha = \min_{l=1, \ldots, L-1} \frac{1 + |s_{l+1}|^2}{1 + |s_l|^2},
\]
we will have,
\[
1 + |s_{l+1}|^2 \geq \alpha (1 + |s_l|^2) \Rightarrow 1 + |s_l|^2 \geq \alpha^{l-1} (1 + |s_1|^2), \quad l = 1, \ldots, L, \quad (V.2)
\]
or
\[
L + \sum_{l=1}^L |s_l|^2 \geq \left( \sum_{l=1}^L \alpha^{l-1} \right) (1 + |s_1|^2) \Rightarrow \frac{1 - \alpha^L}{1 - \alpha} \leq \frac{L + \sum_{l=1}^L |s_l|^2}{1 + |s_1|^2} \leq \frac{L + LP}{1 + |s_1|^2},
\]
(using the average power constraint). Now, since \( \frac{1 - \alpha^L}{1 - \alpha} \) is a monotonically increasing function of \( \alpha \), it is clear that the maximum of \( \alpha \) 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 (V.2) hold with equality. Therefore, the optimum signal set can be obtained by setting
\[
|s_l|^2 = \alpha^{l-1} - 1,
\]
where \( \alpha \) 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 VI**

**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 [12] as
\[
d_c^2(W_i, W_j) = \sum_{m=0}^M \sin^2 (\angle S_{im}, S_{jm}),
\]
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
\[
(S_{im}, S_{jm}) = \arg \max_{(u, v) \in W_i \times W_j} u \cdot v, \quad \text{for } m = 1, \ldots, M. \quad (VI.2)
\]
\[
\|u\| = \|v\| = 1, u \cdot S_{ik} = v \cdot S_{jk} = 0 \text{ for } k < m
\]

We will use the following lemma to prove the equivalence of the *Euclidean* and *chordal* distances.
Lemma 4: If $W_i$ and $W_j$ are two $M$-dimensional subspaces of $\mathbb{C}^r$, 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

\begin{enumerate}
  \item $\{S_{im}\}_{m=1}^M$ and $\{S_{jm}\}_{m=1}^M$ form orthonormal bases for $W_i$ and $W_j$, respectively, and
  \item $S_{im} \cdot S_{jk} = 0$ for $m \neq k$.
\end{enumerate}

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 $W_m^m = \text{span}(S_{jm}, \cdots, S_{jM})$. By definition,

$$\text{Proj}_{W_m^m}(S_{im}) = (S_{im} \cdot S_{jm})S_{jm}, \quad \text{(VI.3)}$$

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

$$\begin{align*}
(S_{im} - (S_{im} \cdot S_{jm})S_{jm}) \cdot S_{jk} &= 0, \quad \text{for } k \geq m \\
\Rightarrow S_{im} \cdot S_{jk} &= (S_{im} \cdot S_{jm})(S_{jm} \cdot S_{jk}) = 0 \quad \text{for } k > m.
\end{align*} \quad \text{(VI.4)}$$

Similarly,

$$\text{Proj}_{W_i^k}(S_{jk}) = (S_{jk} \cdot S_{ik})S_{ik}, \quad \text{(VI.5)}$$

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

$$\begin{align*}
(S_{jk} - (S_{jk} \cdot S_{ik})S_{ik}) \cdot S_{im} &= 0, \quad \text{for } m \geq k \\
\Rightarrow S_{jk} \cdot S_{im} &= (S_{jk} \cdot S_{ik})(S_{ik} \cdot S_{im}) = 0 \quad \text{for } m > k.
\end{align*} \quad \text{(VI.6)}$$

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

$$d_E^2(W_i, W_j) = \sum_{m=1}^M d_E^2(S_{im}, W_j) = \sum_{m=1}^M \left\{ \|S_{im}\|^2 - \frac{\|S_{im} \cdot S_{jm}\|^2}{\|S_{jm}\|^2} \right\}$$

$$= \sum_{m=1}^M \left\{ 1 - |S_{im} \cdot S_{jm}|^2 \right\} = \sum_{m=1}^M \left\{ 1 - \cos^2(\angle S_{im}, S_{jm}) \right\} \quad \text{(VI.7)}$$

$$= \sum_{m=1}^M \sin^2(\angle S_{im}, S_{jm}) = d_c^2(W_i, W_j).$$
REFERENCES


Constellations for Imperfect Channel State Information at the Receiver *

Mohammad Jaber Borran, Ashutosh Sabharwal, and Behnaam Aazhang
ECE Department, MS-366, Rice University, Houston, TX 77005-1892
Email: {mohammad,ashu,aaz}@rice.edu

Abstract

We consider the problem of digital communication in a Rayleigh flat fading environment using a multiple antenna system. We assume that the transmitter doesn’t know the channel coefficients, and that the receiver has only an estimate of them with some known estimation error. Inspired by the Stein’s lemma, we propose to use the Kullback-Leibler distance between distributions assigned to the transmitted symbols as a performance criterion. Using this performance criterion, we derive a design criterion based on maximizing the minimum KL distance between constellation points. As an example, we design constellations for a single transmit antenna system using the above criterion, and through simulation, show that the new constellations can provide a substantial improvement in the performance over existing and widely-used constellations.

1 Introduction

Exploiting propagation diversity by using multiple antennas at the transmitter and receiver in wireless communication systems has been proposed and studied using different approaches [1–11]. 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 the 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 a large number of transmit antennas are being used by the system, which requires very long training sequences to accurately estimate the fading coefficients. 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 non-coherent systems has been studied in [8, 9], where it has been shown that at high SNR’s, 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

*This work was supported in part by Nokia, TATP under Grant 1999-003604-080, and NSF under Grant ANI-9979465
matrices (i.e. with orthonormal columns). The decoding complexity of optimal unitary constellations is linear in the size of the constellation (exponential in rate). Some low-complexity unitary designs are given in [10, 11]. Also a differential detection scheme for transmit diversity is studied in [7]. In [12], based on the Kullback-Leibler (KL) [13] distance between conditional distributions assigned to the signal points, a new design criterion for non-coherent space-time constellations is derived. The resulting constellations coincide with the unitary designs at very high SNR’s or very low rates, but for high rate codes or at low SNR’s, the KL-based criterion results in different signal sets which show better probability of error performance.

