Linear Dispersion Codes for MIMO Systems Based on Frame Theory

Robert W. Heath, Jr., *Member, IEEE*, and Arogyaswami J. Paulraj, *Fellow, IEEE*

***Abstract*—Multiple-input multiple-output (MIMO) wireless communication systems provide high capacity due to the plurality of modes available in the channel. Existing signaling techniques for MIMO systems have focused primarily on multiplexing for high data rate or diversity for high link reliability. In this paper, we present a new linear dispersion code design for MIMO Rayleigh fading channels. The proposed design bridges the gap between multiplexing and diversity and yields codes that typically *perform well both in terms of ergodic capacity as well as error probability*. This is important because, as we show, designs performing well from an ergodic capacity point of view do not necessarily perform well from an error probability point of view. Various techniques are presented for finding codes with good error probability performance. Monte Carlo simulations illustrate performance of some example code designs in terms of ergodic capacity, codeword error probability, and bit error probability.**

***Index Terms*—Diversity methods, MIMO systems, smart antennas, space-time codes.**

I. Introduction and Outline

MULTIPLE-INPUT multiple-output (MIMO) wireless communication systems, i.e., wireless systems with multiple transmit and receive antennas, are important due to their potential for significant spectrum efficiency [1]–[3]. Of particular interest are those schemes that assume channel knowledge at the receiver but no knowledge at the transmitter [4] since training sequences are typically available. Practical modulation schemes for MIMO systems with receive-only channel knowledge fall principally into two areas known as diversity and multiplexing [5]. Diversity modulation, or space-time coding [6]–[9], uses specially designed codewords that maximize the diversity advantage or reliability of the transmitted information. In fading channels, such codes maximize the diversity gain at the expense of a loss in capacity [1]. Spatial multiplexing [10] (or BLAST [4]), on the other hand, transmits independent data streams from each transmitting antenna. Multiplexing designs allow capacity to be achieved but at the expense of a loss in diversity advantage [11] in fading channels. In practical systems, it is desirable to provide both high spectrum efficiency and high reliability; thus, new space-time signaling techniques are needed.

Recognizing that orthogonal space-time block codes [12] designed to maximize diversity advantage do not achieve full channel capacity in MIMO channels (this was also observed in [13]), Hassibi and Hochwald proposed the revolutionary linear dispersion codes (LDCs) [1]. These codes use a linear matrix modulation framework, similar to that in [9], in which the transmitted codeword is a linear combination of certain dispersion matrices with the weights determined by the transmitted symbols. The key to the LDC design is that the basis matrices are chosen such that the resulting codes maximize the ergodic capacity of the equivalent MIMO system. Unfortunately, the LDCs proposed in [1] only optimize the ergodic capacity; thus, corresponding good error probability performance is not strictly guaranteed [14].

In this paper, we present a family of LDC designs based on frame theory [15]. Our designs are tailored for the frequency flat independent and identically distributed (i.i.d.) spatially white complex Gaussian channel known perfectly at the receiver but not at the transmitter. Maximum likelihood (ML) detection is assumed at the receiver. Instead of sending uncoded symbols, our frame-based codes convey the coefficients of a frame expansion of a vector of symbols. Most existing linear codes, for example, spatial multiplexing [10], the Alamouti code, other orthogonal designs [8], [9], [12], and the previously proposed LDCs all have a frame-based structure. We show that with suitable choice of parameters, frame-based LDCs have equivalent channels that achieve the full-ergodic capacity. We extend this design to find low-rate LDCs that are near optimal in terms of capacity yet have a frame-theoretic interpretation. The proposed LDCs can be conveniently represented using the theory of unitary matrices and tight frames [15]. Since we do not require a numerical optimization, as in [1], we can instead optimize over the space of codes with similar ergodic capacity to find those that also have good performance in terms of error probability. To illustrate, we show how to improve our frame-based LDCs using the rank and determinant criteria [6], [7]. Different techniques are presented for finding good codes using optimization with various initial conditions.

The original LDCs proposed in [1] are designed via a numerical optimization, to maximize the mutual information between transmit and receiver. In contrast, we present a closed-form code design that, in some cases, produces an equivalent channel that maximizes the ergodic capacity. We show that capacity optimization, as performed in [1], does not necessarily guarantee good performance in terms of diversity advantage. Therefore,

Manuscript received October 15, 2001; revised May 15, 2002. A portion of these results were presented at the International Conference on Acoustics, Speech, and Signal Processing, Salt Lake City, UT, 2001 and at the International Symposium on Signals, Systems, and Electronics, Tokyo, Japan, July 2001. The associate editor coordinating the review of this paper and approving it for publication was Dr. Bertrand M. Hochwald.

R. W. Heath, Jr., is with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA (e-mail: rheath@ece.utexas.edu).

A. J. Paulraj is with the Information Systems Laboratory, Stanford University, Stanford, CA 94305 USA (e-mail: apaulraj@stanford.edu).

Publisher Item Identifier 10.1109/TSP.2002.803325.

to guarantee a minimum diversity advantage, we further optimize our choice of codewords using the rank and determinant criterion [7].

LDCs that maximize the received signal-to-noise ratio (SNR) are described in [9]. Such codes, which are also known as orthogonal designs [8], [12], satisfy an orthogonality constraint on each matrix and a skew Hermitian cross-relation. The resulting design decouples the symbol streams at the receiver to simplify ML detection. Unfortunately, codes satisfying the orthogonality constraint exist for only a few choices of parameters and do not achieve the ergodic capacity when multiple receive antennas are available [1]. In contrast, the code design proposed herein imposes the less restrictive cross relation that the basis matrices should be orthogonal in terms of the Frobenius norm matrix inner product. Our design is general enough that codes can be found for many combinations of transmit antennas, receive antennas, codeword lengths, and rates. On the other hand, our designs do not simplify ML detection.

This paper is organized as follows. Section II first reviews the channel model and presents our version of the LDC framework. Section III reviews the ergodic capacity with LDCs and presents the frame-based LDC design. Error probability considerations are addressed in Section IV. Section V contains a description of some numerical procedures that can be used to find good frame-based LDCs. Section VI presents example code designs, comparisons, and simulation results. Finally, Section VII presents our conclusions.

II. MIMO Communication Using Matrix Modulation

In this section, we review the MIMO communication system considered in this paper. We start with a brief description of the channel and the assumptions that enable this channel model. We then review the linear dispersion code description of linear space-time block codes [1].

A. Channel Model and Assumptions

A MIMO communication system with M_t transmit antennas and M_r receive antennas is illustrated in Fig. 1. The space-time encoder takes input symbols and generates a codeword matrix, that is, a codeword with dimension in both space and time. The codeword is launched into the propagation environment from M_t transmit antennas and arrives at the M_r receive antennas. The receiver is assumed to have perfect channel knowledge while the transmitter has no channel knowledge. To describe the input–output relationship of the system in more detail, we first describe the propagation channel between the transmitter and receiver.

Suppose that the transmission bandwidth is much less than the coherence frequency of the channel (thus, the channel is frequency flat), the antenna spacing is larger than the coherence distance (thus, antennas are decorrelated), the codewords are separated by at least the coherence time of the channel (thus, the channel is independent from observation to observation), and assume sufficient scattering in the environment (so that the each element of the matrix is independent). This gives rise to the so-called block fading Gaussian matrix channel model, where the channel is described by an $M_r \times M_t$ matrix $\mathbf{H}$ whose elements $[\mathbf{H}]_{m,k}$ are i.i.d. circularly symmetric complex Gaussian random variables with distribution $\mathcal{CN}(0,1)$. The matrix is constant over the duration of the codeword of interest but varies independently from codeword to codeword. Thus, each codeword sees a different channel realization, and the channel coefficients at each realization are independent in space and time.

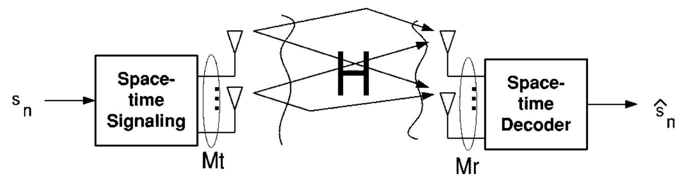

Fig. 1. MIMO\communication link employing space-time signaling and decoding.

The i.i.d. block fading complex Gaussian channel model has seen extensive use in the past, e.g., [6], [7], and [16]. Extension to a more general channel model that includes correlation [17] or delay spread [18] is a subject for future work.

B. Linear Dispersion Codes

Consider a space-time block code that transmits RN bits in T periods across M_t transmit antennas. The code is specified by the codebook and the rule for mapping the incoming bit string to the codewords. For space-time block codes, the codebook is comprised of 2^{RN} space-time codewords, each of which is a matrix with dimensions $M_t \times T$. The rule for mapping bits to space-time codewords is generally the one that minimizes the bit error probability for the given codebook. Without structure in the codewords, decoding a general space-time block code may be difficult due to the significant complexity and storage requirements that grow with larger rates. To overcome these difficulties, in this paper, we focus on linear dispersion codes [1] in which the codewords are a linear function of the data symbols.

The linear dispersion encoder derives space-time codewords from linear combinations of certain basis matrices. The encoder may operate using complex modulation, in which each complex symbol modulates a different complex codeword matrix, or separate modulation, in which the real and imaginary components of a complex symbol each modulate a separate possibly complex codeword matrix. The original LDCs were based on separate modulation since it allows the conjugation operation, which is a key feature of the Alamouti scheme [8], as well as other orthogonal space-time block codes [12]. For nonorthogonal space-time codes, we do not always have significant performance differences between separate and complex codes (see examples in [19]); therefore, to simplify the explanation, we focus on complex LDCs in this paper.

Let $\{s_n\}_{n=0}^{N-1}$ be a set of scalar symbols from some complex constellation that are to be transmitted. Let $\{\mathbf{M}_n\}_{n=0}^{N-1}$ be the set of $M_t \times T$ codeword matrices. Assuming that $\mathcal{E}s_n = 0$ and $\mathcal{E}|s_n|^2 = 1$, the basis matrices should satisfy the power constraint[1]

$$\operatorname{tr}\left\{\sum_{n=0}^{N-1} \mathbf{M}_n \mathbf{M}_n^H\right\} = T \quad (1)$$

[1]In this paper, $\mathcal{E}$ stands for expectation, * for elementwise conjugation, ^{T} for transpose, ^{H} for Hermitian transpose, vec$(\cdot)$ for the operator that forms a vector from successive columns of a matrix, and $\otimes$ is the Kronecker product.

though more practically, each basis matrix should contain the same power

$$\text{tr}\left\{\mathbf{M}_n\mathbf{M}_n^H\right\} = \frac{T}{N} \quad n = 0, 1, \ldots, N-1.$$

More discussion on normalization is available in [1].

