Design of Linear Dispersion Codes: Asymptotic Guidelines and Their Implementation

Ramy H. Gohary and Timothy N. Davidson, *Member, IEEE*

***Abstract*—In this paper, a design method is developed for the class of linear-dispersion (LD) codes—a diverse set of space-time codes that subsumes several standard designs. The development begins by showing that for systems that employ a large number of transmit antennas, LD codes constructed from unitary coding matrices are asymptotically optimum from different design perspectives, viz., minimum mean square error (MMSE), mutual information, and average pairwise error probability (PEP). Those measures have a direct impact on the detection complexity, data rate, and error performance that a space–time code can achieve. Using the insight generated by the asymptotic result, a structured design technique for the LD coding matrices, that suits a broad class of configurations is provided. The resulting codes can support high data rates and provide performance advantages over current designs when decoded with a standard detector. Based on the asymptotic results, a row interleaving scheme is proposed, and it is shown to result in significant performance enhancement.**

***Index Terms*—Efficient detection, ergodic capacity, multiple-input–multiple-output (MIMO) communication systems, space-time coding, Stiefel manifold.**

I. Introduction

Wireless communication systems with multiple antennas at both the transmitter and receiver have the potential to provide reliable transmission at high data rates [1]. In particular, for a sufficiently rich scattering environment, the capacity of a coherent communication channel employing M transmit and N receive antennas grows linearly as $\min(M, N)$, [1]–[3]. The design of coding schemes for multiple-antenna systems operating at high signal-to-noise ratios (SNRs) involves a tradeoff between the achievable rate (normalized by $\log(\text{SNR})$) at which the system capacity grows and the normalized rate at which the error probability decays [3]. This inherent tradeoff provides a distinction between schemes like the orthogonal space–time block codes (OSTBCs) [4] that sacrifice achievable rate for maximum reliability, and the Bell Labs layered space-time (BLAST) schemes (e.g., [5]) that can support rates close to channel capacity, but do so without benefiting from the diversity of the channel. Since OSTBCs and BLAST achieve the extremes of the tradeoff, there is considerable interest in developing design methods for schemes that provide intermediate performance in terms of achievable rate and error probability, and are applicable for a broad range of antenna configurations. The set of linear-dispersion (LD) codes [6] is a diverse class of space-time codes that subsumes many existing schemes, including the OSTBC and BLAST schemes, and hence, is a natural framework in which such design problems can be posed. While some recently developed codes, such as the threaded algebraic space time (TAST) codes [7] and those in [8], possess many desirable performance features, they remain proper subsets of the LD framework. Given the generality of the LD coding framework and the abundance of degrees of design freedom it possesses, the focus of this paper is on the development and implementation of guidelines for the design of LD codes through the study of their asymptotic properties.

The original LD codes in [6] were designed to maximize the ergodic capacity of the system. However, it has recently been pointed out that such capacity-optimal LD codes do not necessarily perform well in practice [9]. Moreover, the maximization of the ergodic capacity is performed under an implicit assumption that maximum-likelihood (ML) detection will be performed at the receiver—a task that requires an exhaustive search that is often computationally infeasible. These observations prompt the search for codes that jointly achieve high data rates and perform well when only a suboptimal detector is available at the receiver. In the present paper, we identify a class of LD codes that approach optimality from both performance and rate perspectives as the number of transmit antennas M grows. In particular, the class of codes presented herein asymptotically ensures minimum output mean-square error (MSE) without incurring any information loss. Minimizing the MSE is a desirable feature for the class of suboptimum sphere detectors that employ a linear front end [10]. In addition to this property, we show that this class of codes asymptotically minimizes the high-SNR average pairwise error probability (PEP).

Since we consider both capacity and performance in our designs, the underlying philosophy of our approach is similar to that followed in the concurrent development of frame-based LD codes [9]. However, our methodology is substantially different. In particular, we derive a different objective function for the code-design problem, and the set of feasible codes in [9] intersects that of the codes proposed herein, but neither is a proper subset of the other. As we will show in Section VI, our design method can generate codes that provide slightly better performance than those in [9]. Furthermore, our codes

Manuscript received October 31, 2003; revised July 29, 2004; accepted November 28, 2004. The editor coordinating the review of this paper and approving it for publication is H. Li. This work was supported in part by the National Science and Engineering Research Council of Canada. The work of T. N. Davidson was supported in part by the Canada Research Chairs program.

A preliminary version of this paper appeared in the *Proc. IEEE SP Workshop on Signal Processing Advances in Wireless Communications*, Rome, June 2003.

The authors are with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: davidson@mcmaster.ca).

Digital Object Identifier 10.1109/TWC.2005.858296

possess a structure that enables us to search the space of feasible codes efficiently without the need to perform the random search techniques employed in [9].

The design criteria for our LD codes are based on several observations. First, we show that LD codes with a certain unitary structure simultaneously: 1) minimize a lower bound on the MSE at the output of a linear preprocessing stage in the receiver; 2) minimize a lower bound on the high-SNR average PEP; and 3) maximize an upper bound on the mutual information that an LD code can achieve. We then show that as the number of transmit antennas grows, these optimized bounds are achieved. Guided by this asymptotic result, we impose this unitary structure on finite-sized LD codes and show how the resulting equivalent channel matrix can be optimized. Imposing this unitary structure dramatically reduces the number of design parameters, which subsequently reduces the computational cost of determining the codes. In particular, our design procedure is substantially simpler than that in [6] and does not require the additional manipulations employed therein. We will demonstrate via simulation that even for systems with a small number of antennas, our codes perform better than many existing codes, including the standard LD codes [6], the TAST codes [7] and the codes in [8].

The high-SNR performance of a communication system is dominated by the outage probability [3], which is the probability that a specific channel realization fails to support the required data rate. In order to reduce the outage probability, we propose a row interleaving procedure that improves the high SNR performance of the code by transmitting the rows of the code matrix over different channel realizations. Asymptotic analysis and supporting simulations demonstrate how our unitarily structured LD codes can benefit from row interleaving. In fact, we provide simulation results that demonstrate that row interleaving can substantially increase the performance advantage of our codes over the original LD codes.