In practice, even if the fading is not very fast, due to the finite length of the training sequence, there will always be some errors in the channel estimates. In order to maintain a given data rate, one would need shorter training sequences for more rapidly fading channels, resulting in even less reliable channel estimates. Having multiple transmit antennas adds up to this problem by requiring longer training sequences for the same estimation performance. Therefore, the usual assumption of known channel parameters at the receiver in designing optimal codes/constellations is not exactly valid in practice. In the presence of channel estimation errors, constellations which are designed using the statistics of the estimation error are more desirable than the ones designed for perfect channel state information at the receiver. In [6], the design criteria of the space-time codes in the absence of perfect channel state information are studied, and it is shown that for constellations with constant energy, the design criteria of the space-time codes remain valid in the presence of channel estimation errors. In [14], the capacity and shaping gain of the partially coherent communication in AWGN channel is studied, and it is shown that the capacity achieving distribution in the presence of phase estimation errors is not Gaussian. In this paper, using a similar approach to [12], we derive a design criterion for space-time constellations when only partial (imperfect) channel state information is available at the receiver.

In Section 2, we introduce the model for the system being considered throughout this paper. In Section 3, we derive the KL distance between distributions assigned to transmitted signals, and propose the design criterion based on that. In Section 4, 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 simulation results that show significant improvement in the performance by using the new constellations instead of the existing and widely-used constellations. Finally, we draw some conclusions in Section 5.

## 2 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

$$X = SH + W,$$  \hspace{1cm} (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 $CN(0, 1)$. 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 \( \hat{H} \) has i.i.d. elements from \( \mathcal{C}\mathcal{N}(0, \sigma^2_E) \), and is statistically independent from \( \hat{H} \) (this can be obtained, e.g., by using an LMMSE estimator).

With the above assumptions, the conditional probability density of the received signal can be written as

\[
p(X \mid S, \hat{H}) = \mathbb{E}_{\hat{H}} \left\{ p(X \mid S, \hat{H}, \tilde{H}) \right\} = \exp\left\{ -\text{tr}\left[ (I_T + \sigma^2_E S S^H)^{-1} (X - S\hat{H})(X - S\hat{H})^H \right] \right\} \frac{\pi^{TN}}{\det(I_T + \sigma^2_E S S^H)}.
\]  

(2)

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

\[
\hat{S}_{ML} = S_{\text{arg max} L_i \in \{1, \ldots, L\} p_l(X)}.
\]  

(3)

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

\[
\Pr(S_1 \rightarrow S_2) = \Pr\{p_2(X) > p_1(X) \mid S_1\}.
\]  

(4)

For \( L > 2 \), even though (4) is no longer exact, we still use it as an approximation for the pairwise error probability.

### 3 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 (4) over the signal set. Therefore, like most of the other constellation/code design techniques, we use the maximum of (4) 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 (4) in general seems to be intractable. Therefore, inspired by the Stein’s lemma [13], we 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.

Using (2), the KL distance between \( p_i \) and \( p_j \) can be calculated as

\[
\mathcal{D}(p_i \| p_j) = N\text{tr}\left\{ (I_T + \sigma^2_E S_i S_i^H) (I_T + \sigma^2_E S_j S_j^H)^{-1} \right\} - NT
\]

\[
- N \ln \det\left\{ (I_T + \sigma^2_E S_i S_i^H) (I_T + \sigma^2_E S_j S_j^H)^{-1} \right\}
\]

\[+ N \ln \det\left\{ (I_T + \sigma^2_E S_j S_j^H)^{-1} (S_i - S_j) \hat{H} \hat{H}^H (S_i - S_j)^H \right\}.
\]

(5)

Using (5) as the error exponent and taking expectation with respect to \( \hat{H} \), we find the following expected KL distance between signal points \( S_i \) and \( S_j \),

\[
\overline{\mathcal{D}}(S_i \| S_j) = N\text{tr}\left\{ (I_T + \sigma^2_E S_i S_i^H) (I_T + \sigma^2_E S_j S_j^H)^{-1} \right\} - NT
\]

\[- N \ln \det\left\{ (I_T + \sigma^2_E S_i S_i^H) (I_T + \sigma^2_E S_j S_j^H)^{-1} \right\}
\]

\[+ N \ln \det\left\{ (I_T + \sigma^2_E S_j S_j^H)^{-1} (S_i - S_j) (I_T + \sigma^2_E S_j S_j^H)^{-1} (S_i - S_j)^H \right\}.
\]

(6)
We emphasize that (6) is obtained by taking the expectation of the error bound using (5) as the exponent, not by taking the expectation of (5). It is interesting to notice that in the two extreme cases of \( \sigma^2_E = 0 \) and \( \sigma^2_E = 1 \), (6) reduces to the existing performance criteria for coherent and non-coherent space-time codes. For \( \sigma^2_E = 0 \) (perfect channel state information at the receiver, i.e., coherent communication), (6) reduces to

\[
\mathcal{D}(S_i \| S_j) = N \ln \det \left( I_M + (S_i - S_j)^H(S_i - S_j) \right),
\]

which is exactly the same performance criterion given in [4] for coherent space-time codes, and results in the rank and determinant design criteria. For \( \sigma^2_E = 1 \) (no channel state information at the receiver, i.e., non-coherent communication), (6) reduces to

\[
\mathcal{D}(S_i \| S_j) = N \text{tr} \left\{ (I_T + S_i S_i^H)(I_T + S_j S_j^H)^{-1} \right\} - NT - N \ln \det \left( (I_T + S_i S_i^H)(I_T + S_j S_j^H)^{-1} \right),
\]

which is exactly the same performance criterion given in [12] for non-coherent space-time codes. For the intermediate values of \( \sigma^2_E \), 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

\[
\max \frac{1}{T} \sum_{t=1}^{T} \| S_i \|^2 = TP \quad i \neq j \\
\min \mathcal{D}(S_i \| S_j),
\]

where \( \| S_i \|^2 = \sum_{t=1}^{T} \sum_{m=1}^{M} |(S_i)_{tm}|^2 \) is the total power used to transmit \( S_i \). Since the actual value of \( N \) does not affect the maximization in (9), in designing the optimal signal sets we can always assume \( N = 1 \).

4 Single Antenna Example

In order to demonstrate the new design technique and the effect of channel estimation error in the structure of resulting constellations, we 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 \( S_i \) is simply a complex scalar. The expression for the expected KL distance in (6) reduces to

\[
\mathcal{D}_1(s_i|s_j) = \frac{1 + \sigma_E^2 |s_i|^2}{1 + \sigma_E^2 |s_j|^2} - 1 - \ln \left( \frac{1 + \sigma_E^2 |s_i|^2}{1 + \sigma_E^2 |s_j|^2} \right) + \ln \left[ 1 + \left( 1 - \sigma_E^2 \right) \frac{|s_i - s_j|^2}{1 + \sigma_E^2 |s_j|^2} \right]. \tag{10}
\]

Using the same idea of multilevel unitary (circular, in this case) constellations of [12], 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 radiuses, and the number of constellation points on each circle. It can be shown that the actual minimum KL distance of the resulting constellations will be greater than or equal to the one guaranteed by this approach, whereas the number of the parameters of the simplified optimization problem is much smaller than the complete problem.