A codeword corresponding to $\{s_n\}_{n=0}^{N-1}$ is constructed by taking the corresponding linear combination of basis matrices

$$\mathbf{S}(s_0, s_1, \ldots, s_{N-1}) = \sum_{n=0}^{N-1} \mathbf{M}_n s_n. \tag{2}$$

The coefficient of the codeword $[\mathbf{S}(s_0, s_1, \ldots, s_{N-1})]_{k,m}$ gives the symbol to be transmitted on the kth transmit antenna during the mth symbol period. The code is fully determined by the set of codeword matrices $\{\mathbf{M}_n\}_{n=0}^{N-1}$ that are i) *known to both transmitter and receiver* and ii) *independent of the channel realization*.

After matched filtering and symbol-rate sampling, the receiver concatenates T observations to form

$$\mathbf{Y} = \sqrt{E_s}\mathbf{H}\mathbf{S}(s_0, s_1, \ldots, s_{N-1}) + \mathbf{V} \tag{3}$$

$$= \sqrt{E_s}\mathbf{H}\sum_{n=0}^{N-1}\mathbf{M}_n s_n + \mathbf{V} \tag{4}$$

where $\mathbf{Y}$ is a $M_r \times T$ matrix constructed by concatenating the T receive vectors, and $\mathbf{V}$ is a $M_r \times T$ matrix whose columns represent realizations of an i.i.d. circular complex additive white Gaussian noise (AWGN) process with distribution $\mathcal{CN}(0, N_o\mathbf{I}_{M_r})$.

It is often desirable to write the matrix input–output relationship in (4) in an equivalent vector notation. Define the linear transformation matrix $\mathcal{X} := [\text{vec}(\mathbf{M}_0), \text{vec}(\mathbf{M}_1), \ldots, \text{vec}(\mathbf{M}_{N-1})]$ and the stacked channel matrix $\mathcal{H} := \mathbf{I}_T \otimes \mathbf{H}$. Taking the vec() of both sides of (3) gives

$$\mathbf{y} = \sqrt{E_s}\mathcal{H}\mathcal{X}\mathbf{s} + \mathbf{v} \tag{5}$$

where $\mathbf{y} := \text{vec}(\mathbf{Y})$, $\mathbf{s} := [s_0, s_1, \ldots, s_{N-1}]^T$, and $\mathbf{v} := \text{vec}(\mathbf{V})$. Essentially, matrix modulation transforms the $M_r \times M_t$ linear system into an expanded $M_rT \times N$ system. The linear nature of the encoding operation is evident since, in the absence of noise, the input $\mathbf{s}$ and output $\mathbf{y}$ are related by the linear transformation $\mathcal{H}\mathcal{X}$.

In this paper, the ML decoding rule, optimal assuming equally likely transmitted symbols, is used at the receiver. In a vector AWGN channel, the detected vector symbol obtained using the ML decoder is the solution of

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}\in\mathcal{S}} \|\mathbf{y} - \sqrt{E_s}\mathcal{H}\mathcal{X}\mathbf{s}\|_2^2$$

where $\mathcal{S}$ is the set of all possible vector symbols $\mathbf{s}$. Note that if $[\mathbf{s}]_n$ comes from a constellation with 2^R points, there are $|\mathcal{S}| = 2^{RN}$ possible vector symbols. Thus straightforward implementation is exponential in RN. Compared with a general space-time block code, decoding complexity is still exponential; however, the storage requirements are dramatically reduced for LDCs since the code is fully specified by the dispersion matrices $\{\mathbf{M}_n\}_{n=0}^{N-1}$. Lower complexity decoding with little loss is possible using spherical decoding techniques based on the theory of lattice decoders [20]–[22]. In either case, however, decoding complexity grows with increasing N; thus, the drawback of linear dispersion codes is that they increase the decoding complexity as well as an increase decoding delay due to choice of $T > 1$.

III. Near-Capacity Optimal Linear Dispersion Codes

In this section, we introduce a closed-form solution for a set of LDCs that are capacity-optimal with appropriate choice of parameters and near capacity-optimal otherwise. First, we summarize the ergodic capacity of the MIMO communication system without LDCs and with capacity-optimal LDCs. We then present a LDC code design that produces capacity-optimal LDCs for certain choices of parameters. We extend this design to find low-rate codes that are near capacity optimal.

A. Summary of Ergodic Capacity Results

In this paper, we consider the ergodic capacity, which is the capacity obtained assuming it is possible to code over many independent channel realizations. This is relevant because code designs for practical systems often include interleavers that enable the code to experience many different channel realizations. Thus, the ergodic capacity achieved by codes that experience an arbitrarily large number of channel realizations is a valuable upperbound on realistic code performance.

The ergodic capacity of an $M_r \times M_t$ AWGN channel with Rayleigh fading has been derived by a number of authors (e.g., [2], [3]) and is given by

$$C_{M_r\times M_t} = \mathcal{E}_{\mathbf{H}} \log\det\left(\mathbf{I}_{M_r} + \frac{E_s}{M_tN_o}\mathbf{H}\mathbf{H}^H\right) \tag{6}$$

where the choice of input distribution which maximizes the mutual information is circular complex Gaussian with covariance $\mathbf{R}_s = 1/M_t\mathbf{I}_{M_t}$. The significant rate and capacity advantages due to the multivariate nature of the channel are well known.

B. Capacity-Optimal Linear Dispersion Codes

Obtaining the ergodic capacity using a capacity-optimal LDCs—by definition—requires maximizing the mutual information with respect to both the input distribution and the coefficients of the LDC. Such LDCs have the largest possible ergodic capacity that can be achieved by any LDC with the same parameters. Maximizing the ergodic capacity is one potential LDC design criterion and codes designed under this criterion are described more thoroughly in [1].

Using the input–output relationship in (5) and applying the results in [2], the ergodic capacity of the AWGN system in (5) with Rayleigh fading for capacity-optimum complex LDCs is given by

$$C_c = \max_{\text{tr}(\mathcal{X}\mathcal{X}^H)\le T} \frac{1}{T}\mathcal{E}_{\mathbf{H}} \log\det\left(\mathbf{I}_{M_rT} + \frac{E_s}{N_o}\mathcal{H}\mathcal{X}\mathcal{X}^H\mathcal{H}^H\right). \tag{7}$$

In expressing (7), we have taken $\mathbf{R}_s := \mathcal{E}_{\mathbf{s}}\mathbf{s}\mathbf{s}^H = \mathbf{I}$ without loss of generality since for any $\mathbf{R}_s$, $\mathcal{X}\mathbf{R}_s\mathcal{X}^H = \hat{\mathcal{X}}\hat{\mathcal{X}}^H$ for some $\hat{\mathcal{X}}$. Using some manipulations, (7) can also be obtained from equivalent expressions in [1].

Comparing (6) and (7), it is clear that the effect of the linear code is to color the covariance of the input. An *optimum LDC*, however, colors the input in such a way that the mutual information in (7) is maximized. It is clear that $C_c \leq C_{M_r \times M_t}$, depending on the choice of parameters.

C. Capacity-Optimal Linear Dispersion Codes

In general, finding a code design that induces an equivalent channel with full channel capacity $C_{M_r \times M_t}$ is difficult since the mutual information cost function is nonconvex. In [1], optimization is used to solve for capacity-optimal LDCs numerically. The resulting solution is not guaranteed to be the global maximum of the cost function, although it is claimed that it is typically near the global maximum. From (7), it is possible to obtain a capacity-maximizing LDC by choosing N according to M_t and T. Consider $N \geq M_t T$. It is easy to show by substituting into (7) that any $\mathcal{X}$ such that $\mathcal{X}\mathcal{X}^H = 1/M_t \mathbf{I}_{TM_t}$ satisfies the power constraint $\text{tr}\{\mathcal{X}\mathcal{X}^H\} = T$ and produces an equivalent channel with a capacity of $C_{M_r \times M_t}$. Essentially, this solution spreads the input signal across all transmit antennas in all time periods. Since decoding complexity grows with N, it is sufficient to take $N = M_t T$ to achieve full capacity. We summarize in the following.

Theorem 1: Let $N = M_t T$. Any $\mathcal{X}$ such that $\mathcal{X}\mathcal{X}^H = 1/M_t \mathbf{I}_{M_t}$ is a capacity-optimal LDC. □

Therefore, capacity-optimal LDCs for $N = M_t T$ have a $\mathcal{X}$ that is simply a scaled unitary matrix. While this follows from (7), this solution is not as obvious from the capacity expressions in [1] due to the difference in the functional relationship therein.

Theorem 1 is of extreme importance in the design of capacity-optimal LDCs. First, it provides a sufficient and necessary condition to check if a set of codes to achieve full capacity. Second, it shows that there are an infinite number of such codes that are candidate capacity-optimized LDCs. This enables an improvement in another feature of the code, for example, the error probability performance, without a reduction in capacity advantage.

D. Frame-Based Linear Dispersion Codes

In many cases, it will be desirable to take $N < M_t T$ due to decoding complexity, memory, or latency constraints. To accommodate this scenario, we propose to modify the capacity optimal design, where $\mathcal{X}$ is a scaled unitary matrix by removing the appropriate number of columns and rescaling to give an $\mathcal{X}$ such that

$$\mathcal{X}^H \mathcal{X} = \frac{T}{N} \mathbf{I}_N. \tag{8}$$

A tall matrix $\mathcal{X}$ that satisfies this relationship is known as a *tight frame* [15]. A tight frame allows an overcomplete representation of a signal. For example, write $\mathbf{s} = N/T \mathcal{X}^H (\mathcal{X}\mathbf{s})$ to express $\mathbf{s}$ as a linear function of the columns of $\mathcal{X}^H$. The redundancy of the frame is the ratio of rows to columns: in this case $M_t T/N$. When $\mathcal{X}$ is square and orthogonal, the redundancy is 1. Large redundancy factors reduce the space spanned by the codewords and lower the overall data rate.

