# Optimizing Space-time Codes via Stochastic Optimization

Xiaodong Wang * Vikram Krishnamurthy † Jibing Wang ‡

## Abstract

Linear dispersion (LD) codes are a good candidate for high data rate MIMO signaling. Traditionally LD codes were designed by maximizing the average mutual information, which cannot guarantee good error performance. This paper presents a design scheme for LD codes that directly minimizes bit/block error probability in (possibly spatially correlated) fading MIMO channels with an arbitrary receiver. Specifically we propose a simulation-based optimization methodology for the design of LD codes through stochastic approximation and simulation-based gradient estimation. The gradient estimation is done using the score function method originally developed in the discrete event system community. The proposed method can be applied to design the minimum error-rate LD codes for a variety of decoder structures. It can also take into account the knowledge of spatial fading correlation at the transmitter and receiver ends. Simulation results show that codes generated by the new algorithm generally outperform the codes designed based on algebraic number theory.

## 1 Introduction

Linear dispersion (LD) codes introduced in [1] are a good candidate for high data rate MIMO signaling over wireless channels. Traditionally LD codes only optimize the average mutual information; and therefore cannot guarantee good error performance. More recently, another scheme called threaded algebraic space-time (TAST) coding has been proposed [2, 3]. The design of the TAST focuses on the worst-case pairwise error probability (PEP). The PEP, however, may not be the best performance metric, since it is not true in general that the codes optimized with respect to the worst case PEP will end up with the minimum bit or block error rate.

In this paper, we we design LD codes with minimum bit/block error probability based on stochastic approximation together with gradient estimation. Most work on space-time code design assumes maximum likelihood decoding. For very high data rate signaling, even the sphere decoder might be too complicated to implement in practice. It is difficult to design space-time codes where a suboptimal decoder (such as the nulling and cancellation decoder) as the performance analysis seems intractable. One advantage of the proposed method is that we can optimize the LD codes for an arbitrary decoder. On the other hand, in MIMO wireless systems, the individual antennas could be correlated due to insufficient antenna spacing and lack of scattering. We demonstrate how to design optimal LD codes if the long-term spatial correlation can be measured beforehand. Finally, we present simulation results to demonstrate that the LD codes obtained using the proposed design procedure generally outperform the codes designed based on the algebraic number theory, especially when a suboptimal decoder is employed or when the MIMO channels are spatially correlated.

## 2 LD Code Design

In this section we present the signal model for MIMO systems employing LD codes, and formulate the LD design problem as a constrained stochastic optimization problem.

*Dept. of Electrical Engineering, Columbia Univ., New York, NY 10027.

†Dept. of Elect. and Comp. Engineering, Univ. of British Columbia, Vancouver, Canada.

‡Qualcomm Inc., San Diego, CA 92121.

## 2.1 Signal Model

Consider a MIMO system with $M_T$ transmit antennas and $M_R$ receive antennas. Assume that the channel is flat fading and remains constant for $\tau$ symbol intervals. The input output relation is given in matrix form by

$$\boldsymbol{Y} = \sqrt{\rho}\boldsymbol{X}\boldsymbol{H} + \boldsymbol{W}, \tag{1}$$

where $\boldsymbol{Y}$ is the $\tau \times M_R$ matrix of the received signal, $\boldsymbol{X}$ is the $\tau \times M_T$ matrix of the transmitted signal, $\boldsymbol{W}$ is the $\tau \times M_R$ matrix of the additive white Gaussian noise, and $\boldsymbol{H}$ is the $M_T \times M_R$ MIMO channel matrix composed of i.i.d. circularly symmetric complex Gaussian random variables with zero mean and unit variance.