The resulting 8- and 16-point constellations with average power of 10 and for different values of \( \sigma_E^2 \) are shown in Figure 1. The symbol error rate performance of the above constellations at \( \sigma_E^2 = 0.5 \) are simulated for different values of \( N \) and compared with the commonly-used 8PSK and 16QAM constellations. The results are shown in Figure 2, where by coherent we refer to the multilevel circular constellations designed for \( \sigma_E^2 = 0 \). As expected, due to the larger minimum KL distance of the new constellations, the exponential decay of the symbol error rate vs. \( N \) is much higher for the new constellations. It is also interesting to notice that at \( \sigma_E^2 = 0.5 \), 16QAM constellation has better performance compared to the multilevel circular constellation designed for coherent communication. The reason is that 16QAM, if considered as a multilevel circular constellation, is in fact a three level constellation as compared to the coherent constellation which has only two levels. This is not the case for the 8-point constellations, where the coherent constellation (with two levels) still performs better, at \( \sigma_E^2 = 0.5 \), than 8PSK (with only one level).

5 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 distributions assigned to the transmitted symbols. Through a single-antenna example, we demonstrated the effect of the channel estimation error in the shape and structure of the optimal constellations. The new constellations are shown to provide significant improvement in the performance as compared to the existing, commonly-used constellations.

References


EM-Based Multiuser Detection in Fast Fading Multipath Environments

Mohammad Jaber Borran and Behnaam Aazhang
Department of Electrical and Computer Engineering
Rice University
6100 Main St., MS-366
Houston, TX 77251-1892
Email: {mohammad,aaz}@rice.edu

Abstract
We address the problem of multiuser detection in fast fading multipath environments for DS-CDMA systems. In fast fading scenarios, temporal variations of the channel cause significant performance degradation even with the RAKE receiver. We use a previously introduced Time-Frequency (TF) RAKE receiver based on a canonical formulation of the channel and signals to simultaneously combat fading and multipath effects. This receiver uses the Doppler spread caused by rapid time-varying channel as another means of diversity. In dealing with multiaccess interference and as an attempt to avoid the prohibitive computational complexity of the optimum Maximum-Likelihood (ML) detector, we use the Expectation Maximization (EM) algorithm to derive an approximate ML detector. The new detector turns out to have an iterative structure very similar to the well-known multistage detector with some extra parameters. At the two extreme values of these parameters, the EM detector reduces to either one-shot TF RAKE or generalized multistage detector. For the intermediate values of the parameters, it combines the two estimates to obtain a better decision for the bits of the users. Because of using the EM algorithm, this detector has better convergence properties than the multistage detector; the bit estimates always converge, and if an appropriate initial vector is used, they converge to the global maximizer of the likelihood function. As a result, the new detector provides significantly improved performance while maintaining the low complexity of the multistage detector. Our simulation results confirm the expected performance improvements compared to the base case of the TF RAKE as well as the multistage detector used with the TF RAKE.

Keywords
CDMA Systems, Multiuser Detection, EM Algorithm, Multipath-Doppler Diversity, Time-Frequency RAKE

I. INTRODUCTION

Multipath, fading, and multiple-access interference are the major factors that limit the performance of the existing mobile wireless communication systems. Fading of the received signal caused by wireless channels, coupled with the interference from other transmitters using the same channel, significantly degrades the performance of the receiver.

Wideband Code Division Multiple Access (W-CDMA), the accepted technology for next generation cellular networks, provides intrinsic protection against the multipath effects of the channel. A RAKE receiver structure is used to exploit the large time-resolution of the wideband signal and capture the information in its multipath components.

In fast-fading scenarios, temporal variations of the channel cause significant performance degradation even with the RAKE receiver. The Doppler spread caused by rapid time-varying channel can be used as another means of diversity in such environments. Joint multipath-Doppler diversity schemes [1, 2, 3], use a canonical representation of the channel and signals to capture the multipath-Doppler components of the signal.

In multiple-access environments, the minimum probability of error reception can be achieved by a Maximum Likelihood (ML) receiver [4]. Even though this optimal receiver shows significant performance gains over the conventional detector, its computational complexity which grows exponentially with the number of users, prohibits its practical implementation. Therefore, some practical sub-optimum detectors have been introduced for multiuser detection [5-21].

Lupas and Verdu [5] describe a family of linear detectors called decorrelator. These detectors eliminate multiuser interference at the expense of increased noise power. Furthermore, the linear decorrelating detectors require the correlation matrix inversion which may be difficult to perform in real time, especially for asynchronous systems. Some suboptimal approaches have been taken to implement the decorrelating detector for asynchronous systems [6, 7, 8, 9]. The most important advantage of the decorrelating detector is that it does not require the estimation of the received amplitudes.

Madhow et al. [10] and Xie et al. [6] describe a Minimum Mean-Squared Error (MMSE) linear detector which minimizes the mean-squared error between the actual data and the conventional detector soft outputs. Because of taking the background noise into account, the MMSE detector generally performs better than the decorrelating detector, and converges to the decorrelating detector as the background noise goes to zero.
Duel-Hallen in [11] presents a nonlinear multiuser detector called decorrelating decision-feedback detector (DDFD) in which the users are ranked according to their signal strengths from the strongest one to the weakest one. This detector is based on a white noise channel model whose noise-whitening filter is obtained by the Cholesky decomposition of the cross-correlation matrix. The detector performs successive interference cancellation at the output of the noise-whitening filter using past decisions. For the strongest user, this detector performs similar to the decorrelator, but as the user’s power decreases compared to the power of interferers, the detector outperforms the decorrelator and its performance approaches the single user bound. However, its important difficulty is the need for computing the Cholesky decomposition. Other successive interference cancellation detectors are described in [12, 13].

In [14, 15], Varanasi and Aazhang describe a parallel interference cancellation detector called multistage detector in which the tentative decisions obtained from the previous stage are used to estimate and subtract the multiuser interference. The first stage decisions are usually obtained from the conventional detector. This detector, like the DDFD of [11], outperforms the decorrelator when interfering users are stronger than the user under consideration, but its performance degrades as the energies of the interfering users decrease.

The Expectation Maximization (EM) algorithm has also been applied for multiple-access interference suppression in CDMA systems [16, 17, 18, 19], as well as for channel estimation [20, 21, 22]. In [16, 19], an iterative interference cancellation method in Additive White Gaussian Noise (AWGN) channels based on the EM algorithm is proposed. Since the likelihood function is bounded above, and since the EM estimates monotonically increase in likelihood, the suggested receiver is convergent. Also, because of taking into account the previous decision about the data symbol of each user in making new decision for that user, this detector outperforms the parallel interference cancellation detector of [15] for strong users, while having similar performance for the other users.