Since we do not explicitly optimize over the ergodic capacity, the resulting LDC design is not guaranteed to be capacity optimal. Therefore, it is of interest to determine how "far" these codes are from true capacity-optimal designs. For a given $\mathcal{X}$, using [2], the maximum ergodic capacity of the equivalent channel is given by

$$C_{c|\mathcal{X}} = \max_{\text{tr}(\mathbf{Q}) \leq N} \frac{1}{T} \mathcal{E}_{\mathbf{H}} \log\det \left(\mathbf{I}_{M_r T} + \frac{E_s}{N_o} \mathcal{H}\mathcal{X}\mathbf{Q}\mathcal{X}^H \mathcal{H}^H \right)$$

where the power constraint on $\mathbf{Q}$ is a result of the symbol energy normalization. Clearly, $C_{c|\mathcal{X}} \leq C_c \leq C_{M_r \times M_t}$.

For comparison purposes, it is not necessarily desirable to optimize over the input distribution. For example, in practice, the input symbols are uncorrelated, and $\mathbf{Q} = \mathbf{I}_N$ is a good assumption. In addition, $\mathbf{Q} = \mathbf{I}_N$ is the input distribution that maximizes the mutual information for the capacity-optimal LDC. In these cases, the performance of a given $\mathcal{X}$ can be evaluated in a natural way by the mutual information

$$I_{c|\mathcal{X}} := \frac{1}{T} \mathcal{E}_{\mathbf{H}} \log\det \left(\mathbf{I}_{M_r T} + \frac{E_s}{N_o} \mathcal{H}\mathcal{X}\mathcal{X}^H \mathcal{H}^H \right). \tag{9}$$

It is apparent that $I_{c|\mathcal{X}} \leq C_{c|\mathcal{X}}$.

Typically, when $N \leq M_t T$, there is a loss of ergodic capacity since $\mathcal{X}$ becomes tall, and it is not possible to excite all the modes of the effective $M_r T \times M_t T$ MIMO channel. We can show, however, that these codes have an ergodic capacity that grows asymptotically in proportion to N/T.

To show this, we use a result from in matrix theory known as the Poincare Separation Theorem [23, p. 190]. Let $\mathcal{B}$ be an arbitrary $M \times M$ Hermitian matrix, and let $\mathcal{X}$ be a $M \times N$ matrix with $N \leq M$ that satisfies $\mathcal{X}^H \mathcal{X} = \mathbf{I}_N$. The Poincare Separation Theorem says that the eigenvalues of $\mathcal{X}^H \mathcal{B} \mathcal{X}$ in decreasing order (μ_1 is the largest) satisfy the following set of inequalities:

$$\mu_{k+M-N}(\mathcal{X}^H \mathcal{B} \mathcal{X}) \leq \mu_k(\mathcal{B}) \leq \mu_k(\mathcal{X}^H \mathcal{B} \mathcal{X}). \tag{10}$$

Note that $\mathcal{H}^H \mathcal{H}$ is Hermitian, and let $\mathcal{B} = \mathcal{H}^H \mathcal{H}$. Let λ_k be the kth singular value of $\mathcal{H}$ (recall that $\mu_k(\mathcal{B}) = \lambda_k^2$). Now we can prove the following theorem.

Theorem 2: The mutual information achieved by using any frame-based $\mathcal{X}$ is bounded by

$$\frac{1}{T} \mathcal{E} \sum_{k=1}^{N} \log \left(1 + \frac{E_s}{M_t N_o} \lambda_{k+M_t T - N}^2 \right) \leq I_{c|\mathcal{X}} \leq \frac{1}{T} \mathcal{E} \sum_{k=1}^{N} \log \left(1 + \frac{E_s}{M_t N_o} \lambda_k^2 \right) \tag{11}$$

where λ_k is the kth singular value of $\mathcal{H}$, and the expectation is with respect to the distribution of the singular values. □

Theorem 2 shows that a frame-based $\mathcal{X}$ excites N of the $M_t T$ modes of the full channel, which is desirable since the code can excite at most N modes. It is easy to find nonframe-based $\mathcal{X}$ that excite less than N modes (take $\mathcal{X}$ to be the all ones matrix for example). In Fig. 2, we plot the upper and lower bounds for $M_t = M_r = 3$ and $N = 3, 4, 5, 6$. Note that the bounds are loose, which is expected since the best and worst case are low-probability events. At high SNR, however, each pair of upper and lower bounds have the same slope, confirming that each approach uses the full N modes. Due to the factor of $1/T$ in

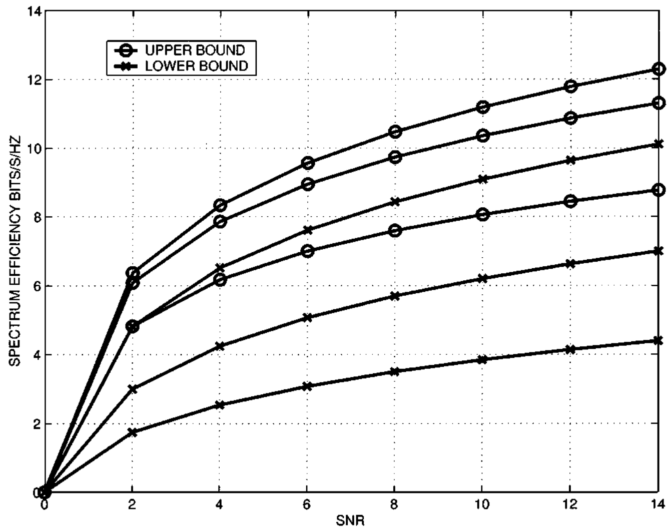

Fig. 2. Mutual information bounds for frame-based codes with $M_t = M_r = 3$ and $N = 4, 5, 6$. The curves increase with N.

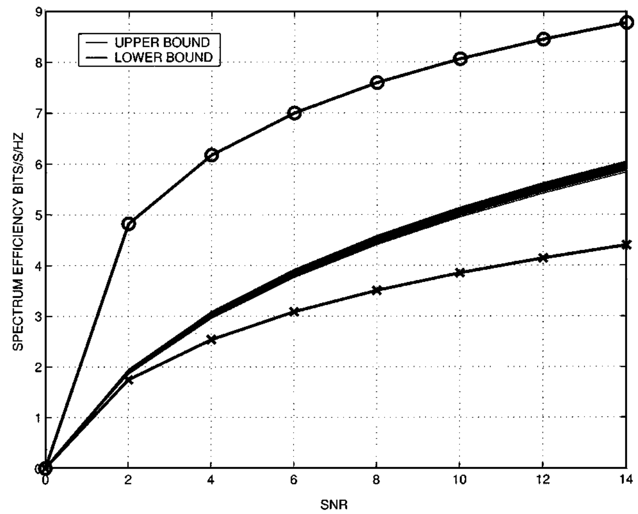

Fig. 3. Comparison of the mutual information bounds with the average mutual information bounds for various randomly chosen frame-based codes with parameters $M_t = M_r = 3, T = 2$, and $N = 4$.

(11), asymptotically the capacity grows in proportion to N/T for frame-based codes.

To get an idea how good or bad frame-based codes perform with respect to capacity, we performed the following experiment. We randomly generated a series of $\mathcal{X}$ for $M_t = M_r = 3$ and $N = 4$ and estimated the ergodic capacity for using 10 000 Monte Carlo simulations. We plot the curves for all 100 candidate codes as well as the upper and lower bounds in Fig. 3. Note that there is not much difference between the best and worst codes. Although the bounds are loose, the upper, lower, and proposed bounds appear to increase with the same rate.

IV. Error Probability-Based Linear Dispersion Codes

In this section, we refine our frame-based code design by incorporating error probability considerations. First, we show that error probability performance is not guaranteed by capacity maximization. Then, we review some criteria for evaluating space-time code performance in terms of error rate. Finally, we propose a linear dispersion code design in which the dispersion matrices are also optimized with respect to the rank and determinant space-time code design criteria.

A. Motivation to Include Error Probability

The frame-based code structure can guarantee full or nearly full capacity depending on the choice of N and T. It does not, however, guarantee good performance in terms of error probability [14]. To motivate consider the following example.

Example 1—Error Probability Comparison: Consider two different codes designed for $M_t = M_r = 2, T = 3$, and $N = 6$. The first code uses the linear transformation matrix

$$\mathcal{X} = \frac{1}{\sqrt{2}}\mathbf{I}_6.$$

The second code uses the linear transformation matrix shown in the equations at the bottom of the page, which were found using a numerical optimization to be described in the sequel. The capacity of the equivalent channel induced by each code is depicted in the left plot of Fig. 4, whereas the bit error rate performance, estimated over 25 000 Monte Carlo simulations, is displayed in the right plot of Fig. 4. □

Each code satisfies Theorem 1 and, thus, induces a channel with the full ergodic capacity. It can be shown, however, that the first code has a second-order diversity advantage, whereas the optimized code exhibits a fourth-order advantage for the same overall rate. Thus, codes with the same asymptotic performance

$$\mathcal{Y}_1 = \begin{bmatrix} 0.1367 & 0.2428 & 0.2521 \\ 0.1015 + j0.2424 & -0.0033 + j0.1009 & 0.2696 - j0.1559 \\ 0.3698 - j0.0297 & -0.2291 + j0.2090 & -0.1960 + j0.1943 \\ -0.2160 - j0.2385 & -0.3112 + j0.0926 & -0.1163 - j0.3345 \\ -0.3036 + j0.0386 & -0.3005 - j0.0676 & 0.2706 + j0.1926 \\ -0.2163 - j0.1749 & 0.2360 + j0.2805 & 0.0221 - j0.1648 \end{bmatrix}$$

$$\mathcal{Y}_2 = \begin{bmatrix} 0.2867 & 0.4655 & -0.2447 \\ -0.3853 + j0.0223 & 0.0568 + j0.1497 & -0.0123 + j0.3860 \\ 0.2500 - j0.0552 & -0.0720 + j0.0066 & -0.0667 + j0.3389 \\ 0.0591 - j0.1259 & 0.2264 + j0.3016 & -0.0495 + j0.0402 \\ 0.1883 + j0.2585 & 0.0857 - j0.1440 & 0.1944 + j0.1818 \\ 0.1083 + j0.2645 & -0.2754 - j0.0795 & -0.2609 + j0.1694 \end{bmatrix}$$

$$\mathcal{X} = [\mathcal{Y}_1 \mathcal{Y}_2]$$

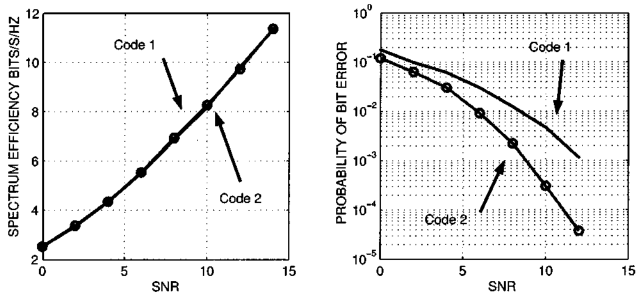

Fig. 4. Comparison of two different BPSK codes with $M_t = M_r = 2, T = 3$, and $N = 6$ in terms of ergodic capacity and probability of bit error.

in terms of ergodic capacity may require further optimization to achieve good error rate performance.