LD codes use a linear modulation framework and the transmitted codeword is a linear combination of certain dispersion matrices with the transmitted symbols as the combining coefficients. Assume we transmit $Q$ $r$-QAM symbols $\{s_q\}_{q=1}^{Q}$ over $\tau$ symbol intervals, the LD transmission matrix $\boldsymbol{X}$ is given by [1]

$$\boldsymbol{X} = \sum_{q=1}^{Q} \alpha_q \boldsymbol{A}_q + j\beta_q \boldsymbol{B}_q, \tag{2}$$

where we have decomposed the transmitted symbols $s_q$ into their real and imaginary parts, i.e., $s_q = \alpha_q + j\beta_q,\ q = 1, \cdots, Q$, and $\{\boldsymbol{A}_q, \boldsymbol{B}_q\}_{q=1}^{Q}$ are complex-valued dispersion matrices of dimension $\tau \times M_T$ that specify the code. The rate of the LD code is $R = \frac{Q \log_2 r}{\tau}$. We also assume that the dispersion matrices $\{\boldsymbol{A}_q, \boldsymbol{B}_q\}_{q=1}^{Q}$ satisfy the following energy constraint

$$\sum_{q=1}^{Q} \mathrm{Tr}\left(\boldsymbol{A}_q^H \boldsymbol{A}_q + \boldsymbol{B}_q^H \boldsymbol{B}_q\right) = 2\tau M_T. \tag{3}$$

As in [1], we denote $\boldsymbol{Y}_R = \mathrm{Re}\{\boldsymbol{Y}\}$ and $\boldsymbol{Y}_I = \mathrm{Im}\{\boldsymbol{Y}\}$. Denote the columns $\boldsymbol{Y}_R$, $\boldsymbol{Y}_I$, $\boldsymbol{H}_R$, $\boldsymbol{H}_I$, $\boldsymbol{W}_R$ and $\boldsymbol{W}_I$, respectively, by $\boldsymbol{y}_{R,n}$, $\boldsymbol{y}_{I,n}$, $\boldsymbol{h}_{R,n}$, $\boldsymbol{h}_{I,n}$, $\boldsymbol{w}_{R,n}$ and $\boldsymbol{w}_{I,n}$; and define $\mathcal{A}_q = \begin{bmatrix} \boldsymbol{A}_{R,q} & -\boldsymbol{A}_{I,q} \\ \boldsymbol{A}_{I,q} & \boldsymbol{A}_{R,q} \end{bmatrix}$, $\mathcal{B}_q = \begin{bmatrix} -\boldsymbol{B}_{I,q} & -\boldsymbol{B}_{R,q} \\ \boldsymbol{B}_{R,q} & -\boldsymbol{B}_{I,q} \end{bmatrix}$, $\underline{\boldsymbol{h}}_i = \begin{bmatrix} \boldsymbol{h}_{R,i} \\ \boldsymbol{h}_{I,i} \end{bmatrix}$. Then we gather equations in $\boldsymbol{Y}_R$ and $\boldsymbol{Y}_I$ to form the single real-valued system of equations [1]

$$\underbrace{\begin{bmatrix} \boldsymbol{y}_{R,1} \\ \boldsymbol{y}_{I,1} \\ \vdots \\ \boldsymbol{y}_{R,M_R} \\ \boldsymbol{y}_{I,M_R} \end{bmatrix}}_{\boldsymbol{y}} = \sqrt{\rho}\boldsymbol{\mathcal{H}} \underbrace{\begin{bmatrix} \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_Q \\ \beta_Q \end{bmatrix}}_{\boldsymbol{x}} + \underbrace{\begin{bmatrix} \boldsymbol{w}_{R,1} \\ \boldsymbol{w}_{I,1} \\ \vdots \\ \boldsymbol{w}_{R,M_R} \\ \boldsymbol{w}_{I,M_R} \end{bmatrix}}_{\boldsymbol{w}} \tag{4}$$

where the equivalent $2M_R\tau \times 2Q$ real-valued channel matrix $\boldsymbol{\mathcal{H}}$ is given by