In [17], Nelson and Poor propose some other iterative multiuser receivers for CDMA systems, based on the EM algorithm and its generalized versions, such as Space Alternating Generalized EM (SAGE), and missing-parameter space-alternating algorithm. The suggested multiuser detectors have structures similar to the parallel interference cancellation method of [14], except that updates of the estimates are made sequentially, rather than in parallel. For the same reason mentioned above, these algorithms are also convergent. The MPEM receiver suggested in this paper has a computational complexity which is proportional to the square of the number of the users, whereas the computational complexity of the original parallel interference cancellation method grows only linearly with the number of users.

In [18], the EM algorithm is applied to maximize the likelihood function over a non-discrete set. The discrete sequence is obtained by quantizing the sequence iterate at convergence. Since the non-discrete maximization problem has a closed form solution, namely the decorrelator, the performance of this scheme is expected to be upper bounded by the performance of the decorrelating receiver. However, depending on the number of the iterations used, the computational complexity of this scheme might be lower. The proposed receiver also iterates between path component estimation and maximal-ratio combining to refine the non-discrete sequence estimate.

In this paper, we first review the canonical representation of the signal and channel in fast fading multipath environments [1, 3]. Then, in Section III, we review some of the multiuser detection techniques in fast fading channels using this representation. These include the optimal (Minimum Probability of Error) and the linear suboptimal decorrelating and MMSE receivers, rederived in [23] for the Time-Frequency (TF) RAKE, as well as a generalization of the multistage detector of [15].

As mentioned earlier, we intend to use the EM algorithm to find an iterative approximate ML solution for the multiuser detection problem. For this, we first, in Section IV, review the EM algorithm, and then, in Section V, in a similar way to [16], derive the new detection scheme for fast fading multipath environments with canonical representation. The proposed detector uses the two-dimensional Time-Frequency RAKE receiver [1, 3] to combat the fading and multipath effects. The simulation results are reported in Section VI, and show the superior performance of the proposed detector compared to the original TF RAKE, as well as the generalized multistage detector. Finally, Section VII contains the conclusions.

II. Canonical Time-Frequency Representation of the Signals and Channel

The Time-Frequency canonical representation [1, 3] exploits the multipath and Doppler effects for obtaining diversity and results in a two-dimensional RAKE receiver which extracts Doppler components in addition to multipath components. This representation reduces the channel to a set of independent channels for the different time-delayed frequency-shifted versions of the signal for each user. Figure 1 illustrates the locations of canonical coordinates in the “time delay”-“Doppler shift” plane, used for time-frequency representation of the channel.

In a multiuser system, the received signal is a superposition of the signals of different users and noise. In this work, we consider a synchronous CDMA system in which the signature sequences of different users are aligned in time.
With this assumption, if the delay spread of the channel is much smaller than the symbol interval, we can ignore the correlation terms between the symbols of different users in adjacent time intervals, and use a one-shot detector for estimating the data bits of different users, as in [23]. Therefore we can restrict ourselves to only the first time interval and assume that the received signal is as follows

\[
r(t) = \sum_{k=1}^{K} b_k x_k(t) + n(t) \quad \text{for } 0 \leq t \leq T_s,
\]

where \( K \) is the number of users, \( b_k \) denotes the data bit of the \( k \)-th user, \( n(t) \) is a white Gaussian noise with zero mean and variance \( \sigma^2 \), \( T_s \) is the symbol interval, and

\[
x_k(t) = \int_0^{T_m} h_k(t, \tau) s_k(t - \tau) d\tau \quad \text{for } k = 1, 2, \cdots, K.
\]

In this equation, \( s_k(t) \) and \( h_k(t, \tau) \) are, respectively, the signature signal and the time-varying channel impulse response for the \( k \)-th user, and \( T_m \) denotes the multipath (delay) spread of the channel.

An equivalent representation for the signal \( x_k(t) \) in terms of the channel spreading function \( H_k(\theta, \tau) \) [24] (i.e., the Fourier transform of \( h_k(t, \tau) \) with respect to \( \tau \)), is

\[
x_k(t) = \int_0^{T_m} \int_{-B_d}^{B_d} H_k(\theta, \tau) e^{j2\pi\theta \tau} s_k(t - \tau) d\theta d\tau,
\]

where \( \theta \) corresponds to Doppler shifts introduced by the channel and \( B_d \) denotes the Doppler spread of the channel. We use the Wide-Sense Stationary Uncorrelated Scatterer (WSSUS) [24] model for the channel, which assumes that \( H(\theta, \tau) \) is a two-dimensional uncorrelated Gaussian process.

For a spread spectrum signal \( s(t) \) of duration \( T_s \) and chip interval \( T_c \), and with the WSSUS assumption for the channel, using the canonical coordinates \([1, 3]\), we can rewrite the signal \( x_k(t) \) as

\[
x_k(t) \approx \sum_{l=0}^{L} \sum_{m=-M}^{M} H_{k}^{ml} s_{k}^{ml}(t) \quad \text{for } 0 \leq t \leq T_s,
\]

where

\[
s_{k}^{ml}(t) = s_k(t - lT_c)e^{j2\pi ml/T_s} \quad \text{for } l = 0, 1, \cdots, L, m = -M, -M + 1, \cdots, M,
\]

\[
H_{k}^{ml} = \frac{T_c}{T_s} \tilde{H}_k(\frac{m}{T_c}, lT_c) \quad \text{for } k = 1, 2, \cdots, K.
\]
with the number of multipath components \( L = \lfloor T_m / T_c \rfloor \), and the number of Doppler components \( M = \lfloor B_d T_a \rfloor \).