B. Codeword Error Probability

Based on Example 1, a complete code design also optimizes the codeword matrices based on the error rate. To solve this problem, we need to determine how a given LDC influences the error rate. The probability of codeword error can be upper bounded by the largest pairwise error probability using the union bound [24]. Therefore, we will use the maximum pairwise error probability as the performance metric.

For a general space-time code with perfect channel knowledge at the receiver, [6] and [7] use the Chernoff upper bound to derive an upper bound on the average probability that matrix codeword $\mathbf{S}$ is misdecoded as $\hat{\mathbf{S}}$. The average is taken with respect to the channel and the resulting expression is given by

$$P(\mathbf{S} \rightarrow \hat{\mathbf{S}}) \leq \frac{1}{\left|\mathbf{I}_{M_r M_t} + \frac{E_s}{4N_o}\mathbf{R}_s \otimes \mathbf{I}_{M_r}\right|} \qquad (12)$$

where $\mathbf{R}_s = (\mathbf{S} - \hat{\mathbf{S}})(\mathbf{S} - \hat{\mathbf{S}})^H$. Let λ_n be the nth eigenvalue of $\mathbf{R}_s$. Note that the eigenvalues of $\mathbf{R}_s \otimes \mathbf{I}_{M_r}$ are simply λ_n with multiplicity M_r. For high SNR, it is possible to rewrite (12) as

$$P\left(\mathbf{S}^{(m)} \rightarrow \mathbf{S}^{(k)}\right) \leq \frac{1}{\left(\frac{E_s}{4N_o}\right)^{\text{rank}(\mathbf{R}_s)M_r} \prod_{n=1}^{\text{rank}(\mathbf{R}_s)} \lambda_n^{M_r}}. \qquad (13)$$

The diversity advantage [25] of the code, which is the anticipated improvement in the slope of the probability of error curve, is determined by the smallest product $\text{rank}(\mathbf{R}_s)M_r$. The coding advantage, or shift in *SNR*, in the probability of error curve is determined by the smallest product $\prod_{n=1}^{\text{rank}(\mathbf{R}_s)} \lambda_n^{M_r}$, which, for full rank $\mathbf{R}_s$, is a function of the determinant of $\mathbf{R}_s$.

For LDCs, $\mathbf{R}_s$ is computed using the representation in (2). Let $\{s_n\}_{n=0}^{N-1}$ denote the transmitted sequence corresponding to $\mathbf{S}$ and $\{r_n\}_{n=0}^{N-1}$ denote the erroneous received sequence corresponding to $\hat{\mathbf{S}}$. The codeword difference matrix is then

$$\mathbf{R}_s = \left(\sum_{n=0}^{N-1} \mathbf{M}_n(s_n - r_n)\right)^* \times \left(\sum_{n=0}^{N-1} \mathbf{M}_n(s_n - r_n)\right)^T \qquad (14)$$

$$= \sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1} \mathbf{M}_{n_1}^* \mathbf{M}_{n_2}^T e_{n_1}^* e_{n_2} \qquad (15)$$

where $e_n := s_n - r_n$ is a difference between two constellation symbols. Essentially, the rank and determinant depend on all possible different linear combinations of the matrices $\mathbf{M}_l \mathbf{M}_m^H$, where the linear combinations are determined by products of the error. In terms of $\{\mathbf{X}_t\}_{t=0}^{T-1}$ and $\mathbf{e}^T = [e_0, e_1, \ldots, e_{N-1}]$., it can be shown (we skip the details) that

$$\mathbf{R}_s = \begin{bmatrix} \mathbf{e}^H \mathbf{X}_0^H \mathbf{X}_0 \mathbf{e} & \cdots & \mathbf{e}^H \mathbf{X}_0^H \mathbf{X}_{T-1} \mathbf{e} \\ \vdots & \ddots & \vdots \\ \mathbf{e}^H \mathbf{X}_{T-1}^H \mathbf{X}_0 \mathbf{e} & \cdots & \mathbf{e}^H \mathbf{X}_{T-1}^H \mathbf{X}_{T-1} \mathbf{e} \end{bmatrix}. \qquad (16)$$

Equation (16) shows how the subspaces of $\{\mathbf{X}_t\}_{t=0}^{T-1}$ play a role. Since, for high rate codes, N is typically greater than M_t, $\mathbf{X}_t$ will be fat and will have a null space. $\mathbf{R}_s$ is in principle Hermitian semi-definite. A necessary condition for $\mathbf{R}_s$ to be full rank is that every principle submatrix has a determinant that is greater than zero. Clearly, if $\mathbf{e}^H \mathbf{X}_t^H \mathbf{X}_t \mathbf{e} = \mathbf{0}$ for some error vector $\mathbf{e}$, then the resulting code is not full rank.

Based on the parameters of the code the diversity order of a linear code is bounded by the dimensions of the codeword matrices and is summarized in the following.

Theorem 3: The diversity order of a linear code is less than or equal to $M_r \min(M_t, T)$.

Proof: $\mathbf{R}_s$ is the product of a $M_t \times T$ matrix with its Hermitian; therefore, the rank of $\mathbf{I}_{M_r} \otimes \mathbf{R}_s$ is upper bounded by the minimum of M_t and T times the factor of M_r due to the Kronecker product. □

Fully diverse codes achieve equality in Theorem 3. A necessary condition, which is obvious from the structure of the error vectors, is illustrated in the following proposition.

Proposition 4: A fully diverse code has full-rank codeword matrices.

Proof: $\mathbf{R}_s$ should be full rank for all possible error sequences. One such set of error sequences are those with a single nonzero value, i.e., $e_n \neq 0$ and $e_m = 0$ $n \neq m$. For such a set of error sequences, $\mathbf{R}_s = \alpha \mathbf{M}_n^* \mathbf{M}_n^T$ (α is some scaling factor), and it is necessary that $\mathbf{M}_n$ is full rank if the code is fully diverse. □

The key message of Theorem 3 is that the diversity advantage of a linear code can be limited by designs that use short block lengths $T < M_t$. While it is tempting to always make T as large as possible, increases in T require corresponding increases in N to maintain the same rate of the code. Larger N introduces more possible error differences in (15), increasing difficulty in designing a fully diverse code. It also increases the decoding memory requirements, latency, and complexity. Thus, for linear codes, there is a fundamental tradeoff between the achievable diversity, capacity, and the resulting decoding complexity.

C. Rank and Determinant-Based Linear Dispersion Codes

We are interested in finding codes that have good performance in terms of ergodic capacity yet also provide a low error rate for small block lengths. The rank and determinant criteria require a search over all possible error vectors and is constellation specific. They do not *a priori* reveal a good design structure. The ergodic capacity, on the other hand, provides a rich structure that is constellation independent (the capacity calculation assumes Gaussian signals).

Unfortunately, as shown in Fig. 4, capacity-optimal LDCs do not necessarily optimize an error probability metric. Further, given the nonconvex optimization that must be solved to find a general capacity-optimal LDC, joint error rate and capacity optimization are difficult. Using our closed-form solution for $N = M_tT$, however, we can obtain capacity-optimal codes that also satisfy the rank and determinant criterion. When it is required that $N < M_tT$, we use the frame-based structure to obtain codes with both good error probability and good ergodic capacity performance.

Let a $(M_t, M_r, N, T, \mathcal{C})$ code be one designed for M_t transmit antennas, M_r receive antennas, a block length of T symbol periods, and N symbols from constellation $\mathcal{C}$. Assuming that the spectral efficiency of the constellation is R bits per symbol, the overall rate of this code will be NR/T bits per symbol period. The parameters M_t and T determine the dimension of the dispersion matrices, whereas N determines the number of such matrices. The parameters M_r and $\mathcal{C}$ will be used in the selection of the coefficients of the codeword matrices according to the rank and determinant criteria.

We summarize our design as follows.

Design Criterion 1—High SNR Near-Optimal LDCs: For a $(M_t, M_r, N, T, \mathcal{C})$ code, choose $\{\mathbf{M}_n\}_{n=0}^{N-1}$ to satisfy the tight frame relationship in (8). Within this class of codes, search for the design that maximizes the minimum rank and product of nonzero singular values. □

Codes designed according to Design Criterion 1 are substantially different than the capacity-optimal LDCs presented in [1]. To improve error probability performance, [1] imposes various side constraints that try to introduce additional structure in their LDCs that promote good error probability performance. For complex LDCs, the equivalent design constraints are (assuming that $T \le M_t$)

i) $\sum_{n=0}^{N-1} \mathrm{tr}\left\{\mathbf{M}_n\mathbf{M}_n^H\right\} = T$

ii) $\mathrm{tr}\left\{\mathbf{M}_n\mathbf{M}_n^H\right\} = \frac{T}{N}\; n = 0, 1, \ldots, N-1$

iii) $\mathbf{M}_n^H\mathbf{M}_n = \frac{1}{N}\,\mathbf{I}_T\; n = 0, 1, \ldots, N-1.$

The first constraint is simply the standard power constraint. The second constraint requires that each dispersion matrix contain the same average power. There is no obvious relationship between the first two constraints and error probability. The third constraint, however, produces codewords that satisfy the necessary condition in Proposition 4 for a code to achieve full diversity. It was found in [1] that codes designed according to constraint iii) often have good error probability performance, although it does not guarantee that performance.

Codes designed according to Design Criterion 1 differ in the following way. First, to the extent that the rank and determinant criterion is valid (at high SNR), our codes for $N = M_tT$ will be *both* capacity-optimal and error probability-optimal. Of course, we could modify Design Criterion 1 based on a better error performance metric, say the bit error rate, to obtain good bit error rate performance. We leave this to future work.