II. LD Codes

We consider a system with M transmit and N receive antennas by which one wishes to transmit Q data symbols, $s_1, \dots, s_Q$, from a given constellation $\mathcal{C}$ over T time slots. If we represent the channel symbols transmitted from the antennas at each time slot as the rows of a $T \times M$ matrix S, then each LD codeword can be written in the form [6]

$$S = \sum_{q=1}^{Q} (\alpha_q A_{2q-1} + j\beta_q A_{2q}) \tag{1}$$

where α_q and β_q are the real and imaginary parts of s_q, respectively. The matrices A_q are $T \times M$ fixed coding matrices that define the code. Typically, they are normalized so that the average power transmitted by each antenna is unity; i.e.,

$$\mathrm{E}_s \left\{ \mathrm{Tr}(SS^{\mathrm{H}}) \right\} = TM \tag{2}$$

where $(\cdot)^{\mathrm{H}}$ denotes the Hermitian transpose. If r denotes the constellation size of the data symbols s_q, then the transmission rate of the code is $R = (Q/T)\log_2 r$ bits/channel use.

We will adopt the standard frequency-flat block-fading model of the channel [4], [6] in which the slow fading of the propagation coefficients is approximated by coefficients that are constant over a given block of T time slots and then fade independently in the next block. This can be a model for many transmission strategies including frequency hopping, ideally interleaved time-division multiple access (TDMA), or packet-based transmission in which each frame of data sees an independent realization of the channel but the channel is constant within each frame [11]. Under the assumed model, the received symbol matrix Y can be written as

$$Y = \sqrt{\frac{\rho}{M}} SH + V \tag{3}$$

where ρ is the SNR per receive antenna, and H is the $M \times N$ channel matrix whose (m, n)th element is the complex gain between the mth transmit and the nth receive antennas. We will assume that the scattering is rich enough for the channel coefficients $[H]_{mn}$ to be modeled as independent complex circular Gaussian random variables of unit variance. The matrix V in (3) denotes the additive white Gaussian noise. Its elements are also independent complex circular Gaussian random variables of unit variance. Since we have assumed that the channel changes slowly, we will assume that the receiver has been able to acquire an accurate model of the channel. By rearranging the elements in the matrices in (3), we can rewrite that equation as [6]

$$\underbrace{\begin{bmatrix} y_{R,1} \\ y_{I,1} \\ \vdots \\ y_{R,N} \\ y_{I,N} \end{bmatrix}}_{y} = \sqrt{\frac{\rho}{M}} \mathcal{H} \underbrace{\begin{bmatrix} \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_Q \\ \beta_Q \end{bmatrix}}_{s} + \underbrace{\begin{bmatrix} v_{R,1} \\ v_{I,1} \\ \vdots \\ v_{R,N} \\ v_{I,N} \end{bmatrix}}_{v} \tag{4}$$

where, the subscripts $(\cdot)_{R,i}$ and $(\cdot)_{I,i}$ denote the real and imaginary parts of the ith column, respectively. The $2NT \times 2Q$ equivalent-channel matrix $\mathcal{H}$ is given by

$$\mathcal{H} = [I_N \otimes \mathcal{A}_1 \quad I_N \otimes \mathcal{A}_2 \quad \cdots \quad I_N \otimes \mathcal{A}_{2Q}][I_{2Q} \otimes \underline{h}] \tag{5}$$

where $\otimes$ denotes the Kronecker product, and

$$\mathcal{A}_{2q-1} = \begin{bmatrix} A_{R,2q-1} & -A_{I,2q-1} \\ A_{I,2q-1} & A_{R,2q-1} \end{bmatrix}$$
$$\mathcal{A}_{2q} = \begin{bmatrix} -A_{I,2q} & -A_{R,2q} \\ A_{R,2q} & -A_{I,2q} \end{bmatrix}. \tag{6}$$

In (5), $\underline{h} = [h_1^{\mathrm{T}}, \dots, h_N^{\mathrm{T}}]^{\mathrm{T}}$, where $h_n = [h_{R,n}^{\mathrm{T}}, h_{I,n}^{\mathrm{T}}]^{\mathrm{T}}$, and $h_{R,n}$ and $h_{I,n}$ are, respectively, the real and imaginary component vectors of the nth column of the physical channel matrix H.

It can be seen from (5) that the columns of $\mathcal{H}$ are linearly transformed versions of each other. Therefore, unless the coding matrices A_q are carefully chosen, it may be quite likely that the resulting $\mathcal{H}$ is "close" to being degenerate. In that event, the performance of most detectors deteriorates substantially. Hence, an effective design technique for the coding matrices ought to guard against these occurrences, in addition to optimizing other design criteria. A candidate technique for doing so is provided in Section IV.

III. Performance Bounds and Asymptotic Optimality

In this section, we derive the design criteria for our LD codes, and demonstrate the asymptotic optimality of codes with a certain unitary structure. It was shown in the previous section that the structure of LD codes allows us to explicitly manipulate the equivalent-channel matrix (5). A question that arises is how to choose the coding matrices A_q so that the equivalent-channel matrix ensures good average receiver performance at a given data rate. Obviously, this design problem depends on the receiver structure, the number of available antennas at each end of the communication link, the block length T, and the choice of the number of symbols Q in a block.

Optimum signal detection involves maximization of the likelihood function over the discrete set of the code alphabet—an NP-hard integer least-squares problem; e.g., [12]. For an ML receiver to detect the transmitted symbols reliably at high SNR, it is sufficient for the singular values of the equivalent-channel matrix to be bounded away from zero in order to avoid ambiguity errors produced by system degeneracy [3]. However, it is not immediately clear what desirable conditions the equivalent-channel matrix has to meet should we choose to use a less computationally expensive receiver. For example, consider the version of the sphere detector in [12] and [13]. For the vectorized model in (4), this detector performs the QR decomposition of the equivalent-channel matrix $\mathcal{H}$, rotates y, and performs a tree-structured search over (a certain subset) of the space of all possible transmitted vectors $\mathcal{C}^Q$ to determine

$$\hat{s} = \arg \min_{s_q \in \mathcal{C}'_q \subseteq \mathcal{C}} \|\tilde{y} - \mathcal{R}s\| \tag{7}$$

where $\mathcal{H} = \mathcal{Q}\mathcal{R}$ and $\tilde{y} = \mathcal{Q}^{\mathrm{T}}y$. An alternative to sphere detection is the "nulling and cancellation" approach [2], [5], [14] in which decisions on the elements of s in (7) are made sequentially and are fed back to remove interference from subsequent decisions. Some interesting insight into our design approach (described below) can be gained by observing from (7) that if the columns of $\mathcal{H}$ can be made orthogonal, both the search in (7) and the "nulling and cancellation" approach reduce to simple detection of the individual entries of $\tilde{y}$, i.e., the entries of $\tilde{y}$ are decoupled. Having such a column orthogonal $\mathcal{H}$ dramatically reduces the search space of the sphere detector and avoids the potential for error propagation in the nulling-and-cancellation approach. While we do not deal with the column orthogonality of $\mathcal{H}$ directly, our design approach (implicitly) generates equivalent-channel matrices $\mathcal{H}$ with "increasingly" orthogonal columns as the size of the system grows. More precisely, $\mathcal{H}\mathcal{H}^{\mathrm{T}}$ approaches a scaled identity for almost every channel realization.