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \mathcal{A}_1\underline{\boldsymbol{h}}_1 & \mathcal{B}_1\underline{\boldsymbol{h}}_1 & \dots & \mathcal{A}_Q\underline{\boldsymbol{h}}_1 & \mathcal{B}_Q\underline{\boldsymbol{h}}_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{A}_1\underline{\boldsymbol{h}}_{M_R} & \mathcal{B}_1\underline{\boldsymbol{h}}_{M_R} & \dots & \mathcal{A}_Q\underline{\boldsymbol{h}}_{M_R} & \mathcal{B}_Q\underline{\boldsymbol{h}}_{M_R} \end{bmatrix}. \tag{5}$$

From (2) we can see that LD codes are very simple to encode. Furthermore, LD codes can be decoded very efficiently either by a polynomial-time maximum likelihood decoder, i.e., the sphere decoder; or by suboptimal decoders, e.g., the nulling and cancellation decoder.

## 2.2 Problem Formulation

The LD codes introduced in [1] are designed to maximize the average mutual information between the input and output. However maximizing the average mutual information does not necessarily lead to better performance in terms of error rate. Unfortunately, the average bit or block error rate is hard to analyze for arbitrary LD codes. Simulation optimization turns out to be useful is this scenario. In this work, we demonstrate how to optimize the average bit error rate (BER) for LD codes through simulation optimization with gradient estimation. The block error performance can be optimized similarly. First, we denote $\boldsymbol{h} = \left[\underline{\boldsymbol{h}}_1^T, \dots, \underline{\boldsymbol{h}}_{M_R}^T\right]^T$, and denote the set of dispersion matrices as $\boldsymbol{\theta} \triangleq \{\boldsymbol{A}_q, \boldsymbol{B}_q, q = 1, \dots, Q\}$. With a slight abuse of notation, we also use $\boldsymbol{\theta}$ to denote the column vector that stacks all the columns of $\boldsymbol{A}_{R,q}$, $\boldsymbol{A}_{I,q}$, $\boldsymbol{B}_{R,q}$, and $\boldsymbol{B}_{I,q}$, for $q = 1, \dots, Q$. Note that $\boldsymbol{\theta}$ is a $(4\tau M_T Q)$-dimensional vector.

Let $\boldsymbol{y}_n$ denote the $n$th block of the received signal corresponding to the $n$th block of transmitted signal $\boldsymbol{x}_n$ and channel $\boldsymbol{h}_n$, $n = 1, 2, \dots$.

Let $\gamma(\boldsymbol{y}_n, \boldsymbol{x}_n, \boldsymbol{h}_n, \boldsymbol{\theta})$ denote the *empirical BER* for a given set of dispersion matrices $\boldsymbol{\theta}$. That is,

$$\gamma(\boldsymbol{y}_n, \boldsymbol{x}_n, \boldsymbol{h}_n, \boldsymbol{\theta}) = \frac{\#\text{ of bit errors in } \boldsymbol{y}_n}{\#\text{ of bits in } \boldsymbol{y}_n}. \quad (6)$$

Note that for fixed $\boldsymbol{\theta}$, since $\{\boldsymbol{y}_n, \boldsymbol{x}_n, \boldsymbol{h}_n\}$ is an i.i.d. sequence, $\{\gamma(\boldsymbol{y}_n, \boldsymbol{x}_n, \boldsymbol{h}, \boldsymbol{\theta})\}$ is also an i.i.d. sequence of random variables. For a given set of dispersion matrices $\boldsymbol{\theta}$, the *average BER* denoted by $\Upsilon(\boldsymbol{\theta})$ is given by

$$\Upsilon(\boldsymbol{\theta}) = E\{\gamma(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{h}, \boldsymbol{\theta})\} \quad (7)$$
$$= \int\int\int \gamma(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{h}, \boldsymbol{\theta})\, p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{h} \mid \boldsymbol{\theta})\, d\boldsymbol{y} d\boldsymbol{x} d\boldsymbol{h},$$