When $N < M_tT$, our codes are not capacity-optimal, although they still have good capacity properties. This is not a huge disadvantage since it is not easy to find the LDCs that globally maximize the capacity criterion in this case due to the nonconvex nature of the capacity cost function [1]. Further, our codes still have good performance, at least asymptotically, in terms of ergodic capacity, and they have the added bonus that they improve the error probability. No matter what the choice of N, note that our codes automatically satisfy side constraints i) and ii). Our design is flexible enough that we can evenincorporate constraint iii) *and still optimize over the rank and determinant criterion* to even further improve performance.

V. Numerical Techniques for Finding Good Codes

In this section, we present different techniques for finding dispersion matrices with good error probability performance. First, we describe how we can use results from frame theory to parameterize the proposed family of LDCs. Then, we propose a number of different procedures that can be used to obtain good coefficients for this parameterization.

A. Characterization of Frames

The benefit of using the proposed LDC structure is that we can rely on the rich literature on frame expansions [15] to assist in the selection of optimal codewords. In this section, we present two different parameterizations that will be useful in the sequel.

1) Projection: Given any full-rank $N \times M$ tall matrix $\mathbf{A}$

$$\mathcal{X} = \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1/2} \tag{17}$$

is a tight frame, as can be confirmed by checking

$$\mathcal{X}^H\mathcal{X} = (\mathbf{A}^H\mathbf{A})^{-H/2}\mathbf{A}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1/2} \tag{18}$$
$$= (\mathbf{A}^H\mathbf{A})^{-H/2}(\mathbf{A}^H\mathbf{A})^{H/2}(\mathbf{A}^H\mathbf{A})^{1/2} \times (\mathbf{A}^H\mathbf{A})^{-1/2} \tag{19}$$
$$= \mathbf{I}. \tag{20}$$

For an $M_tT \times N$ matrix $\mathbf{A}$, by scaling the power, $\mathcal{X}{=}\sqrt{T/N}\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1/2}$ forms a candidate linear transformation matrix that can be evaluated for pairwise error probability performance. This important result allows construction of candidate codeword matrices from any suitably sized arbitrary matrix $\mathbf{A}$. Let $\mathbf{A} = \mathbf{QR}$ be the QR-decomposition of $\mathbf{A}$, where $\mathbf{Q}$ is $M_tT \times N$, and $\mathbf{R}$ is $N \times N$. It is easy to check that $\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1/2} = \mathbf{Q}$; thus, computing $\mathcal{X}$ requires computing the QR decomposition of $\mathbf{A}$. Computational complexity and implementation issues are detailed in [26].

While it is useful to construct a tight frame given a random matrix, for the purpose of optimization, it is useful to be able to parameterize the family of tight frames. This can be accomplished by using analogous results for parameterizing unitary matrices. The space of unitary matrices is known in the mathematics literature as the Stiefel manifold, and a good summary of representations can be found in [27].

2) Unitary-Based Parameterization: It is possible to construct an $M_tT \times N$ tight frame from an $M_tT \times M_tT$ unitary matrix $\mathbf{U}$ by taking the first N columns of $\mathbf{U}$ as

$$\mathcal{X} = \mathbf{UZ}$$

where

$$\mathbf{Z} = \begin{bmatrix} \mathbf{I}_N \\ \mathbf{0}_{M_tT-N\times N} \end{bmatrix}. \tag{21}$$

This representation can characterize any tight frame since given a tight frame $\mathcal{X}$, an $\mathcal{X}$ can be found (for example, by augmenting with a random matrix of dimension $M_tT \times M_tT - N$ and using the Gram–Schmidt procedure) such that

$$\mathbf{U} = [\mathcal{X} \quad \hat{\mathcal{X}}]$$

is square and unitary. Of course, the representation is not unique since any permutations of the columns of $\hat{\mathcal{X}}$ also leads to a possible $\mathbf{U}$.

Since parameterizations of $\mathbf{U}$ exist in the literature, they can be used in turn to specify $\mathcal{X}$. In this paper, we use the fact that any unitary matrix can be written as a product of Householder reflections to develop an equivalent representation for tight frames. Let $\mathbf{v}^{(m)}$ for $m = 1, 2, \ldots, M_tT$ be a length M_tT vector of the form

$$\mathbf{v}^{(m)} = \left[0, \ldots, 0, 1, v_1^{(m)}, \ldots, v_{M_tT-m}^{(m)}\right]^T. \quad (22)$$

Then, the corresponding Householder reflection is

$$\mathbf{V}_m := \mathbf{I}_{M_tT} - \frac{2\mathbf{v}^{(m)}\mathbf{v}^{(m)H}}{\|\mathbf{v}^{(m)}\|^2}. \quad (23)$$

We can represent any $M_tT \times M_tT$ unitary matrix as

$$\mathbf{U} = \mathbf{D}\mathbf{V}_1 \cdots \mathbf{V}_{M_tT} \quad (24)$$

where $\mathbf{D}$ is a diagonal matrix of arbitrary complex exponentials. To make $\mathcal{X}$ tall, we use $\mathbf{Z}$ in (21)

$$\mathcal{X} = \sqrt{\frac{T}{N}}\mathbf{D}\mathbf{V}_1\mathbf{V}_2 \cdots \mathbf{V}_{M_tT}\mathbf{Z}. \quad (25)$$

When $N < M_tT$, we can reduce the number of terms in (25) by recognizing that $\mathbf{v}^{(m)}\mathbf{Z} = 0$ for $m > N$; thus, $\mathbf{V}_m\mathbf{Z} = \mathbf{Z}$ for $m > N$. Therefore, any potential frame-based code matrix can be written

$$\mathcal{X} = \sqrt{\frac{T}{N}}\mathbf{D}\mathbf{V}_1\mathbf{V}_2 \cdots \mathbf{V}_N\mathbf{Z}. \quad (26)$$

The total number of real parameters necessary to specify $\mathcal{X}$ is

$$M_tT + 2\sum_{k=M_tT-N}^{M_tT-1} k \;=\; M_tT + 2NM_tT - N - N^2 \quad (27)$$

which can be significantly less than the $2M_tTN$ real parameters required to specify an arbitrary complex $M_tT \times N$ matrix. Computing each $\mathbf{V}_m$ requires $2(M_tT - m - 1)^2$ real multiplies to compute the outer product and the product with the scaling factor, $(M_tT - m - 1)$ multiplies to compute the scaling factor $\|\mathbf{v}^{(m)}\|^2$, and one division. Ignoring the structure of $\mathbf{V}_m, \mathcal{X}$ can be computed using $(N+1)M_tT \times M_tT$ matrix multiplies.

The Householder representation is nice because it reduces the number of parameters of the search space and because it provides an unconstrained representation of $\mathcal{X}$ in terms of the coefficients of the Householder matrices. Other representations are also possible, for example, using the Given's rotation [26]. A good summary of these representations for unitary and orthogonal matrices is available in [27].

B. Algorithms for Finding Candidate Error-Probability-Optimal LDCS

In this section, we provide some algorithms for computing dispersion matrices that satisfy the frame-constraint and are further optimized with respect to the rank and determinant criteria.

1) Global Optimization: For a given N, M_t, and T, the linear dispersion matrices are exactly determined by $\mathcal{X}$. Thanks to our frame-based design, $\mathcal{X}$ has a parsimonious parameterization using Householder reflections. Therefore, obtaining an error-probability-optimal LDC (in the rank and determinant sense) requires optimizing the rank and determinant criterion with respect to the constellation.

Consider a $(M_t, M_r, N, T, \mathcal{C})$ code. Let $\mathcal{E} = \{\mathbf{s}_k - \mathbf{s}_l | \mathbf{s}_k \in \mathcal{S}, \mathbf{s}_l \in \mathcal{S}, k \neq l\}$ be the set of all possible 2^{RN} $N \times 1$ error vectors where $R = |\mathcal{C}|$. For any error vector $\mathbf{e}_k \in \mathcal{E}$, the error difference matrix $\mathbf{R}_s(\mathbf{e}_k)$ is determined from (15) (we have changed the notation to clarify that the error difference matrix is a function of a particular error vector).

The optimization corresponding to Design Criterion 1 that produces codes that are optimal with respect to the rank and determinant criteria is summarized in the following.

Optimization I: *Let* $\hat{\mathbf{D}}$ *and* $\{\hat{\mathbf{v}}^{(m)}\}_{m=1}^N$ *be an initial set of Householder parameters with the structure in (22). Let* $K \leq \min(M_t, T)$ *be the minimum required rank of the LDC. Find* $\{\mathbf{v}^{(m)}\}_{m=1}^N$ *such that*

$$J(\mathcal{X}) = \min_{\mathbf{e}\in\mathcal{E}} \prod_{i=0}^{K-1} \lambda_i(\mathbf{R}_s(\mathbf{e})) \quad (28)$$

is maximized subject to the constraint

$$\forall\, \mathbf{e} \in \mathcal{E} \text{ rank } \mathbf{R}_s(\mathbf{e}) \geq K. \quad (29)$$

A target minimum rank of $K \leq \min(M_t, T)$ fixes the minimum diversity order achieved by the code. If a code cannot be found with a minimum rank K, then the optimization is infeasible. Practically, it is best to start with small K and gradually increase the size until the optimization terminates with an infeasible solution. This ensures that the maximum diversity advantage will be achieved with the resulting code.

The optimization can be implemented with any number of numerical methods, for example, a gradient search. The complexity of any numerical method is high since calculating $J(\mathcal{X})$ requires computing the SVD of $2^{RN}(2^{RN} - 1)/2$ error covariance matrices. Clearly, this is prohibitive for larger rates (as is the ML decoder). The overall complexity depends on the exact algorithm chosen for implementation but is quite high since most algorithms require a numerical estimate of the gradient, which is obtained by perturbing the cost function, in addition to cost function evaluation at each step in the optimization.

Unfortunately, convergence of Optimization 1 is not readily guaranteed. First, the cost function is a nonlinear function of the Householder parameters. Second, it is nonconvex since it maximizes the minimum taken over a discrete space. Thus, a solution obtained using Optimization 1 will typically be the best rank and determinant solution in the neighborhood of the initial condition. This is not a serious drawback since the capacity-optimal LDCs in [1] do not consider any such optimizations with respect to error probability.

To improve the performance of Optimization 1, it is desirable to find a good initial condition.

2) Basis Selection: An extremely simple procedure to obtain an initial condition for Optimization 1 is to choose an $M_tT \times M_tT$ unitary matrix and pick the N columns whose corresponding codeword matrices are the best in terms of Design Criterion 1.