A. Linear Preprocessing and Mean Square Error

Another approach used in sphere detection linearly processes y to form $\tilde{s} = Gy$, and then performs an ML search in the neighborhood of $\tilde{s}$ [10]. Typically, G is chosen as the "zero-forcing" equalizer,[1] $G_{\mathrm{ZF}} = \mathcal{H}^{\dagger}$, or the linear minimum-mean-square-error (MMSE) equalizer, $G_{\mathrm{MMSE}} = \sqrt{\rho/M}\mathcal{H}^{\mathrm{T}}((\rho/M)\mathcal{H}\mathcal{H}^{\mathrm{T}} + I_{2NT})^{-1}$. To ensure good performance at reasonable complexity from this type of sphere detector, one should design the coding matrices A_q in (1) to ensure that $\tilde{s}$ is, on average, "close" to s. For a given channel H, the distance between $\tilde{s}$ and s can be measured in terms of the MSE. For the case where $G = G_{\mathrm{MMSE}}$

$$\begin{aligned} \mathrm{MSE} &= \mathrm{E}_{s,v}\left\{\mathrm{Tr}\left((s-\tilde{s})(s-\tilde{s})^{\mathrm{T}}\right)\right\} \\ &= Q - NT + \frac{1}{2}\mathrm{Tr}\left(\left(I_{2NT} + \frac{\rho}{M}\mathcal{H}\mathcal{H}^{\mathrm{T}}\right)^{-1}\right). \end{aligned} \tag{8}$$

We will show later (in Section III-D) that it is desirable to have Q as large as possible (from both symbol-rate and asymptotic-performance perspectives). However, the MSE expression (8) reinforces the standard practice that Q should be chosen to be no greater than NT [6].

Since we have assumed that the transmitter does not know the channel matrix H, an appropriate design strategy would be to choose the coding matrices A_q embedded in $\mathcal{H}$ in order to minimize the expected MSE over the distribution of H. That is

$$\min_{A_q} \quad \mathrm{E}_H\left\{\mathrm{Tr}\left(\left(I_{2NT} + \frac{\rho}{M}\mathcal{H}\mathcal{H}^{\mathrm{T}}\right)^{-1}\right)\right\} \tag{9a}$$

$$\text{subject to} \quad \mathrm{Tr}\left(A_q A_q^{\mathrm{H}}\right) \le \frac{TM}{Q}. \tag{9b}$$

The constraint in (9b) not only ensures that the coding matrices meet the overall power constraint in (2), it also guarantees that each data symbol satisfies the same bound on its transmitted power. The set of matrices $\{A_q\}$ satisfying (9b) is smaller than that satisfying (2), and hence, enforcing (2) rather than (9b) may result in improved performance (e.g., [15]). However, we will show in Section III-D and Appendix II that any performance loss due to (9b) vanishes as M increases. In fact, the satisfaction of (9b) is a necessary condition for the optimized bounds derived below to be asymptotically achieved.

The problem in (9) can be quite awkward to solve directly because the objective involves an expectation and the design parameters enter $\mathcal{H}$ in a highly structured fashion. Rather than attempting to solve (9) directly, we now develop bounds on the objective function and show that, by exploiting the asymptotic properties of the bounds, we can obtain "good" LD codes in a relatively straightforward manner. The key to this development is Jensen's inequality [16].

Lemma 1: If f is a strictly convex function and X is a random variable, then $\mathrm{E}\{f(X)\} \ge f(\mathrm{E}\{X\})$ with equality iff $X = \mathrm{E}\{X\}$ with probability 1.

[1] Here, $(\cdot)^{\dagger}$ denotes the Moore-Penrose pseudoinverse.

The function $\mathrm{Tr}((I+X)^{-1})$ is strictly convex in X over $\{X|X=X^{\mathrm{T}}\succeq 0\}$. Therefore, using Jensen's inequality

$$\mathrm{E}_H\left\{\mathrm{Tr}\left(\left(I_{2NT}+\frac{\rho}{M}\mathcal{H}\mathcal{H}^{\mathrm{T}}\right)^{-1}\right)\right\} \geq \mathrm{Tr}\left(\left(I_{2NT}+\frac{\rho}{M}\mathrm{E}_H\{\mathcal{H}\mathcal{H}^{\mathrm{T}}\}\right)^{-1}\right). \quad (10)$$

We now attempt to find coding matrices A_q that minimize the lower bound in (10). We will show in Section III-D that the minimized lower bound is asymptotically achieved.

Observe that

$$\frac{1}{M}\mathcal{H}\mathcal{H}^{\mathrm{T}} = \frac{1}{M}\sum_{q=1}^{2Q}(I_N\otimes\mathcal{A}_q)\underline{h}\,\underline{h}^{\mathrm{T}}\left(I_N\otimes\mathcal{A}_q^{\mathrm{T}}\right) \quad (11)$$

where $\mathcal{A}_q$ is defined in (6), and $\underline{h}$ in (5). Using (11), we have

$$\mathrm{E}_H\{\mathcal{H}\mathcal{H}^{\mathrm{T}}\} = \frac{1}{2}\left(I_N\otimes\sum_q\mathcal{A}_q\mathcal{A}_q^{\mathrm{T}}\right) \quad (12)$$

and hence, the lower bound in (10) can be written as

$$\mathrm{Tr}\left(\left(I_{2NT}+\frac{\rho}{2M}\left(I_N\otimes\sum_q\mathcal{A}_q\mathcal{A}_q^{\mathrm{T}}\right)\right)^{-1}\right). \quad (13)$$