Here, \( \hat{H}_k(\theta, \tau) \) is the time-frequency smoothed version of \( H(\theta, \tau) \) \([23]\) given by the following expression:

\[
\hat{H}_k(\theta, t) = \frac{T_s}{T_c} \int_0^{T_m} \int_{-B_d}^{B_d} H_k(\theta, \tau) e^{-j \pi (\theta - \theta') T_s} \text{sinc} \left((\theta - \theta') T_c\right) \text{sinc} \left((\tau - \tau') / T_c\right) d\theta' d\tau'.
\]  

(6)

In order to simplify the mathematical expressions, we use the following vector notation for the time-delayed and frequency-shifted versions of the signature waveforms of the users,

\[
s(t) = [s_1(t)^T s_2(t)^T \cdots s_K(t)^T]^T,
\]

where

\[
s_k(t) = [s_k^{M_0}(-M+1)^0(t) \cdots s_k^{M_0}(-M+1)L(t) \cdots s_k^{M_0}(t) \cdots s_k^{ML}(t)]^T
\]

for \( k = 1, 2, \ldots, K \). Using this representation, the \( K(L+1)(2M+1) \times K(L+1)(2M+1) \) cross-correlation matrix of the components of the signature waveforms of different users is

\[
R = \int_0^{T_s} s^*(t) s(t)^T dt = \begin{bmatrix}
R_{11} & R_{12} & \cdots & R_{1K} \\
R_{21} & R_{22} & \cdots & R_{2K} \\
\vdots & \vdots & \ddots & \vdots \\
R_{K1} & R_{K2} & \cdots & R_{KK}
\end{bmatrix},
\]

(7)

where

\[
R_{kl} = \int_0^{T_s} s_k^*(t) s_l(t)^T dt \quad \text{for } k, l = 1, 2, \ldots, K.
\]

We also define the channel matrix \( H \) as

\[
H = \begin{bmatrix}
h_1 & 0 & \cdots & 0 \\
0 & h_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & h_K
\end{bmatrix},
\]

where

\[
h_k = \begin{bmatrix}
H_k^{M_0} & \cdots & H_k^{ML} & H_k^{M_0} & \cdots & H_k^{ML}
\end{bmatrix}^T
\]

for \( k = 1, 2, \ldots, K \).

Using the above notations, (4) and (1) can be rewritten as

\[
z_k(t) \approx s_k(t)^T h_k \quad \text{for } 0 \leq t \leq T_s,
\]

(9)

and

\[
r(t) \approx s(t)^T H b + n(t) \quad \text{for } 0 \leq t \leq T_s,
\]

(10)

where \( b = [b_1 \ b_2 \ \cdots \ b_K]^T \).

In the next section, we will see that the outputs of the Time-Frequency RAKE receiver, given as

\[
z_k = \int_0^{T_s} s_k^*(t) r(t) dt \quad \text{for } k = 1, 2, \ldots, K,
\]

(11)

form a set of sufficient statistics for Maximum Likelihood multiuser detection. We collect all of these vectors in one vector \( z = [z_1^T z_2^T \cdots z_K^T]^T \). Using (10) and (7), it can be easily shown that

\[
z = \int_0^{T_s} s^*(t) r(t) dt \approx R H b + w,
\]

(12)

where

\[
w = \int_0^{T_s} s^*(t) n(t) dt
\]

(13)

is a zero-mean complex Gaussian noise vector with \( \mathbb{E}[w w^H] = \sigma^2 R \).
III. REVIEW OF SOME MULTIUSER DETECTION SCHEMES

In this section, we review the optimal and linear suboptimal multiuser detectors rederived in [23] for fast fading channels. We also consider the generalization of the well-known multistage detector to fast fading channels using the TF RAKE.

A. Conventional Single-User Receiver

The single-user receiver assumes that there is no multiaccess interference, i.e., either there are no interfering users, or the signature codes of all of the users and their shifted versions are orthogonal. It can be easily shown ([2, 1]) that, in this case, the TF RAKE receiver with Maximal Ratio Combining (MRC), given by the following expression, is the optimal (i.e., minimum probability of error) receiver.

\[ \hat{b}_k = \text{sgn} \left\{ R[h_k^H z_k] \right\} \quad \text{for } k = 1, 2, \ldots, K. \]  

(14)

This receiver coherently combines the different multipath-Doppler shifted components of the signal to achieve a diversity of order \( (L + 1)(2M + 1) \). Of course, it is assumed that the receiver has complete Channel State Information (CSI). In practice, channel coefficients, \( H_k^T \), may be estimated through a pilot signal transmission.

In the presence of multiaccess interference, i.e., when the signature codes of the interfering users are not completely orthogonal, the above receiver is no longer optimal, and doesn’t show acceptable performance. The optimal multiuser detector is discussed in the next subsection, and has a much more computational complexity.

B. Minimum Probability of Error Receiver

Initially introduced by Verdu [4], the Maximum Likelihood (ML) multiuser receiver achieves the minimum probability of error and is optimal in this sense. For the problem under consideration, the log-likelihood function of the received signal (1) can be written as

\[ \log f_R(r; \mathbf{b}) = \frac{1}{2\sigma^2} \int_0^{T_s} \left| r(t) - \sum_{k=1}^{K} b_k \bar{z}_k(t) \right|^2 dt, \]  

(15)

where \( A \) is a constant. The ML receiver finds the vector \( \hat{b}_{\text{opt}} = [\hat{b}_1 \hat{b}_2 \cdots \hat{b}_K]^T \), such that the above log-likelihood function is maximized for \( \mathbf{b} = \hat{b}_{\text{opt}} \).

Ignoring the constant terms and the terms which do not depend on the unknown bits of the users, and using (7–12), we define the simplified log-likelihood function as

\[ \Lambda(r; \mathbf{b}) = \sum_{k=1}^{K} 2 \Re \{ h_k^H z_k \} b_k - \sum_{k=1}^{K} \sum_{l=1}^{K} b_k h_k^H \mathbf{R}_{kk} h_l b_l = 2 \Re \{ \mathbf{b}^T \mathbf{H}^H z \} - \mathbf{b}^T \mathbf{H}^H \mathbf{R} \mathbf{H} \mathbf{b}. \]  

(16)

Therefore, the decision rule for the ML receiver can be written as

\[ \hat{b}_{\text{opt}} = \arg \max_{\mathbf{b} \in \{-1, 1\}^K} \Lambda(r; \mathbf{b}). \]  

(17)

We observe that the outputs of the TF RAKE, \( z_k \) for \( k = 1, 2, \ldots, K \), form a set of sufficient statistics for the detection problem. We also observe that, still, maximal ratio combining of the outputs of the TF RAKE is necessary, though not sufficient.

The above maximization is a \( K \)-dimensional discrete optimization problem and requires a search over \( 2^K \) possibilities. As a result, the computational complexity of the receiver increases exponentially with the number of users, which makes its real-time implementation prohibitive for large number of users. Therefore several suboptimal approaches have been proposed. In the next few subsections, we will review some of these suboptimal receivers. Later, in Section V, we will introduce a new detection scheme which iteratively solves the above optimization problem, and even with a few number of stages, shows better performance compared to the existing schemes with similar complexity.
C. Linear Suboptimal Multiuser Receivers

Having established that $z = [z_1, z_2, \cdots, z_K]$ is a sufficient statistic for the detection problem, we can try other low complexity processings of this vector to obtain some suboptimal receivers. The approach is motivated by the fact that, in the absense of multiaccess interference, i.e., when the noise free output of the correlators for the $k$th user is equal to $h_k b_k$, the maximal ratio combining is optimal. Therefore, we first try to find a reliable estimate for the vectors $h_k b_k$ for $k = 1, 2, \cdots, K$, given the observation $z$, and then, to coherently combine them to obtain the bit estimate for each user. In [23], based on the above idea, the well-known decorrelating and MMSE receivers are rederived for the TF RAKE. Since the noise vector at the outputs of these linear processings is correlated, a whitening operation is performed before maximal ratio combining of these outputs.