Basis Selection: *Let $\mathcal{I}$ be the set of all possible subsets of N columns of Q. By construction, $|\mathcal{I}| = C_N^{M_t T}$, where $C_k^n = n!/k!(n-k)!$. Let $\mathcal{X}_i$ denote the matrix formed by the ith subset in $\mathcal{I}$. The basis selection algorithm chooses $\mathcal{X} = \sqrt{T/N}\mathcal{X}_i$ such that the codeword matrices constructed from the columns of $\mathcal{X}_i$ maximize $J(\mathcal{X})$ subject to the side constraint in (29).*

The coefficients of the Householder representation corresponding to the Basis Selection optimization can be found as a byproduct of the QR decomposition. See [26] for details. Basis selection has the advantage of low complexity since it requires computing $J(\mathcal{X})$ only for $n!/k!(n-k)!$ possible combinations of columns. Of course, it is not helpful when $N = TM_t$ since there is only one possible combination of columns.

3) Random Search: An alternative to improve sampling of the space of frames is to use a random search. This has the advantage of providing a better sampling of the space of possible frame-based LDCs and typically provides good initial conditions for Optimization 1. In fact, codes obtained using the random search often perform well without further optimization. Of course, since the search is random, they do not guarantee even a local maximum of the determinant cost function.

One way to perform a random search is to use the projection-based representation in (17) to generate L arbitrary $M_t T \times N$ matrices $\mathbf{A}$ and to use the projection in (17) to construct a corresponding set of $\mathcal{X}$ matrices. The $\mathcal{X}$ that maximizes $J(\mathcal{X})$ subject to the rank side constraint is chosen as the initial condition for Optimization 1.

Random Search I: *Generate L realizations of $\mathbf{A}$ from some distribution, for example, the multivariate complex Gaussian distribution. From all the possible $\mathbf{A}$ matrices, select $\mathcal{X} = \sqrt{T/N}\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1/2}$ such that $\mathcal{X}$ maximizes $J(\mathcal{X})$ subject to the side constraints in (29).*

The coefficients for Householder representation can again be found as a byproduct of the QR decomposition. See [26] for details.

For comparison with codes from [1] it is often useful to enforce the side constraint iii). How we enforce iii) depends on the choice of M_t. For example, if $M_t > T$, we choose $\mathbf{M}_n^H\mathbf{M}_n = 1/N\mathbf{I}_T$, whereas for $T \leq M_t$, we choose $\mathbf{M}_n\mathbf{M}_n^H = T/NM_t\mathbf{I}_{M_t}$. This additional side constraint has been shown to also improve the performance of frame-based LDCs in [28].

To impose side constraints, we propose the following modification to Random Search I.

Random Search II: *Consider L candidate realizations of a random matrix $M_t T \times N$ matrix $\mathbf{A}$, and let ϵ be some stopping value. For each candidate matrix $\mathbf{A}$, let $\mathcal{X} = \sqrt{T/N}\,\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1/2}$. Then, extract $\{\mathbf{M}_n\}_{n=0}^{N-1}$ from $\mathcal{X}$, and let $\hat{\mathbf{M}}_n = \mathbf{M}_n(\mathbf{M}_n^H\mathbf{M}_n)^{-1/2}$ and scale appropriately. Repeat the procedure for $\mathcal{X} = [\text{vec}(\hat{\mathbf{M}}_0), \text{vec}(\widehat{\mathbf{M}}_1), \ldots, \text{vec}(\widehat{\mathbf{M}}_{N-1})]$ until $\left\|\mathcal{X}^H\mathcal{X} - T/N\,\mathbf{I}_N\right\| < \epsilon$. Choose the $\mathcal{X}$ generated from the L* realizations of $\mathbf{A}$ *that maximizes $J(\mathcal{X})$ subject to the side constraints in (29).*

Again, the coefficients for Householder representation can again be found as a byproduct of the QR decomposition. Note that we are alternately enforcing a convex constraint with respect to the dispersion matrices, and thus, the iteration in Random Search II typically converges quickly.

VI. Design Examples

In this section, we present a number of example code designs obtained through joint capacity and error probability optimization. We compare performance in terms of ergodic capacity, codeword error probability, and bit error probability. Gray labeling was assumed for the bit-to-symbol mapping.

A. Comparison of Linear Dispersion Codes for Four-Transmit Antenna One-Receive Antenna

In this example, we compare two (4, 1, 4, 4, 4—*QAM*) linear dispersion code designs. The reference LDC is capacity-optimal and was derived in [1]. The proposed frame-based code was generated using Random Search II with a search over 100 000 realizations and the side constraint that the rank of $\mathbf{R}_s$ should be 3. The linear transformation matrix for the proposed code is shown in the equation at the bottom of the page.

$$\mathcal{X} = \begin{bmatrix}
-0.0360 - j0.0140 & -0.1156 - j0.4268 & -0.0583 - j0.1780 & -0.0659 - j0.1163 \\
0.0048 + j0.4032 & 0.0929 + j0.0153 & -0.0975 - j0.1916 & -0.0508 + j0.1726 \\
-0.0246 + j0.0043 & 0.1657 - j0.1040 & -0.1332 + j0.2814 & -0.3343 + j0.0493 \\
-0.0394 - j0.2894 & 0.0796 + j0.0313 & -0.1593 - j0.2154 & -0.0463 + j0.2889 \\
0.1847 + j0.0048 & 0.0880 - j0.0752 & -0.0380 - j0.0685 & -0.2601 - j0.3587 \\
0.1370 - j0.1306 & -0.1296 + j0.0497 & -0.3499 + j0.2283 & 0.1243 - j0.0701 \\
-0.3409 - j0.0637 & -0.0824 - j0.3047 & -0.0908 - j0.0130 & 0.0320 - j0.1437 \\
0.1476 - j0.1949 & -0.3386 - j0.0556 & 0.2237 - j0.1041 & 0.0864 - j0.0646 \\
-0.1233 + j0.1169 & 0.1143 + j0.1568 & 0.2468 - j0.3387 & -0.0693 + j0.0554 \\
-0.1089 - j0.1982 & 0.2081 - j0.0851 & -0.0235 + j0.0812 & -0.2386 + j0.2902 \\
0.0808 - j0.2558 & 0.1048 - j0.2848 & 0.1622 - j0.1432 & 0.0890 - j0.1766 \\
-0.1831 - j0.2540 & 0.0930 + j0.2470 & -0.1391 - j0.0329 & 0.2460 - j0.0362 \\
-0.3397 - j0.2648 & -0.0574 - j0.0096 & -0.0025 + j0.1820 & -0.0800 - j0.1468 \\
0.0196 + j0.0093 & -0.3445 - j0.2295 & -0.1465 - j0.0248 & -0.1780 + j0.1563 \\
-0.2037 - j0.1250 & 0.1368 + j0.0360 & -0.1406 - j0.2795 & 0.0511 + j0.2691 \\
0.0082 + j0.0825 & -0.0188 + j0.2344 & -0.2826 - j0.1305 & -0.2044 - j0.2216
\end{bmatrix}.$$

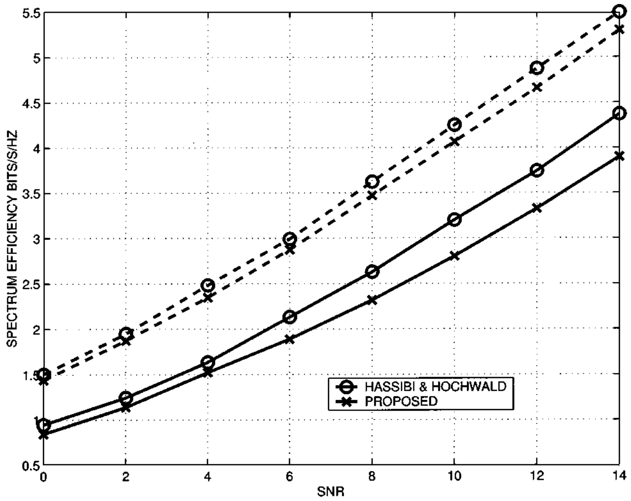

Fig. 5. Comparison of the two different (4, 1, 4, 4, 4—*QAM*) linear dispersion codes in terms of spectrum efficiency for one and two receive antennas. The Hassibi and Hochwald code is capacity optimal, whereas the proposed code is a noncapacity optimal frame-based code.

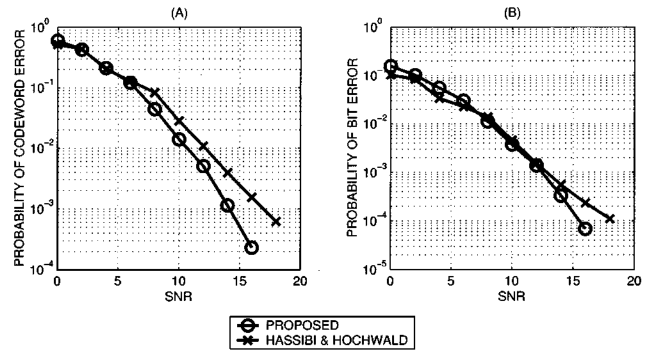

Fig. 6. Comparison of the two different (4, 1, 4, 4, 4—*QAM*) linear dispersion codes in terms of (a) codeword error probability and (b) bit error probability. The proposed code has been optimized for good error probability performance.

The singular values for the $\mathbf{R}_s$ with worst determinate are (2.1141, 1.6267, 0.9405, 0.0050). As designed, the codeword error matrix has a rank of three so it would be expected that this code exhibits a third-order diversity advantage in this example since $M_r = 1$.

The details for the Hassibi–Hochwald code can be found in [1]. The singular values for the $\mathbf{R}_s$ matrix with worst rank and determinant are (1.7899,1.6719, 0.0279, 0.0183). Whereas $\mathbf{R}_s$ has four nonzero singular values, practically speaking, it has a rank of two since the other two singular values are quite small. Thus, we would expect this code to exhibit only a second order diversity advantage.

We illustrate the ergodic capacity of the equivalent channel induced by each linear dispersion code in Fig. 5. As expected, the capacity of the frame-based code is close but does not quite maximize the ergodic capacity. This is because in this case, $N < M_t T$, and our design does not guarantee capacity optimality. We could have chosen larger N, but then, this would not have been a fair comparison. The performance difference is small, particularly as the number of receive antennas increases.