For any positive definite matrix X, we have that [17]

$$\mathrm{Tr}(X^{-1}) \geq \sum_i\frac{1}{[X]_{ii}}$$

with equality holding if and only if X is diagonal. Therefore, when minimizing the lower bound in (13) we can restrict our attention to the case where $\sum_q\mathcal{A}_q\mathcal{A}_q^{\mathrm{T}}$ is diagonal. In order to find the matrices $\mathcal{A}_q$ that minimize the right-hand side of (10), we solve the following problem

$$\min_{\mathcal{A}_q}\quad \sum_i\frac{1}{1+\frac{\rho}{2M}\left[\sum_q\mathcal{A}_q\mathcal{A}_q^{\mathrm{T}}\right]_{ii}} \quad (14a)$$

$$\text{subject to}\quad \mathrm{Tr}\left(\mathcal{A}_q\mathcal{A}_q^{\mathrm{T}}\right)\leq\frac{2TM}{Q}. \quad (14b)$$

By differentiating the Lagrangian function of (14) and setting it to zero, we find that the optimal $\mathcal{A}_q$'s satisfy

$$\left[\sum_q\mathcal{A}_q\mathcal{A}_q^{\mathrm{T}}\right]_{ii} = 2M \qquad \forall i. \quad (15)$$

Hence, the optimal choice of $\{\mathcal{A}_q\}$ is one for which

$$\sum_q\mathcal{A}_q\mathcal{A}_q^{\mathrm{T}} = 2MI_{2T}. \quad (16)$$

With this choice of $\mathcal{A}_q$, the inequality in (10) becomes

$$\mathrm{E}_H\left\{\mathrm{Tr}\left(\left(I_{2NT}+\frac{\rho}{M}\mathcal{H}\mathcal{H}^{\mathrm{T}}\right)^{-1}\right)\right\} \geq \frac{2NT}{(1+\rho)}. \quad (17)$$

In the case when we have $M\geq T$, the matrices $\mathcal{A}_q$ are square or "fat," and a simple choice that satisfies (16) and the power constraint on each transmitted symbol is given by

$$\mathcal{A}_q\mathcal{A}_q^{\mathrm{T}} = \frac{M}{Q}I_{2T}. \quad (18)$$

In order to satisfy (18), the matrices A_q have to be unitary. That is, $A_qA_q^{\mathrm{H}}=(M/Q)I_{\mathrm{T}}$ for $1\leq q\leq Q$. When we have $M<T$, we cannot satisfy the restricted optimality condition in (18), but if $T\leq 2MQ$, it is possible to satisfy the general optimality condition in (16), as we will show in Example 2 in Section VI.

B. Mutual Information

In the previous section, we considered code design from a performance perspective with a certain class of suboptimum detectors in mind, and we argued that coding matrices with a unitary structure minimize a lower bound on the MSE. In this section, we discuss the status of these codes from a mutual information perspective.

If s assumes the capacity achieving standard circular Gaussian distribution, the model in (4) implies that the mutual information between transmitted and received signals is given by [1], [6]

$$I(s;y) = \frac{1}{2T}\mathrm{E}_H\left\{\log\det\left(I_{2NT}+\frac{\rho}{M}\mathcal{H}\mathcal{H}^{\mathrm{T}}\right)\right\}. \quad (19)$$

Since $\log\det(I+X)$ is strictly concave in X over $\{X|X=X^{\mathrm{T}}\succeq 0\}$, it follows from Jensen's inequality (Lemma 1) that

$$\frac{1}{2T}\mathrm{E}_H\left\{\log\det\left(I_{2NT}+\frac{\rho}{M}\mathcal{H}\mathcal{H}^{\mathrm{T}}\right)\right\} \leq \frac{1}{2T}\log\det\left(I_{2NT}+\frac{\rho}{M}\mathrm{E}_H\{\mathcal{H}\mathcal{H}^{\mathrm{T}}\}\right). \quad (20)$$

We will now show that the coding matrices A_q that maximize this upper bound possess the same unitary structure as those that minimized the lower bound on the MSE. In Section III-D, we will show that such coding matrices asymptotically achieve the maximized upper bound.[2]

Using the expression in (12), to maximize the upper bound in (20) we need to solve the following optimization problem

$$\max\log\det\quad\left(I_{2NT}+\frac{\rho}{2M}\left(I_N\otimes\sum_q\mathcal{A}_q\mathcal{A}_q^{\mathrm{T}}\right)\right) \quad (21a)$$

$$\text{subject to}\qquad \mathrm{Tr}\left(\mathcal{A}_q\mathcal{A}_q^{\mathrm{T}}\right)\leq\frac{2TM}{Q}. \quad (21b)$$

Using Hadamard's inequality, we have that for any positive definite matrix X

$$\det(X)\leq\prod_i X_{ii} \quad (22)$$

[2]The upper bound in (20) and the observation that it is maximized by unitary coding matrices appeared independently in [18]. In Section III-D, we will enhance that result by showing that unitary coding matrices not only maximize the upper bound, but also asymptotically achieve it.

with equality if and only if X is diagonal. By restricting our attention to the set of matrices $\{\mathcal{A}_q\}$ for which $\sum_q \mathcal{A}_q\mathcal{A}_q^{\mathrm{T}}$ is diagonal, and setting the derivative of the Lagrangian of (21) equal to zero, we find that the optimal $\mathcal{A}_q$'s must satisfy

$$\left[\sum_q \mathcal{A}_q\mathcal{A}_q^{\mathrm{T}}\right]_{ii} = 2M.$$

Hence, the optimal choice of the set $\{\mathcal{A}_q\}$ is one for which

$$\sum_q \mathcal{A}_q\mathcal{A}_q^{\mathrm{T}} = 2MI_{2T}. \tag{23}$$

It is clear from (23) that the coding matrices that minimize the lower bound of the MSE also maximize the upper bound on the mutual information that can be achieved by an LD code. Using (23) and (20), we can write the maximized bound on the mutual information as

$$I(s;y) \le N\log(1+\rho). \tag{24}$$

In Appendix I, we show that the right-hand side of (24) is not only a bound on the maximum mutual information that an LD code can achieve, but is also an upper bound on the ergodic channel capacity. Furthermore, we show that as the number of transmit antennas grows, the ergodic channel capacity approaches the right-hand side of (24). In Section III-D, we will show that as M (and T) grows, the actual mutual information achieved by an LD code that satisfies (23) also approaches the bound in (24), and hence, it approaches the ergodic capacity of the channel.