where $p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{h} \mid \boldsymbol{\theta})$ is the joint probability density function (pdf) of $(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{h})$ for a given $\boldsymbol{\theta}$. The integrals in (7) are over the space $\mathbb{R}^{2M_R\tau}$ (for $\boldsymbol{y}$), $\mathcal{A}^{2Q}$ (for $\boldsymbol{x}$), where $\mathcal{A}$ is the discrete set of real-valued constellation symbols that elements of $\boldsymbol{x}$ take value from, and $\mathbb{R}^{2M_RM_T}$ (for $\boldsymbol{h}$), respectively. For notational simplicity we subsequently omit the space over which these integrals are defined.

*Aim*: The design goal is to solve the following constrained stochastic optimization problem: Given the sequence of empirical BER measurements $\gamma(\boldsymbol{y}_n, \boldsymbol{x}_n, \boldsymbol{h}_n, \boldsymbol{\theta})$ for any choice of $\boldsymbol{\theta}$, find the optimal set of dispersion matrices to minimize the average BER, i.e., compute

$$\min_{\boldsymbol{\theta}\in\Theta} \Upsilon(\boldsymbol{\theta}), \quad (8)$$

with the energy constraint set $\Theta$ given by

$$\Theta = \left\{\boldsymbol{\theta} \in \mathbb{R}^{4\tau M_T Q} : \boldsymbol{\theta}^T\boldsymbol{\theta} = 2\tau M_T.\right\} \quad (9)$$

Let $\boldsymbol{\theta}^*$ denote a local minimum of (8) subject to the constraint (9). From (7), we have

$$\Upsilon(\boldsymbol{\theta}) = E_{\boldsymbol{x}} E_{\boldsymbol{h}} E_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{h},\boldsymbol{\theta}} \{\gamma(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{h}, \boldsymbol{\theta})\}. \quad (10)$$

*Remark 1:* Because $p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{h}, \boldsymbol{\theta})$ is Gaussian (as we will show later), and it is continuously differentiable in $\boldsymbol{\theta}$, it follows that $\Upsilon(\boldsymbol{\theta})$ is continuously differentiable in $\boldsymbol{\theta}$. ( This point will be clear in the next section.) Hence $\Upsilon(\boldsymbol{\theta})$ attains a minimum on the compact set $\Theta$ and the optimization problem (8)-(9) is well posed.

*Remark 2:* Note that an explicit closed-form expression for the average BER $\Upsilon(\boldsymbol{\theta})$ is usually not available. Indeed, $\Upsilon(\boldsymbol{\theta})$ also depends on the particular decoder employed (e.g., ML decoder or suboptimal decoder). We will use a stochastic gradient algorithm that uses measurements of the empirical BER $\gamma(\boldsymbol{y}_n, \boldsymbol{x}_n, \boldsymbol{h}_n, \boldsymbol{\theta})$ to compute the optimal LD code $\boldsymbol{\theta}^*$. On the other hand, $\nabla_{\boldsymbol{\theta}}\Upsilon(\boldsymbol{\theta})$ cannot be computed analytically. We therefore need to devise a scheme that estimates the gradient $\nabla_{\boldsymbol{\theta}}\Upsilon(\boldsymbol{\theta})$ using the empirical BER measurements $\gamma(\boldsymbol{y}_n, \boldsymbol{x}_n, \boldsymbol{h}_n, \boldsymbol{\theta})$.

### 2.3 Spherical Parameterization

Recall that the dimension of $\boldsymbol{\theta}$ is $4\tau M_T Q$. Denote $d \triangleq 4\tau M_T Q$ In this subsection, we parameterize the dispersion matrices $\boldsymbol{\theta}$ by spherical coordinates $\boldsymbol{\psi} \in \mathbb{R}^{d-1}$. We show that under these spherical coordinates $\boldsymbol{\psi}$ the constrained optimization (8)-(9) transforms to the equivalent unconstrained optimization problem. This in turn implies that we can present a simple convergence proof without requiring any form of projection of the estimates which typically makes convergence proofs for stochastic approximation algorithms very difficult.