If the linear operation involved in the linear detector is performed using a matrix $F$, the general form of the overall linear multiuser TF RAKE receiver will be

$$\hat{b} = \text{sgn} \{ \Re [H^H D F z] \}, \quad \text{(18)}$$

where $D$ is a block diagonal whitening matrix. The entries of this matrix depend on the type of the linear processing, i.e., the matrix $F$, as well as the correlation matrix of the signature codes, $R$.

$$D = \begin{bmatrix}
Q_{11}^{-1} & 0 & \cdots & 0 \\
0 & Q_{22}^{-1} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & Q_{KK}^{-1}
\end{bmatrix}, \quad \text{(19)}$$

where

$$Q = \mathbb{E}[F w w^H F^H] = \sigma^2 F R F^H = \begin{bmatrix}
Q_{11} & Q_{12} & \cdots & Q_{1K} \\
Q_{21} & Q_{22} & \cdots & Q_{2K} \\
\vdots & \vdots & \ddots & \vdots \\
Q_{K1} & Q_{K2} & \cdots & Q_{KK}
\end{bmatrix}. \quad \text{(20)}$$

In the next two subsections, we will consider two special cases of the above generic linear detector, called decorrelating and linear MMSE receivers.

C.1 Decorrelating Receiver

From the likelihood function (16), it is easy to show that the maximum likelihood estimate for $u = Hb$ is given by

$$\hat{u}_{ML} = \arg \max_u \{2 \Re [u^H z] - u^H R u \} = R^{-1} z. \quad \text{(21)}$$

Therefore, from (18) by letting $F = R^{-1}$, a generalization of the decorrelating receiver of [5] can be obtained

$$\hat{b}_{\text{dec}} = \text{sgn} \{ \Re [H^H D_{\text{dec}} R^{-1} z] \}, \quad \text{(22)}$$

where $D_{\text{dec}}$ is defined as in (19), with $Q = Q_{ML} = \sigma^2 R^{-1}$. This detector eliminates multiuser interference at the expense of increasing the noise power. It also requires the correlation matrix inversion which may be difficult to perform in real time.

C.2 Linear MMSE Receiver

A generalization of the linear MMSE multiuser detector of [10, 6] results from employing a linear MMSE estimate for $u = Hb$. It is shown in [23] that the corresponding linear operation, $F$, for this detector is given by

$$F_{\text{MMSE}} = \arg \min_F \mathbb{E} \| H b - F z \|^2 = (R + \sigma^2 \Psi^{-1})^{-1}, \quad \text{(23)}$$

where $\Psi = \mathbb{E}[H H^H]$. The resulting linear MMSE TF RAKE receiver is given by

$$\hat{b}_{\text{MMSE}} = \text{sgn} \{ \Re [H^H D_{\text{MMSE}} (R + \sigma^2 \Psi^{-1})^{-1} z] \}, \quad \text{(24)}$$

where $D_{\text{MMSE}}$ is defined as in (19), with $Q = Q_{\text{MMSE}} = \sigma^2 (R + \sigma^2 \Psi^{-1})^{-1} R (R + \sigma^2 \Psi^{-1})^{-1}$ (for a WSSUS channel, $\Psi$ is a real diagonal matrix [23]).

Because of taking the background noise into account, this detector generally performs better than the decorrelating detector. However, like the decorrelating detector, it requires a correlation matrix inversion which may be difficult to perform in real time.
D. Generalized Multistage Receiver

In [14], Varanasi and Aazhang describe a parallel interference cancellation detector called multistage detector, which attempts to iteratively maximize the likelihood function. At each stage, the bit estimate for each user is obtained by maximizing the likelihood function over the possible values of the data bit of that user and by using the bit estimates from the previous stage for all other users. From the likelihood function (16), it is easy to show that for the system with TF RAKE, the \((n+1)\)st stage estimate of the data bit of the \(k\)th user, using this multistage detector will be given by the following expression

\[
\hat{b}_{k}^{(n+1)} = \arg \max_{b_k \in \{-1, 1\}} \Lambda(r; b) = \text{sgn} \left\{ \Re \left[ h_k^H z_k - \sum_{j=1, j \neq k}^{K} \hat{b}_{j}^{(n)} h_j^H R_{kj} h_j \right] \right\}, \tag{25}
\]

As it can be seen from the above expression, the tentative decisions obtained from the previous stage are used to estimate and subtract the multiuser interference. The first stage decisions are usually obtained from the conventional detector, which will be given by

\[
\hat{b}_{k}^{(0)} = \text{sgn} \{\Re[h_k^H z_k]\}, \tag{26}
\]

if the TF RAKE is used. This detector outperforms the decorrelating detector when the interfering users are stronger than the user under consideration, but its performance degrades as the energies of the interfering users decrease. In this case, i.e., when the interfering users are not much stronger than the user under consideration, because of the enormous errors in the estimate of the interference, the performance of the multistage detector can be even worse than the conventional detector, and using more stages may only result in even more degraded performance. Examples of this situation are given in Figures 5 and 3 and discussed in Section VI.

In general, there is no guarantee that the multistage detector will converge, or in convergence, if at all, will produce the global maximizer of the likelihood function. However, its lower computational complexity, which is a result of its iterative nature, is a motivation to look for other iterative methods for maximizing the likelihood function which have better convergence properties. The EM algorithm is one of these methods, and will be reviewed in the next section.