C. Pairwise Error Probability (PEP)

In this section, we point out that the codes that simultaneously optimize the bounds on the MSE and mutual information derived in Sections III-A and B also have desirable properties in terms of a measure of the PEP. The measure that we use is the average PEP for the case where the symbol vectors s are drawn from a Gaussian codebook (cf., [6]), where the average is taken over the channel realizations and all possible codeword pairs of the codebook. More specifically, if $P(s \to s'|\mathcal{H})$ denotes the probability that, conditioned on a specific equivalent channel realization $\mathcal{H}$, the receiver selects the codeword s', given that s was sent, then the average PEP is defined to be [6]

$$\mathrm{PEP}_{\mathrm{av}} = \mathrm{E}_H\left\{E_{s,s'}\left\{P(s\to s'|\mathcal{H})\right\}\right\} \tag{25}$$

where s and s' are drawn from a Gaussian codebook.[3] For the model in Section II, it has been shown that [6]

$$E_{s,s'}\left\{P(s\to s'|\mathcal{H})\right\} = \frac{1}{4\pi}\int_{-\infty}^{\infty}\frac{d\omega}{\omega^2+\frac{1}{4}}\prod_{q=1}^{2Q}\frac{1}{\sqrt{1+\frac{2\rho\sigma_q\left(\omega^2+\frac{1}{4}\right)}{M}}} \tag{26}$$

where $\sigma_q \ge \sigma_{q+1}$ denote the eigenvalues of $\mathcal{H}\mathcal{H}^{\mathrm{T}}$. If we define $r = \mathrm{Rank}(\mathcal{H})$, then when the SNR is sufficiently high, we have $\rho\sigma_r/M \gg 2$. In that case, the unity term under the square root in (26) can be neglected and the integral can be evaluated analytically to obtain

$$E_{s,s'}\{P(s\to s'|\mathcal{H})\} \le \eta_r\left(\frac{\rho}{M}\right)^{-\frac{r}{2}}\prod_{q=1}^{r}\sigma_q^{-\frac{1}{2}} \tag{27}$$

where

$$\eta_r = \frac{1}{4\pi}\int_{-\infty}^{\infty}\frac{d\omega}{\left(\omega^2+\frac{1}{4}\right)^{1+\frac{r}{2}}} = \begin{cases}\frac{(2r-1)(2r-3)\cdots(1)}{r!}, & r \text{ even}\\ \frac{2^{-\frac{r}{2}}}{32\pi}\sum_{m=0}^{r-\frac{1}{2}}\binom{\frac{r-1}{2}}{m}\frac{(-1)^m}{2m+1}, & r \text{ odd}\end{cases}.$$

If $r = \mathrm{Rank}(\mathcal{H}) = \min\{2Q, 2NT\}$, then the product term in (27) can be expressed as

$$\prod_{q=1}^{r}\sigma_q^{-\frac{1}{2}} = \begin{cases}\det(\mathcal{H}\mathcal{H}^{\mathrm{T}})^{-\frac{1}{2}}, & Q \ge NT\\ \det(\mathcal{H}^{\mathrm{T}}\mathcal{H})^{-\frac{1}{2}}, & Q \le NT\end{cases} \tag{28}$$

where we have used the fact that for $1 \le q \le r$, $\sigma_q(\mathcal{H}\mathcal{H}^{\mathrm{T}}) = \sigma_q(\mathcal{H}^{\mathrm{T}}\mathcal{H})$, and the fact that $\det(I_{2NT}+\mathcal{H}\mathcal{H}^{\mathrm{T}}) = \det(I_{2Q}+\mathcal{H}^{\mathrm{T}}\mathcal{H})$. Combining (25), (27), and (28), we obtain

$$\mathrm{PEP}_{\mathrm{av}} \le \begin{cases}\eta_{2NT}\left(\frac{\rho}{M}\right)^{-NT}\mathrm{E}_H\left\{\det(\mathcal{H}\mathcal{H}^{\mathrm{T}})^{-\frac{1}{2}}\right\}, & Q \ge NT\\ \eta_{2Q}\left(\frac{\rho}{M}\right)^{-Q}\mathrm{E}_H\left\{\det(\mathcal{H}^{\mathrm{T}}\mathcal{H})^{-\frac{1}{2}}\right\}, & Q \le NT\end{cases}. \tag{29}$$

The bound in (29) is tight if $r = \mathrm{Rank}(\mathcal{H}) = \min\{2Q, 2NT\}$ with probability 1, and is trivial otherwise.

Our goal in this section is to minimize $\mathrm{PEP}_{\mathrm{av}}$ or, alternatively, to minimize the upper bound in (29), provided that the latter is tight. As in Sections III-A and III-B, we will not minimize the upper bound in (29) directly, but we seek a lower bound on $\mathrm{E}_H\{\det\left(\mathcal{H}\mathcal{H}^{\mathrm{T}}\right)^{-1/2}\}$ when $Q \ge NT$, and a lower bound on $\mathrm{E}_H\{\det\left(\mathcal{H}^{\mathrm{T}}\mathcal{H}\right)^{-1/2}\}$ when $Q < NT$. To that end, we observe that the function $f(X) = \det(X)^{-1/2}$ is strictly convex over $\{X | X = X^{\mathrm{T}} \succ 0\}$.[4] Thus, we can employ Jensen's inequality (Lemma 1) to show that

$$\mathrm{E}\{\det(X)^{-\frac{1}{2}}\} \ge \det(\mathrm{E}\{X\})^{-\frac{1}{2}} \tag{30}$$

where X is defined as

$$X = \begin{cases}\mathcal{H}\mathcal{H}^{\mathrm{T}}, & Q \ge NT,\\ \mathcal{H}^{\mathrm{T}}\mathcal{H}, & Q \le NT.\end{cases}$$

As in the previous sections, this bound is optimized if $\mathrm{E}_H\{\mathcal{H}\mathcal{H}^{\mathrm{T}}\} \to MI_{2NT}$ for $Q \ge NT$ and $\mathrm{E}_H\{\mathcal{H}^{\mathrm{T}}\mathcal{H}\} \to (MNT/Q)I_{2Q}$ for $Q < NT$. In the next section, we will show that when the number of symbols per block Q is appropriately

[3]Although this average PEP measure [6] is significantly different from the conventional PEP [4], it does provide similar insight, as we now show.