Consider $\boldsymbol{\theta}(\boldsymbol{\psi})$ parameterized by $\boldsymbol{\psi} \in \mathbb{R}^{d-1}$. Here the spherical coordinate $\boldsymbol{\psi} = [\Psi_1, \ldots, \Psi_{d-1}]^T \in \mathbb{R}^{d-1}$, such that

$$\theta_p(\boldsymbol{\psi}) = \sqrt{2\tau M_T}\cos\Psi_p \prod_{k=1}^{p-1}\sin\Psi_k\,, \quad (11)$$

$$\theta_d(\boldsymbol{\psi}) = \sqrt{2\tau M_T}\sin\Psi_{d-1} \prod_{k=1}^{d-2}\sin\Psi_k. \quad (12)$$

Note that the transformation from $\boldsymbol{\theta}$ to $\boldsymbol{\psi}$ is invertible. Therefore, the constrained optimization (8)-(9) in $\boldsymbol{\theta}$ is equivalent to an unconstrained one in $\boldsymbol{\psi}$:

$$\min_{\boldsymbol{\theta}\in\Theta} \Upsilon(\boldsymbol{\theta}) = \min_{\boldsymbol{\psi}\in\mathbb{R}^{d-1}} \Upsilon(\boldsymbol{\theta}(\boldsymbol{\psi})). \quad (13)$$

## 3 Optimal Code Design

We propose the following two-stage simulation algorithm to generate the gradient estimate $\widehat{\boldsymbol{g}}(\boldsymbol{\psi})$.

**Algorithm 1** *[Composite-score function algorithm]* Given the set of dispersion matrices $\boldsymbol{\theta}(\boldsymbol{\psi}_k)$ at the $k$-th iteration, perform the following simulation steps:

- Composition method to generate mixture sample :

  1. Draw $M$ symbol vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M$ uniformly from the constellation set $\mathcal{A}^{2Q}$.
  2. Simulate $M$ observations $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_M$ where each $\boldsymbol{y}_i$ is generated according to (4) using symbol vector $\boldsymbol{x}_i$, i.e., $\boldsymbol{y}_i = \sqrt{\rho}\boldsymbol{\mathcal{H}}_i\boldsymbol{x}_i + \boldsymbol{w}_i, \quad i = 1, 2, \ldots, M$.
  3. Using the given decoding algorithm, decode $\boldsymbol{x}_i$ based on the observations $\boldsymbol{y}_i$ and the channel value $\boldsymbol{\mathcal{H}}_i$, $i = 1, 2, \ldots, M$. Compute the empirical BER $\gamma(\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{h}_i, \boldsymbol{\theta}(\boldsymbol{\psi}_k))$.

- Score function method for gradient estimation: Using the empirical BER $\gamma(\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{h}_i, \boldsymbol{\theta}(\boldsymbol{\psi}_k))$, compute the gradient estimate as

$$\widehat{\boldsymbol{g}}(\boldsymbol{\psi}_k) = \frac{1}{M}\sum_{i=1}^{M}\gamma(\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{h}_i, \boldsymbol{\theta}_k) \times \left[\nabla_{\boldsymbol{\psi}} \log p(\boldsymbol{y}_i \mid \boldsymbol{x}_i, \boldsymbol{h}_i, \boldsymbol{\theta}) \mid_{\Psi=\Psi_k}\right] \quad (14)$$

  where the element of $\nabla_{\boldsymbol{\psi}} \log p(\boldsymbol{y}_i \mid \boldsymbol{x}_i, \boldsymbol{h}_i, \boldsymbol{\theta})$ is given by

$$\frac{\partial \log p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{h}, \boldsymbol{\theta})}{\partial \Psi_p} = \left[\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{h}, \boldsymbol{\theta})\right]^T \left(\frac{\partial \boldsymbol{\theta}}{\partial \Psi_p}\right) \quad (15)$$

  An explicit formula for $\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{h}, \boldsymbol{\theta})$ can be obtained.