Next, we illustrate the error rate performance in Fig. 6. The left plot displays the codeword error rate for each code. Note that, as designed, the optimized code has a higher order diversity advantage as reflected by the improvement in the slope of the error rate curve. Even at a codeword error rate of 10^{-3}, the proposed code has a 3-dB advantage. This advantage grows with larger SNR due to the larger diversity advantage of the proposed code.

The corresponding bit error rate curves are displayed on the right plot of Fig. 6. Again, the slope of the optimized curve is improved over that of the unoptimized curve, with the difference becoming more dramatic at high SNR. At a BER of only 10^{-4}, there is again a significant 3–dB difference in the performance of the two codes. This difference grows as the BER decreases due to the larger diversity advantage of the proposed codes.

In both plots, note that there is negligible improvement at low SNR. This is because the rank and determinant criteria are really only good predictors of code performance at high SNR. To obtain better low SNR performance, an alternative error probability criterion should be employed in the optimization. In practice, it is the average error rate that determines the performance of a given space-time code. Thus, despite the fact that the proposed code is not capacity optimal, it is a better LDC than that proposed in [1] in the practical error rate comparison.

B. Linear Dispersion Codes for Three-Transmit Antennas Three-Receive Antennas

In this example, we compare two code different jointly capacity-optimal LDCs that are local minima of Optimization 1.

First, we consider the (3, 3, 3, 1, 4—*QAM*) code with linear transformation matrix

$$\mathcal{X}_1 = \frac{1}{\sqrt{3}}\mathbf{I}_3.$$

Next, we consider the (3, 3, 9, 3, 4—*QAM*) code with linear transformation matrix

$$\mathcal{X}_3 = [\,\mathcal{X}_a \quad \mathcal{X}_b\,]$$

where we have (30) and (31), shown at the bottom of the next page. The second code was designed using Optimization I with a desired rank parameter of 2.

For the first code, the worst $\mathbf{R}_s$ is rank one with singular values (1.1549,0,0). Therefore, we expect the first code to have a $1 * 3 =$ third-order diversity advantage since there are three receive antennas. For the second code, the worst R_s is rank two with minimum singular values (1.1549, 1.1549,0). Therefore, we expect this code to provide a $2 * 3 = 6$th-order diversity advantage since there are three receive antennas.

As displayed in Fig. 7, each code achieves the same (in this case the full) ergodic channel capacity. In Fig. 8, however, there is significant difference between the two codes in terms of bit error rate. The $T = 1$ code obtains a maximum third-order diversity advantage, whereas the $T = 3$ code obtains a dramatic sixth order advantage. The difference is significant, providing, even at a BER of 10^{-3}, an advantage of 3 dB. This illustrates that the benefit of increasing T without decreasing the overall rate is a potential improvement in diversity. Of course, the tradeoff is an increase in decoding complexity since $T = 3$ code requires joint decoding of nine symbols instead of three symbols.

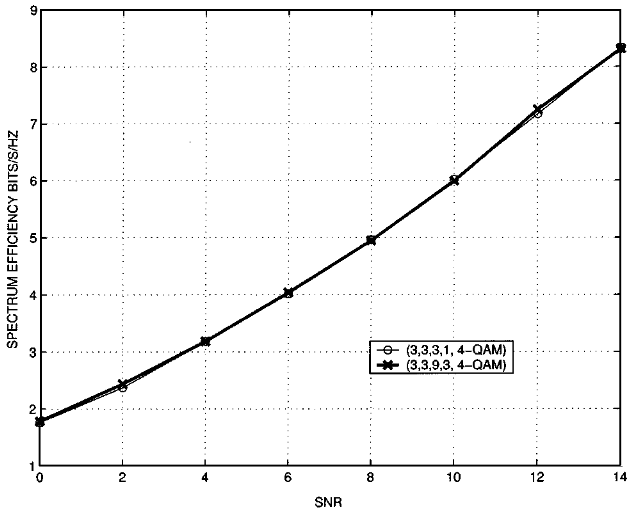

Fig. 7. Ergodic capacity comparison of two different frame-based optimized LDCs with different block lengths but the same rate.

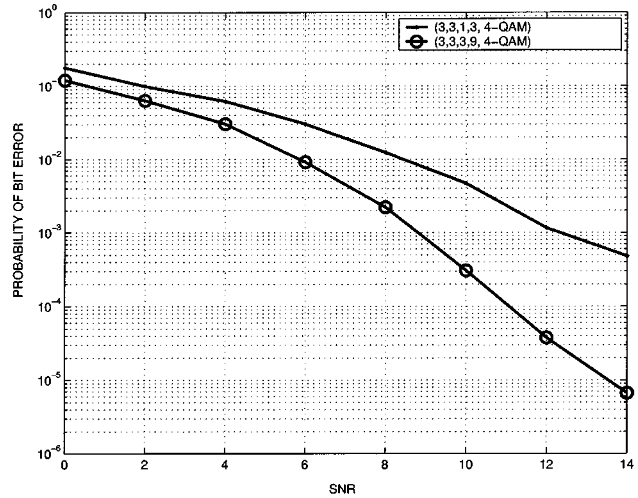

Fig. 8. Bit error probability comparison of two different frame-based optimized LDCs with different block lengths but the same rate.

C. Linear Dispersion Codes for Three-Transmit Antennas and One-Receive Antennas

In this example, we examine the impact of increasing T while keeping N fixed. The proposed frame-based codes were generated using Optimization 1 with maximum possible rank and Random Search II with a search over 1000 realizations to provide the initial conditions. The error rate was estimated using 100 000 Monte Carlo simulations.

For the (3, 1, 3, 1, 4—*QAM*) code, we used the spatial multiplexing LDC. Note that this code is capacity optimal. For the (3, 1, 3, 2, 4—*QAM*) code, we found the second equation shown at the bottom of the page, whereas for the (3, 1, 3, 2, 4—*QAM*) code, we found the equation shown at the bottom of the next page.

In Fig. 9, we plot both the ergodic capacity and the codeword error probability. Unlike the previous example, where the ratio N/T was fixed and capacity was preserved, in this example, the

$$\mathcal{X}_a = \begin{bmatrix} 0.0014 - j0.0000 & 0.3333 + j0.0000 & -0.0013 + j0.0000 & 0.0003 - j0.0000 & -0.0002 + j0.0000 \\ 0.0134 + j0.1767 & 0.0003 - j0.0019 & 0.0866 - j0.2688 & 0.0860 - j0.2695 & 0.1454 - j0.0997 \\ -0.0888 + j0.2680 & 0.0003 - j0.0003 & -0.1148 + j0.1348 & -0.1149 + j0.1336 & 0.2772 - j0.0566 \\ -0.2250 + j0.1711 & 0.0002 - j0.0010 & -0.1341 - j0.1142 & -0.0333 + j0.1736 & -0.2251 + j0.1711 \\ -0.0983 + j0.1129 & 0.1713 + j0.1247 & -0.1287 - j0.0782 & 0.1310 - j0.0718 & -0.0480 - j0.1419 \\ 0.0847 - j0.0403 & 0.0993 - j0.2374 & -0.1404 - j0.1941 & 0.2387 - j0.0236 & -0.0070 + j0.0933 \\ 0.0893 - j0.1524 & -0.0015 + j0.0007 & -0.2690 - j0.0880 & 0.0588 + j0.2764 & 0.0888 - j0.1526 \\ -0.0779 + j0.2263 & -0.0993 - j0.2374 & 0.0918 + j0.0157 & -0.0584 + j0.0732 & -0.1581 - j0.1805 \\ 0.0993 - j0.1129 & 0.1713 - j0.1247 & 0.1274 + j0.0782 & -0.1317 + j0.0718 & 0.0471 + j0.1419 \end{bmatrix} \quad (30)$$

$$\mathcal{X}_b = \begin{bmatrix} -0.0005 - j0.0000 & 0.3333 - j0.0000 & -0.0002 - j0.0000 & 0.3333 - j0.0000 \\ 0.1456 - j0.0998 & -0.0002 + j0.0002 & -0.2762 + j0.0600 & 0.0001 - j0.0004 \\ 0.2770 - j0.0566 & -0.0001 - j0.0002 & 0.1737 + j0.0324 & 0.0002 + j0.0001 \\ -0.0350 - j0.2805 & -0.0002 - j0.0002 & -0.1340 - j0.1157 & -0.0002 - j0.0004 \\ -0.0997 + j0.1118 & -0.3333 + j0.0027 & 0.1315 - j0.0717 & 0.1618 - j0.1282 \\ 0.0844 - j0.0406 & 0.0013 - j0.0044 & 0.2382 - j0.0242 & -0.1013 + j0.2414 \\ 0.0886 + j0.1529 & -0.0001 + j0.0002 & -0.2685 - j0.0876 & 0.0002 - j0.0003 \\ -0.0766 + j0.2272 & -0.0013 - j0.0044 & -0.0585 + j0.0734 & 0.1013 + j0.2414 \\ 0.0995 - j0.1118 & -0.3333 - j0.0027 & -0.1319 + j0.0717 & 0.1618 + j0.1282 \end{bmatrix}. \quad (31)$$

$$\mathcal{X} = \begin{bmatrix} 0.3003 - j0.1273 & -0.2303 - j0.1625 & 0.3651 - j0.1192 \\ 0.1390 + j0.3748 & -0.2530 - j0.0734 & -0.2446 - j0.2106 \\ 0.2137 + j0.1467 & 0.3677 + j0.2220 & 0.1218 - j0.2585 \\ -0.0699 - j0.3776 & 0.2833 - j0.1616 & -0.2783 + j0.0445 \\ -0.2410 + j0.2147 & 0.1819 + j0.3560 & -0.2324 + j0.1240 \\ -0.2401 - j0.1549 & 0.2294 + j0.1207 & 0.4183 - j0.0977 \end{bmatrix}$$

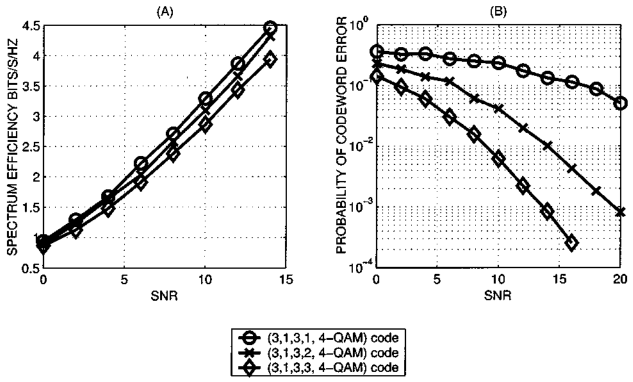

Fig. 9. Comparison of three different frame-based LDCs in terms of (a) mutual information and (b) codeword error rate.

capacity decreases for larger T since N/T is decreasing. This rate reduction, however, comes with a much more substantial improvement in codeword error rate since larger T allows for the error difference matrix to have a larger rank.