[4]In fact, $f(X)$ is log convex over $\{X | X = X^{\mathrm{T}} \succ 0\}$, and hence, convex over this set.

chosen, coding matrices A_q with a unitary structure asymptotically achieve the minimized lower bound of the high-SNR average PEP in (29).

D. Asymptotic Optimality of Unitary Coding Matrices

In this section, we will show that unitary coding matrices A_q not only minimize the lower bound on the MSE in (10) and the PEP bound in (29), as well as maximize the upper bound on the mutual information that can be achieved by an LD code in (20). For any given N, they also asymptotically attain these bounds as T, M, and Q grow, so long as the scattering environment remains sufficiently rich to provide independence between the entries of the channel matrix.

The justification of the claim of asymptotic optimality of the unitary coding matrices, is based on the following lemma [19].

Lemma 2: Let C and $D \in \mathbb{R}^{n\times n}$ be diagonal matrices with $C = \mathrm{Diag}(c_1, \ldots, c_n)$ and $D = \mathrm{Diag}(d_1, \ldots, d_n)$, where $c_1 \geq c_2 \geq \ldots \geq c_n \geq 0$, $d_1 \geq d_2 \geq \ldots \geq d_n \geq 0$, and $\|C\| = \|D\| = n$. Let $\Delta = \Gamma + j\Lambda$ be a unitary matrix uniformly distributed according to the Haar measure.[5] Then, as n increases, $(\mathrm{Tr}(C\Gamma), \mathrm{Tr}(D\Lambda))$ converges in distribution to $(1/\sqrt{2})(Z_1, Z_2)$, where Z_1 and Z_2 are independent identically distributed (i.i.d.) standard normal random variables. That is, $\mathrm{Tr}(C\Gamma + jD\Lambda)$ converges to a complex standard circular normal distribution.

In order to enable the asymptotic analysis, we assume that the coding matrices A_q are randomly chosen from the uniform distribution (in the Haar measure) on the group of unitary matrices. Using the result in Lemma 2, we prove the following proposition in Appendix II.

Proposition 1: Let $\mathcal{H} \in \mathbb{R}^{2NT\times 2Q}$ [cf., (5)] be given by

$$\mathcal{H} = [I_N \otimes \mathcal{A}_1 \quad I_N \otimes \mathcal{A}_2 \quad \cdots \quad I_N \otimes \mathcal{A}_{2Q}]\,[I_{2Q} \otimes \underline{h}]$$

where the entries of $\underline{h} \in \mathbb{R}^{2MN}$ are i.i.d. zero-mean Gaussian random variables with a variance of one half, and $\forall q \in [1, 2Q]$, $\mathcal{A}_q$ is given by (6), where A_q are uniformly distributed with respect to the Haar measure on the group of unitary matrices. Then, for any $N \geq 1$, $Q \geq M$, and with the ratio $T/M \leq 2Q$ held constant

$$\mathcal{H}\mathcal{H}^{\mathrm{T}} \xrightarrow{\text{a.s.}} \mathrm{E}_H\{\mathcal{H}\mathcal{H}^{\mathrm{T}}\} \tag{31}$$

as $M \to \infty$, where $\xrightarrow{\text{a.s.}}$ denotes almost sure convergence.

The condition in (31) is sufficient for the bounds derived in Sections III-A–C to be approached, and hence, randomly selecting unitary coding matrices is asymptotically optimal as M and T grow.

It is to be noted here that although we have considered complex coding matrices A_q, the above asymptotic analysis is valid for real orthogonal matrices as well. An advantage of constraining the coding matrices to be real is that it reduces the number of parameters to be optimized in designing the code. However, this reduction in design complexity comes at the possible risk of a reduction in performance.

It is interesting to compare the result in Proposition 1 with a related result that appeared recently in [21]. There, it was shown that using orthogonal coding matrices is asymptotically information lossless as the SNR $\to 0$. It was conjectured therein that that result carries over to the high-SNR realm. Using Proposition 1, we can obtain a complementary result that states that for any SNR, unitary coding matrices are asymptotically information lossless as the number of transmit antennas grows.

IV. Design Procedure

In the previous section, it was argued that for large systems (with T/M held constant), picking unitary coding matrices A_q randomly is optimal in the sense that it enables $\mathcal{H}\mathcal{H}^{\mathrm{T}}$ to approach $\mathrm{E}_H\{\mathcal{H}\mathcal{H}^{\mathrm{T}}\}$ as the dimensions of $\mathcal{H}$ grow, i.e., $\mathcal{H}\mathcal{H}^{\mathrm{T}}$ concentrates around its mean. However, when the dimension of the coding matrices is not large enough to obtain statistical independence between the entries, picking the coding matrices at random may result in a nearly singular equivalent channel matrix $\mathcal{H}$ for many channel realizations, and hence, poor performance. In order to obtain "good" unitary coding matrices for such practical systems, we need a selection criterion. In this section, we exploit insight from the proof of Proposition 1 to develop a candidate design procedure. In Section VI, we will demonstrate that our design procedure can provide codes that perform well in practice.

The first step in the design procedure is to determine the number of symbols Q to be transmitted over the channel in each block of T channel uses. The fact that $\mathcal{H}$ has dimensions $2NT \times 2Q$ immediately suggests choosing $Q \leq NT$ [6]. This choice of Q is also supported by the MSE expression in (8). However, the proof of Proposition 1 suggests that in order to approach the optimized bounds in Section III, one should choose Q to be as large as possible in order to provide averaging over the largest possible set of coding matrices. Based on these observations, we do not restrict Q to be equal to $\min\{NT, MT\}$ as in [6]. Instead, we propose choosing $Q = NT$, which yields a square equivalent channel matrix $\mathcal{H}$.