IV. EM Algorithm

Expectation Maximization algorithm is an iterative method for maximizing log-likelihood functions. The original problem is formulated as the following optimization problem

\[
\max_{b} \log f_R(r; b), \tag{27}
\]

where \(r\) is the observed data. The vector \(b\) can be any set of parameters. In the problem under consideration, it is the vector of unknown data bits of different users. This is a \(K\)-dimensional discrete optimization problem whose real-time implementation is prohibitive because of exponential complexity in \(K\) (number of users). To construct an iterative sub-optimal solution for this problem, a set of complete data, \(\mathbf{y}\), is defined such that

\[
r = g(y_1, y_2, \cdots, y_K) = g(\mathbf{y}), \tag{28}
\]

where \(g\) is some many-to-one transformation relating the complete data set, \(\mathbf{y}\), to the observation \(r\). Then, instead of solving the problem given in (27), we solve the following maximization problem

\[
\max_{b} \log f_{\mathbf{y}}(\mathbf{y}; b). \tag{29}
\]

However, as mentioned above, \(\mathbf{y}\) is related to \(r\) by a many-to-one transformation and there is no unique \(\mathbf{y}\) for each value of \(r\). Therefore, we replace the log-likelihood function in (29) with its expected value with respect to \(\mathbf{y}\) given \(r\), and maximize the following expression:

\[
\mathbb{E}_r \{ \log f_{\mathbf{y}}(\mathbf{y}; b) | R = r ; b \} = \int \log f_{\mathbf{y}}(\mathbf{y}; b) f_{R | \mathbf{y}}(\mathbf{y} | r ; b) d\mathbf{y}. \tag{30}
\]
Since $b$ is also unknown, we cannot calculate $f_{Y|H}(y|r; b)$ in (30), therefore we replace $b$ in $f_{Y|H}(y|r; b)$ with the current estimate of $b$, i.e. $\hat{b}$, and maximize the following function with respect to its first argument, $b$,

$$U(b, \hat{b}) = \int \log f_Y(y; b) f_{Y|H}(y|r; \hat{b}) dy.$$  \hfill (31)

Using Jensen’s inequality, it can be shown that

$$U(b, \hat{b}) > U(\hat{b}, \hat{b}) \Rightarrow f_R(r; b) > f_R(r; \hat{b}).$$

This provides the following iterative method for maximizing likelihood function and guarantees that the likelihood function does not decrease along the iterations:

- **E-step (Expectation calculation step):** Compute $U(b, \hat{b}^{(n)})$,

$$U(b, \hat{b}^{(n)}) = \int \log f_Y(y; b) f_{Y|H}(y|r; \hat{b}^{(n)}) dy,$$

where $\hat{b}^{(n)}$ is the estimate of $b$ in the $n$th iteration.

- **M-step (Maximization step):** Maximize $U(b, \hat{b}^{(n)})$,

$$\hat{b}^{(n+1)} = \arg \max_b U(b, \hat{b}^{(n)}).$$

Since the likelihood function is bounded above, and since the above estimates monotonically increase in likelihood, we expect the algorithm to converge, to at least a local maximizer. By an appropriate choice of the initial estimates, $\hat{b}^{(0)}$, the algorithm can produce the global maximizer of the likelihood function.

In most cases, if the complete data is chosen properly, the maximization step of the above algorithm can be decomposed into $K$ one-dimensional maximizations, which has linear complexity in $K$ and can be easily implemented for real-time processing.

**V. EM-BASED MULTIUSER DETECTOR**

In order to apply the EM algorithm to the problem in hand, we define the complete data, $y(t) = [y_1(t) \cdots y_K(t)]^T$, where

$$y_k(t) = b_k z_k(t) + n_k(t), \quad \text{for } k = 1, \cdots, K,$$

and $n_k(t), k = 1, \cdots, K$ are independent additive white Gaussian noise with variance $\sigma_k^2$. Then we have $r(t) = \sum_{k=1}^K y_k(t)$, and the log-likelihood function of the complete data is

$$\log f_Y(y; b) = B - \sum_{k=1}^K \frac{1}{2\sigma_k^2} \int_0^{T_n} |y_k(t) - b_k z_k(t)|^2 dt,$$  \hfill (32)

where $B$ is a constant.

In the Appendix, we will show that with this choice of complete data, the result of the **E-step**, i.e. $U(b, \hat{b}^{(n)})$, is given by the following equality:

$$U(b, \hat{b}^{(n)}) = \Re \left\{ \sum_{k=1}^K \frac{b_k}{\sigma_k^2} \left[ \hat{\beta}_k^{(n)} \mathbf{h}_k^H \mathbf{R}_{kk} \mathbf{h}_k + \frac{\sigma_k^2}{\sigma_k^2} \left( \mathbf{h}_k^H z_k - \sum_{j=1}^K \hat{\beta}_j^{(n)} \mathbf{h}_j^H \mathbf{R}_{kj} \mathbf{h}_j \right) \right] \right\}.  \hfill (33)$$

Since the data bit of each user appears only in one of the terms in the summation in (33), we can maximize each term separately in the **M-step**. Therefore, defining $\beta_k = \frac{\sigma_k^2}{\sigma_k^2}$, the iterative equation for updating the $(n+1)$st stage estimate of the data bit of the $k$th user will be

$$\hat{\beta}_k^{(n+1)} = \sgn \left\{ \Re \left[ (1 - \beta_k) \hat{\beta}_k^{(n)} + \frac{\beta_k}{\mathbf{h}_k^H \mathbf{R}_{kk} \mathbf{h}_k} \left( \mathbf{h}_k^H z_k - \sum_{j=1,j \neq k}^K \hat{\beta}_j^{(n)} \mathbf{h}_j^H \mathbf{R}_{kj} \mathbf{h}_j \right) \right] \right\}.  \hfill (34)$$
As mentioned in Section IV, by an appropriate choice of the initial values for the unknown parameters, the algorithm convergences to the global maximizer of the log-likelihood function. As in the well-known multistage detector, a good choice for $\hat{b}_k^{(0)}$ can be the output of the filter matched to the signature signal of the $k$th user, or if, as in our case, multipath and Doppler diversities are available, the maximal ratio combined outputs of the Time-Frequency RAKE receiver for the $k$th user,

$$
\hat{b}_k^{(0)} = \text{sgn} \{\Re[\mathbf{h}_k^H \mathbf{z}_k]\}.
$$

The block diagram of this multiuser detection scheme is shown in Figure 2.

With the above assumption for the initial value for $b$, we can consider two extreme special cases of the new detection scheme as follows

- If $\beta_k = 1$, then the new detector for User $k$ will be the same as the multistage one.
- If $\beta_k = 0$, then the new detector for User $k$ will lose its iterative nature, and will reduce to the Time-Frequency RAKE receiver with maximal ratio combining.