VII. Conclusion

We have introduced a new linear dispersion code design for MIMO Rayleigh fading channels that provides codes with good ergodic capacity and error probability performance. The basis matrices of the code, when suitably rearranged, yield a tight frame. This gives the intuition that instead of sending uncoded symbols, the proposed frame-based codes convey the coefficients of a frame expansion of a vector of symbols. The tight frame-based structure provides a closed-form solution that can achieve the full capacity of the channel with appropriate choice of dimensions.

For a given set of parameters, we showed that there are many choices of frame-based codes that have a similar ergodic capacity performance. Therefore, we were able to use existing parameterizations of the space of tight frames to search for the code that provided the lowest probability of symbol error. This is important because, as we demonstrated, codes that have the same ergodic capacity performance may have different error performance. To ensure good error rate performance, we optimized over the space of tight frames to find the one with the best performance in terms of the rank and determinant criterion. We discussed an algorithm to accomplish this optimization and presented some alternatives to generate good initial conditions. Monte Carlo simulations demonstrated the performance of various proposed codes for different numbers of transmit and receive antennas, constellation sizes, and block lengths.

While we considered the Rayleigh channel that is uncorrelated in space and time, practical MIMO systems will have channels with space, time, as well as frequency selectivity [29]–[31]. Future work is necessary to extend the results presented herein to these practical channels. Further, our designs are optimal only for the ML receiver. Further work is necessary to extend this to other low-complexity receivers, for example, the iterative receiver [4] or the linear receiver [32].

Acknowledgment

The authors would like to thank Prof. H. Boelcskei for his useful comments.

References

[1] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," in *Proc. Allerton Conf. Commum., Cont. Comput.*, Oct. 2000.
[2] I. E. Telatar, "Capacity of Multiantenna Gaussian Channels, Tech. Rep., AT&T Bell Labs. Internal Tech. Memo. 1995," Eur. Trans. Telecommun., 1999.
[3] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
[4] C. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
[5] R. W. Heath Jr and A. Paulraj, "Switching between spatial multiplexing and transmit diversity based on constellation distance," in *Proc. Allerton Conf. Commun. Cont. Comput.*, Oct. 2000.
[6] J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 4, pp. 527–537, Apr. 1999.
[7] V. Tarokh, N. Seshadri, and R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.
[8] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1451–1458, Oct. 1998.
[9] G. Ganesan and P. Stoica, "Space-time block codes: A maximum SNR approach," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1650–1656, May 2001.
[10] A. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception (DTDR)," U. S. #5 345 599, Sept. 1994.
[11] B. A. Bjerke and J. C. Proakis, "Multiple-antenna diversity techniques for transmission over fading channels," in *Wireless Commun. Networking Conf.*, vol. 3, Sept. 1999, pp. 1038–1042.
[12] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1456–1467, July 1999.

$$\mathcal{X} = \begin{bmatrix} 0.1960 + j0.1412 & 0.2942 + j0.3268 & 0.1579 + j0.2381 \\ -0.2860 - j0.2102 & -0.1274 + j0.1802 & 0.3982 + j0.0062 \\ -0.1191 + j0.3671 & -0.0557 - j0.2969 & 0.2215 + j0.2098 \\ 0.0607 - j0.2792 & -0.1607 + j0.1804 & -0.4257 - j0.1103 \\ -0.2731 - j0.2899 & 0.2387 - j0.2627 & 0.0069 - j0.2206 \\ -0.0039 - j0.3051 & -0.3528 - j0.1565 & 0.2102 + j0.2170 \\ 0.4198 + j0.1308 & -0.1904 + j0.2130 & 0.2057 + j0.1267 \\ -0.0176 + j0.2200 & 0.0949 + j0.3868 & -0.3005 - j0.1890 \\ -0.1995 - j0.2269 & -0.1469 + j0.2674 & -0.0923 + j0.3748 \end{bmatrix}.$$

[13] S. Sandhu and A. Paulraj, “Space-time block coding: A capacity perspective,” *IEEE Commun. Letters*, vol. 4, pp. 384–386, Dec. 2000.
[14] R. W. Heath Jr, H. Bölcskei, and A. Paulraj, “Space-time signaling and frame theory,” in *Proc. IEEE ICASSP*, May 2001, pp. 2445–2448.
[15] I. Daubechies, *Ten Lectures on Waveles*. Philadelphia, PA: SIAM, 1992.
[16] R. R. Hammons and K. El Gamal, “On the theory of space-time codes for PSK modulation,” *IEEE Trans. Inform. Theory*, vol. 46, pp. 524–542, Mar. 2000.
[17] Y. Mohasseb and M. Fitz, “A 3D spatio-temporal simulation model for wireless channels,” in *Proc. Int. Conf. Commun.*, June 2001.
[18] L. Zhiqiang and G. B. Giannakis, “Space-time coding with transmit antennas for multiple access regardless of frequency-selective multipath,” in *Proc. Sensor Array Multichannel Signal Process. Workshop*, 2000, pp. 178–182.
[19] R. W. Heath Jr, “Space-time signaling in multi-antenna systems,” Ph.D. dissertation, Stanford Univ., Stanford, CA, 2001.
[20] O. Damen, A. Chkeif, and J.-C. Belfiore, “Lattice code decoder for space-time codes,” *IEEE Commun. Lett.*, vol. 4, pp. 161–163, May 2000.
[21] B. M. Hochwald and S. T. Brink, “Iterative list sphere decoding to attain capacity on a multi-antenna link,” in *Proc. Allerton Conf. Commun., Cont., Comput.*, Oct. 2001.
[22] H. Vikalo and B. Hassibi, “Low-complexity iterative decoding over multiple antenna channels using a modified sphere decoder,” in *Proc. Allerton Conf. Commun., Cont., Comput.*, Oct . 2001.
[23] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[24] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1995.
[25] A. Paulraj, “Diversity methods,” in *CRC Handbook on Mobile Communications*, J. Gibson, Ed. Boca Raton, FL: CRC, 1995, vol. 12, pp. 166–176.
[26] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
[27] B. Hassibi and B. Hochwald. Cayley differential unitary space-time codes. [Online]. Available: http://mars.bell-labs.com/cm/ms/what/mars/index.html
[28] R. W. Heath Jr and A. Paulraj, “Capacity maximizing linear space-time codes,” in *Proc. Int. Symp. Signals, Syst., Electron.*, Tokyo, Japan, July 24–27, 2001.
[29] D. S. Baum, D. Gore, R. Nabar, S. Panchanathan, K. V. S. Hari, V. Erceg, and A. J. Paulraj, “Measurement and characterization of broadband MIMO fixed wireless channels at 2.5 GHz,” in *IEEE Int. Conf. Pers. Wireless Commun.*, 2000, pp. 203–206.
[30] C. C. Martin, J. H. Winters, and N. R. Sollenberger, “Multiple-input multiple-output (MIMO) radio channel measurements,” in *Proc. IEEE Veh. Technol. Conf.*, vol. 2, 2000, pp. 774–749.
[31] A. L. Swindlehurst, G. German, J. Wallace, and M. Jensen, “Experimental measurements of capacity for MIMO indoor wireless channels,” in *Proc. IEEE Third Workshop Signal Process. Adv. Wireless Commun.*, 2001, pp. 30–33.
[32] R. W. Heath Jr and A. Paulraj, “Antenna selection for spatial multiplexing systems based on minimum error rate,” in *Proc. Int. Conf. Commun.*, vol. 7, June 2001, pp. 2276–2280.

Robert W. Heath, Jr. (S’96–M’01) received the B.S. and M.S. degrees from the University of Virginia, Charlottesville, in 1996 and 1997, respectively, and the Ph.D. degree from Stanford University, Stanford, CA, in 2002, all in electrical engineering.

From 1998 to 1999, he was a Senior Member of Technical Staff at Iospan Wireless Inc. (formerly Gigabit Wireless Inc.), San Jose, CA. From 1999 to 2001, he served as a Senior Consultant for Iospan Wireless Inc. In January 2002, he joined the Electrical and Computer Engineering Department, The University of Texas, Austin, where he serves as an Assistant Professor as part of the Wireless Networking and Communications Group. His research group, the Wireless Systems Innovations Laboratory, focuses on the theory, design, and practical implementation of wireless systems. His current research interests are coding, modulation, equalization, and resource allocation for MIMO wireless communication systems.

Arogyaswami J. Paulraj (F’91) received the Ph.D. degree from the Naval Engineering College and the Indian Institute of Technology, Delhi, India, in 1973.

He has been a Professor with the Department of Electrical Engineering, Stanford University, Stanford, CA, since 1993, where he supervises the Smart Antennas Research Group. This group consists of approximately a dozen researchers working on applications of space-time signal processing for wireless communications networks. His research group has developed many key fundamentals of this new field and helped shape a worldwide research and development focus onto this technology. His nonacademic positions included Head, Sonar Division, Naval Oceanographic Laboratory, Cochin, India; Director, Center for Artificial Intelligence and Robotics, Bangalore, India; Director, Center for Development of Advanced Computing; Chief Scientist, Bharat Electronics, Bangalore; and Chief Technical Officer and Founder, Iospan Wireless Inc., San José, CA. He has also held visiting appointments at Indian Institute of Technology, Delhi, India; Loughborough University of Technology, Loughborough, U.K.; and Stanford University. He sits on several board of directors and advisory boards for U.S. and Indian companies/venture partnerships. His research has spanned several disciplines, emphasizing estimation theory, sensor signal processing, parallel computer architectures/algorithms, and space-time wireless communications. His engineering experience includes development of sonar systems, massively parallel computers, and, more recently, broadband wireless systems.

Dr. Paulraj has won several awards for his engineering and research contributions. These include two President of India Medals, CNS Medal, Jam Medal, Distinguished Service Medal, Most Dististinguished Service Medal, VASVIK Medal, IEEE Best Paper Award (Joint), amongst others. He is the author of over 250 research papers and holds eight patents. He a Member of the Indian National Academy of Engineering.