We also require a selection criterion for designing the unitary coding matrices A_q. To develop an appropriate criterion, we will use the analysis of Section III. The PEP bound in (27) suggests that we need to ensure that the minimum eigenvalue of $\mathcal{H}\mathcal{H}^{\mathrm{T}}$ is bounded away from zero for "as many channel realizations as possible." In a complementary way, the MSE and mutual information analyses show that lower and upper bounds on these quantities, respectively, are achieved at their optimized values when $\mathcal{H}\mathcal{H}^{\mathrm{T}}$ approaches scaled identity for every channel realization. Therefore, we expect good performance from the coding scheme if the equivalent channel matrix $\mathcal{H}$ is "close" to being orthogonal for as many realizations as possible. We will use a statistical measure of the proximity of $\mathcal{H}$ to a scaled identity so that a standard stochastic optimization technique

[5]The Haar measure [20] is the unique rotationally invariant measure (up to scalars) associated with compact measurable sets of unitary matrices. Thus, if $\mathcal{S}$ is a measurable subset of the set of unitary $p \times q$ matrices, then the Haar measure associated with this subset, $\mu(\mathcal{S}) = \mu(P\mathcal{S})$ for any $p \times p$ unitary matrix P. For example, in $\mathbb{R}^2$, the Haar measure $\mu(\mathcal{S})$ associated with a measurable set of unit vectors $\mathcal{S}$ is given by the length of the arc of the unit circle spanned by the unit vectors contained in $\mathcal{S}$. A general $M \times N$ unitary matrix is said to be isotropically distributed if it is uniformly distributed with respect to the Haar measure [11].

can be employed. The measure is based on the observation that for any $n \times n$ positive definite matrix X

$$\arg \max_{\mathrm{Tr}(X) \leq cn} \det(X) = cI. \tag{32}$$

Hence, the expected determinant of $\mathcal{H}\mathcal{H}^{\mathrm{T}}$ is an appropriate measure of the proximity of $\mathcal{H}\mathcal{H}^{\mathrm{T}}$ to a scaled identity. Although it is possible to develop other measures that promote the same goal, this measure is consistent with the performance criteria discussed in Section III. Our proximity measure has the advantage that it is a smooth continuous function of the parameters that characterize the unitary coding matrices and, as we will see in Section VI, it results in an effective design algorithm for all configurations that satisfy[6] $T/M \leq 2Q$. Therefore, we will design finite codes that come close to achieving the asymptotically achievable optimized bounds by solving the following stochastic optimization problem:[7]

$$\max_{A_q} \quad \mathrm{E}_H\{\det(\mathcal{H}\mathcal{H}^{\mathrm{T}})\} \tag{33a}$$

$$\text{subject to} \quad A_q A_q^{\mathrm{H}} = \frac{M}{Q} I. \tag{33b}$$

The optimization problem (33) looks rather similar to a high-SNR instance of the capacity maximization problem presented in [6] with coding matrices restricted to the unitary group (cf., (33b), [6, eq. 28]). However, the fact that we do not have the offset provided by the identity matrix in the mutual information expression (19) means that our formulation (33) strongly penalizes bad equivalent channel events. That is, the expression in (19) will be quite insensitive to slight variations in the minimum eigenvalue of $\mathcal{H}\mathcal{H}^{\mathrm{T}}$ when the latter approaches zero, whereas our formulation explicitly incorporates these variations. In Section III-C, we have discussed the impact of the minimum eigenvalue on the PEP [cf., (27)].

In order to explore the relationship between our design approach and the conventional diversity gain in more detail, we recall that the diversity gain of a multiple antenna system is quantified through the rank criterion defined in [22]. For an LD code [cf., (1)], the transmitted codewords take the form $S = \sum_q \alpha_q A_{2q-1} + \beta_q A_{2q}$, and the diversity gain is given by

$$\min_{\alpha+j\beta, \alpha'+j\beta' \in \mathcal{C}} \left\{ \mathrm{Rank} \left[\sum_q (\alpha_q - \alpha'_q) A_{2q-1} + (\beta_q - \beta'_q) A_{2q} \right] \right\}.$$

Now, consider our design problem (33). By maximizing $\det(\mathcal{H}\mathcal{H}^{\mathrm{T}})$, we implicitly attempt to reduce the linear dependence of the different columns of $\mathcal{H}$. Columns of $\mathcal{H}$ are linearly independent if and only if $\gamma_1 \check{h}_j + \gamma_2 \check{h}_k \neq 0$ for all γ_1, $\gamma_2 \neq 0$, $\|\check{h}_j\|$, $\|\check{h}_k\| \neq 0$, where $\check{h}_r$ denotes the rth column of $\mathcal{H}$. That is, if and only if

$$(\gamma_1 (I_N \otimes \mathcal{A}_j) + \gamma_2 (I_N \otimes \mathcal{A}_k)) \, \underline{h} \neq 0 \qquad \forall \underline{h} \neq 0, 1 \leq j, k \leq 2Q.$$

However, if $B = \gamma_1 (I_N \otimes \mathcal{A}_j) + \gamma_2 (I_N \otimes \mathcal{A}_k)$ is rank deficient, then there is always an $\underline{h}$ that lies in the null space of B for which the determinant will be zero, and hence, the columns will be linearly dependent. Hence, by maximizing $E_{\mathrm{H}}\{\det(\mathcal{H}\mathcal{H}^{\mathrm{T}})\}$, we penalize the rank deficiency of B and indirectly attempt to maximize the diversity gain.

An advantage of our formulation is that the objective is independent of the constellation $\mathcal{C}$, and hence, we can apply standard (stochastic) optimization techniques based on analytic expressions for the gradient of the objective (see Section VI and Appendix IV). This is in contrast to an alternative performance-orientated design [9] that uses a constellation-dependent PEP criterion, and hence, as the constellation size increases considerable computational effort is required to merely compute the objective. Actually, our constraint in (33b) that the code matrices be unitary is also slightly different from the constraint that was imposed in [9]. In our notation, the constraint in [9] corresponds to

$$\mathrm{Tr}(A_q^{\mathrm{H}} A_p) = \frac{TM}{Q} \delta_{pq} \tag{34}$$

where δ_{pq} is the Kronecker delta. The set of matrices that satisfy (34) intersects the set of unitary matrices, but neither is a proper subset of the other. Coding matrices that satisfy (34) are capacity optimal when $Q \geq MT$, but they do not necessarily provide good performance. Moreover, if $N < M$, effective detection for systems with $Q \geq MT > NT$ may incur considerable computational cost and performance degradation [cf., (8)]. In contrast, by constraining A_q to be unitary, we almost surely obtain optimality from both mutual information and performance perspectives for any number of receive antennas as the size of the system grows.