With a suitable choice of parameter $\beta$ for different users, we hope to achieve better performance than both TF RAKE and multistage receivers. According to the discussions of Section III-D, we expect that large (close to one) values of $\beta$ will result in good performance for weak (in terms of signal to interference ratio) users, whereas for strong users, smaller values of $\beta$ will provide better performance. This parameter also determines the speed of convergence of the iterative algorithm. In our simulations discussed in the next section, the value of this parameter for each user is chosen by simulation for the best performance. However, further simulations show that the performance of the detector is not very sensitive to the exact values of these parameters, and values from the following heuristic expression

$$
\beta_k = \frac{\text{ISR}_k}{1 + \text{ISR}_k},
$$

where $\text{ISR}_k$ is a measure of the Interference to Signal Ratio, calculated as

$$
\text{ISR}_k = \sum_{l \neq k} \frac{|\mathbf{h}_k^H \mathbf{R}_{kl} \mathbf{h}_l|}{|\mathbf{h}_k^H \mathbf{R}_{kk} \mathbf{h}_k|},
$$

provides similar performance.

VI. SIMULATION RESULTS

We implemented the EM based multiuser detector and compared its performance with the base case of the Time Frequency RAKE as well as the Multiuser detector. The simulations are done for a system with five users with Gold sequences of spreading length 7. In the EM and multistage detectors we obtained the performance curves for two and three-stage cases. The channel was modeled as a three-path channel, with independent Jakes’ models for each path.

Figures 3 and 4 show the plots of Bit Error Rate (BER) vs. the Signal to Noise Ratio (SNR) for a case with Doppler frequency of 100 Hz. We observe that the performance of the EM based detector is better for both users than the base case of the TF RAKE as well as the multistage detector. Notice that for the multistage detector, the performance of the 3-stage detector is worse than the 2-stage detector for User 2, and doesn’t show much improvement in the performance.
for User 5. As a result, the performance of the 2-stage EM detector is better than the 3-stage multistage detector with higher computational complexity. It should be noted that the computational complexities of these two detectors with the same number of stages are similar. Finally we observe that the 3-stage EM provides significant gains with respect to the multistage case.

Similarly, Figures 5 and 6 show that the performance is consistent with other values of the Doppler (200 Hz). EM detector also shows similar performance for other users.

Note that the different users have different $\beta$'s in the different plots. The appropriate value for parameter $\beta$ can result in a rapid convergence of the EM algorithm. In our simulations, these parameters are chosen by simulation for the best performance within two or three stages. As mentioned in Section V, however, even values obtained from the heuristic expression (36) provide satisfactory performance.
We have presented a new multiuser detector for CDMA systems in fast fading multipath channels. The detector uses the time-frequency RAKE receiver at the front end to exploit multipath and Doppler spreads as two sources of diversity. The multiaccess interference cancellation part of the detector is based on the EM algorithm. It has an iterative structure very similar to the generalized multistage detector but with better convergence properties. As a result, unlike the multistage detector whose performance could become very poor for strong users because of the errors in the decisions of the weak users, this detector shows good performance for all users. Our simulation results show that the new EM-based detector can provide a substantial improvement in performance compared to the generalized multistage detector as well as the TF RAKE.

The improvement in the performance comes at the expense of introducing a set of new parameters which have to be chosen appropriately. In this paper, the optimum values for these parameters were found by simulation and exhaustive
search. Finding an analytical expression for the optimum values of these parameters is not addressed in this paper and requires more investigation, but we have provided an ad hoc expression which is shown to provide satisfactory performance, very close to that of optimum values found by simulation.

**APPENDIX**

In this appendix, we apply the *E-step* of the EM algorithm to (32) to obtain (33). Expanding the squared absolute value in (32) and noting that \( \theta_k = 1 \), we have

\[
\log f_y(y; b) = g(y) + \sum_{k=1}^{K} \frac{1}{\sigma_k^2} \int_0^{T_s} \Re \{ y_k(t) x_k^*(t) \} \, dt,
\]

where \( g(y) \) is a function of \( y \) and does not depend on \( b \).

According to the definition of \( U(b, \hat{\theta}(n)) \), we have to compute the conditional expected value of the log-likelihood function in (38) given the observed signal \( r(t) \), at a parameter value \( \hat{\theta}(n) \). Defining \( \mathbf{C}(t) = \left[ \frac{b_1}{\sigma_1^2} x_1^*(t) \, \cdots \, \frac{b_K}{\sigma_K^2} x_K^*(t) \right]^T \) and ignoring the first term \( g(y) \) which has no effect on the maximization process, we have

\[
U(b, \hat{\theta}(n)) = \Re \left\{ \int_0^{T_s} \mathbf{C}^T(t) \mathbf{E} \left\{ y(t) \left| r(t); \hat{\theta}(n) \right. \right\} \, dt \right\}.
\]

Substituting (41)-(44) in (40) we have

\[
\mathbf{C}_{Yr} = \mathbf{E} \left\{ \left( \frac{r(t)}{\hat{\theta}(n)} \right)^2 \right\} = \sigma_1^2 \cdots \sigma_K^2.
\]

It can be easily shown that

\[
\mathbf{C}_{Yr} = \left[ \sigma_1^2 \cdots \sigma_K^2 \right]^T.
\]

Since both \( y(t) \) and \( r(t) \) given \( \hat{\theta}(n) \) are Gaussian, we can write

\[
\mathbf{E} \left\{ y(t) \left| r(t); \hat{\theta}(n) \right. \right\} = \mathbf{E} \left\{ y(t) \left| \hat{\theta}(n) \right. \right\} + \mathbf{C}_{Yr} \mathbf{C}_{rr}^{-1} \left( r(t) - \mathbf{E} \left\{ r(t) \left| \hat{\theta}(n) \right. \right\} \right) \hat{\theta}(n),
\]

where

\[
\mathbf{C}_{Yr} = \mathbf{E} \left\{ \left( y(t) - \mathbf{E} \left\{ y(t) \left| \hat{\theta}(n) \right. \right\} \right)^* \left( r(t) - \mathbf{E} \left\{ r(t) \left| \hat{\theta}(n) \right. \right\} \right) \right\} \hat{\theta}(n),
\]

and

\[
\mathbf{C}_{rr} = \mathbf{E} \left\{ \left( r(t) - \mathbf{E} \left\{ r(t) \left| \hat{\theta}(n) \right. \right\} \right)^2 \right\}.
\]

Substituting (41)-(44) in (40) we have

\[
\mathbf{E} \left\{ y(t) \left| r(t); \hat{\theta}(n) \right. \right\} = \left[ \frac{\hat{\theta}(n)}{\sigma_1^2} x_1(t) \, \cdots \, \frac{\hat{\theta}(n)}{\sigma_K^2} x_K(t) \right]^T,
\]

and Equation (33) can be obtained by substituting (45) in (39) and using Equations (9), (8), and (11).

**ACKNOWLEDGMENT**

The first author wishes to thank Srikrishna Bhashyam for helpful discussions, Željko Čakareski, Ahmad Khoshnevis, and Vishwas Sundaramurthy for providing some of the simulation programs.
REFERENCES