We next use the above gradient estimator in a stochastic approximation algorithm to solve the optimization problem given in (13).

**Algorithm 2** *[Optimal LD codes design]* Assume at the $k$-th iteration the current set of dispersion matrices is $\boldsymbol{\theta}(\boldsymbol{\psi}_k)$ with the coordinate parameterization $\boldsymbol{\psi}_k$, perform the following steps during the next iteration to generate $\boldsymbol{\psi}_{k+1}$ and the corresponding $\boldsymbol{\theta}(\boldsymbol{\psi}_{k+1})$:

- Same as Step 1 and 2 of Algorithm 1.
- Update new set of dispersion matrices: Generate

$$\boldsymbol{\psi}_{k+1} = \boldsymbol{\psi}_k - a_k\widehat{\boldsymbol{g}}(\boldsymbol{\psi}_k), \quad (16)$$

  then update $\boldsymbol{\theta}(\boldsymbol{\psi}_{k+1})$ according to (11) and (12).

The convergence of the above algorithm is given by the following theorem:

**Theorem 1** *Under the conditions: 1) the step size $\{a_n\}$ satisfies $\sum_{n=1}^{\infty} a_n = \infty$, $\sum_{n=1}^{\infty} a_n^2 < \infty$; 2) $\widehat{\boldsymbol{g}}(\boldsymbol{\psi}_k)$ is uniformly integrable; 3) $\nabla_{\boldsymbol{\psi}}\Upsilon(\boldsymbol{\theta}(\boldsymbol{\psi}))$ is Lipschitz for all $\boldsymbol{\psi}$ in $[0, 2\pi]^{d-1}$; then the sequence of estimates $\boldsymbol{\psi}_k$ generated by the stochastic approximation Algorithm 2 converges to $\boldsymbol{\psi}^*$ almost surely, i.e.,*

$$P(\lim_{k\to\infty} \boldsymbol{\psi}_k = \boldsymbol{\psi}^*) = 1. \quad (17)$$

*Thus $\boldsymbol{\theta}(\boldsymbol{\psi}_k)$ converges to $\boldsymbol{\theta}(\boldsymbol{\psi}^*) = \boldsymbol{\theta}^*$ almost surely.*

# 4 Simulation Results

In this section, we give three examples that illustrate the performance of the LD codes obtained by Algorithm 2. As mentioned before, our code design depends on the operating SNR. In the following examples, we design the codes by choosing the SNR so that the BER is around $10^{-2}$. We will see that the codes optimized for a particular SNR work fine for a whole range of SNR of interest.

*Example 1– New LD codes with maximum-likelihood decoder:* We first present simulation results for i.i.d. fading channels using the sphere decoder. We consider a system with three transmit antennas and two receive antennas. Figure 1 shows the BER performance of the new LD codes and the TAST codes employing 16QAM constellations. The rate $R$ is 8 bits/sec/Hz. It is seen that the new codes perform better than the TAST codes for the entire range of SNR. Note that the new LD codes have the same encoding and decoding complexity as the TAST codes.

*Example 2– New LD codes with nulling and cancellation decoder:* In this example we consider

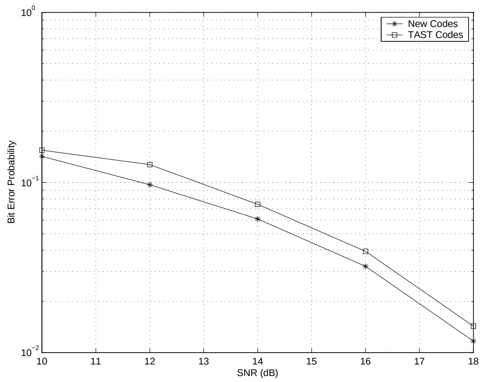


Figure 1: BER performance of the new LD code and the TAST code in $3 \times 2$ uncorrelated MIMO channel with and 16QAM constellation.

designing LD codes for suboptimal decoders, in particular, the zero-forcing nulling and cancellation decoder. We also assume i.i.d. fading channels. In Fig. 2 we present the BER performance of the new code optimized for four transmit antennas and one receive antenna with 16QAM constellation. We also show the performance of the DAST code and a randomly generated LD codes. The gain of the new LD code can be clearly seen. The rate $R$ is kept to be 4 bits/sec/Hz.

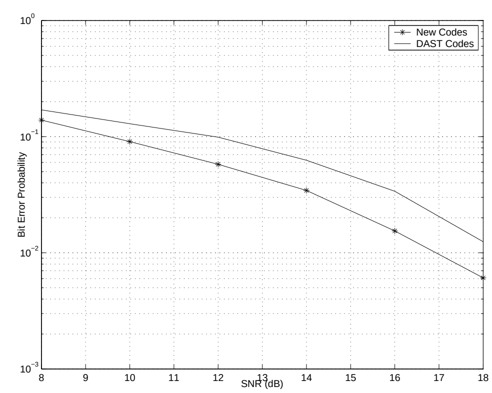


Figure 2: BER performance of the new LD code and the DAST code in $4 \times 1$ uncorrelated MIMO channel with and 16QAM constellation.

*Example 3– New LD codes for spatially correlated Rayleigh fading channels:* Finally we give an example of the optimal LD codes for spatially correlated Rayleigh fading channels. We consider a $2 \times 2$ MIMO channel with spatial correlation at both the transmitter and the receiver. The correlation matrices are given by $\boldsymbol{S} = \boldsymbol{R} = \begin{bmatrix} 1 & 0.7+0.7j \\ 0.7-0.7j & 1 \end{bmatrix}$. We assume QPSK constellation is employed so the data rate is $R = 4$ bits/sec/Hz. Fig. 3 shows the BER performance of the LD code obtained by Algorithm 2 over this correlated MIMO channel. In the same figure we also show the performance of the TAST code. Note that although for $2 \times 2$ i.i.d. fading channels using QPSK constellation, the LD codes obtained by Algorithm 2 performs roughly the same as the TAST code, however, in the presence of spatial correlation, the new LD codes outperforms the TAST code considerably.

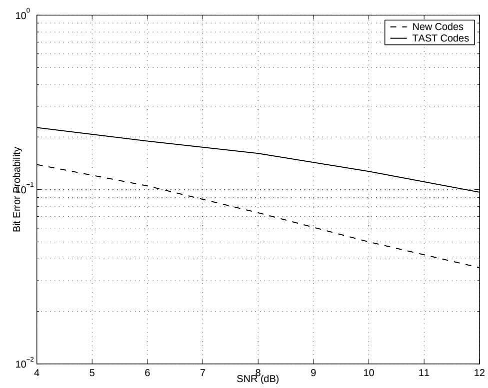


Figure 3: BER performance of the new LD code and the TAST code in $2 \times 2$ correlated MIMO channel with QPSK constellation.

# References

[1] B. Hassibi and B. Hochwald, “High-rate codes that are linear in space and time, ” *IEEE Trans. Inform. Theory,* vol. 48, pp. 1804-1824, July 2002.

[2] H. E. Gamal, and M. O. Damen, “Universal space-time coding,” *IEEE Trans. Inform. Theory,* vol. 48, pp. 1097-1119, May 2003.

[3] M. O. Damen, K. Abed-Meraim, and J. C. Belfiore, “Diagonal algebraic space-time block codes,” *IEEE Trans. Inform. Theory*, vol. 48, Mar. 2002.

[4] H. J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications,* Springer, New York, NY, 2003.