V. Row Interleaving

At high SNR, the error rate performance of a space-time code is dominated by the outage probability—the probability that the channel is unable to support the required data rate. Outage is associated with the event of the channel matrix dropping rank [3]. In order to reduce the probability of outage, one can employ interleaving. The general philosophy of interleaving in communication over fading channels is to first impose structure on the data sequence to be transmitted, and then to interleave the structured sequence so that related components are transmitted over independent channel realizations. The standard implementation of this strategy in the space-time coding literature is to impose the structure on the data sequence using an outer (scalar) code, then interleave the coded stream and pass it to a standard space-time mapper (e.g., BLAST or OSTBC). The alternative approach that we propose in this section is to exploit the structure imposed by the LD code itself. We will show in Section VI that the structure of our codes enables them to significantly outperform the standard LD codes [6] when the interleaving scheme we describe below is employed.

A disadvantage of the standard LD coding framework that we have considered in this paper is that each row of the codeword S is transmitted over the same channel. While maximizing $\mathrm{E}_H\{\det(\mathcal{H}\mathcal{H}^{\mathrm{T}})\}$ penalizes the occurrence of an outage,

[6] Since we have chosen $Q = NT$, this condition is always satisfied.

[7] In relation to (32), the trace constraint in (32) is captured by the power constraint on the coding matrices [cf., (33b)] and the norm of the given channel realization.

it may be insufficient to protect the transmitted codeword if $\|\underline{h}\|$ drops below a certain level. However, by augmenting the LD-code framework by grouping successive codeword matrices together and interleaving their rows, we can benefit from the temporal diversity provided by different channel realizations. This scheme allows us to reduce the outage probability, and hence, improve the high-SNR performance. (Recall that we have assumed an independent block-fading model for the channel.) In particular, given the codewords $S_1, S_2, \ldots, S_J$, [cf., (1)] we construct the row-interleaved codewords $\tilde{S}_1, \tilde{S}_2, \ldots, \tilde{S}_J$, such that

$$\begin{bmatrix} \tilde{S}_1 \\ \vdots \\ \tilde{S}_J \end{bmatrix} = P \begin{bmatrix} S_1 \\ \vdots \\ S_J \end{bmatrix} \tag{35}$$

where J is called the interleaving depth and P is some $JT \times JT$ permutation matrix. Since the channel is assumed to change to an independent realization after a block of T channel uses, the fact that each codeword constitutes T rows implies that J should be chosen so that $J \leq T$. Otherwise, row interleaving will lead to further processing delay without extracting additional temporal diversity. In this paper, we will choose $J = T$. This choice ensures that each codeword can be symmetrically dispersed over the same number of channel realizations, i.e., there exists (at least) one permutation matrix P that shuffles the T^2 rows in such a way that rows are interleaved across (rather than within) blocks. In order to gain some insight into the proposed interleaving scheme, we will choose such a P, namely the one whose (i,j)th entry is given by $P(i,j) = 1$ for all $j = \lceil i/T \rceil + T((i-1) \bmod T)$, $i \in [1, T^2]$, and zero otherwise. The row-interleaved codewords $\tilde{S}_t$ are then transmitted over the respective channel realizations $H^{(t)}$, $t \in [1, T^2]$. This model can be considered as a special case of the general block-fading model of [23]. However, unlike [23], our interleaving scheme does not incur extra detection complexity due to augmentation of the equivalent channel matrix. This fact will become clear as we discuss our proposed scheme in more detail. While we do not claim optimality of the proposed interleaving scheme, we will discuss below the intuition that led to its development, and in Section VI we will demonstrate its effectiveness.

With the row-interleaving scheme in (35), the vectorized received symbol matrix can be written in a form similar to (4) as

$$y = \sqrt{\frac{\rho}{M}} \tilde{\mathcal{H}} s + v$$

where

$$\tilde{\mathcal{H}} = \begin{bmatrix} I_N \otimes \tilde{\mathcal{A}}_1 & I_N \otimes \tilde{\mathcal{A}}_2 & \cdots & I_N \otimes \tilde{\mathcal{A}}_{2Q} \end{bmatrix} \begin{bmatrix} I_{2Q} \otimes \underline{\tilde{h}} \end{bmatrix}.$$

If $\mathcal{A}_q(i,:)$ denotes the ith row of $\mathcal{A}_q$, and if we define the matrix E_{ij} to be the all zero $2T \times T$ matrix with unity in the (i,j)th position [24], then our particular choice of the permutation matrix P will result in

$$\tilde{\mathcal{A}}_q = \sum_{i=1}^{2T} E_{i,\kappa(i)} \otimes \mathcal{A}_q(i,:)$$

where

$$\kappa(i) = 1 + (i-1) \bmod (T) \tag{36}$$

and $\underline{\tilde{h}} = [h_1^{(1)^{\mathrm{T}}} \ldots h_1^{(T)^{\mathrm{T}}} \ldots h_N^{(1)^{\mathrm{T}}} \ldots h_N^{(T)^{\mathrm{T}}}]^{\mathrm{T}}$, where $h_n^{(t)}$ is the vector h_n in (5) at the tth realization of the channel matrix, $1 \leq t \leq T$. Notice that the size of $\tilde{\mathcal{H}}$ is the same as the size of $\mathcal{H}$. This fact guarantees that the complexity of the detector used at the receiver remains the same as for the non-interleaved codes. That is, while the proposed interleaving scheme incurs the standard latency penalty of transmission schemes that code over independent realizations, the complexity of the detection problem is unchanged.

As we show in Appendix III, using unitary coding matrices along with row interleaving enhances diagonal dominance of $\tilde{\mathcal{H}}\tilde{\mathcal{H}}^{\mathrm{T}}$ over that of $\mathcal{H}\mathcal{H}^{\mathrm{T}}$, and hence makes $\tilde{\mathcal{H}}$ "closer" to being orthogonal than $\mathcal{H}$, even for systems of finite size. The arguments that led to the design problem in (33) suggest that row interleaving will improve the high SNR performance of our LD codes. As we will show in the next section, these performance improvements can be substantial.

VI. Numerical Results

In this section, we provide some numerical examples that illustrate the potential of the proposed designs. We have chosen scenarios from Hassibi and Hochwald's work [6] in which their codes (which will be referred to as HHLD codes) were shown to outperform OSTBCs and BLAST. Given that the LD codes framework subsumes other recent designs, we have also chosen scenarios from [7]–[9]. Before we begin, a few remarks regarding the practical implementation of our design approach are provided below.

1) The optimization problem in (33) can be solved by combining the principles of stochastic optimization [25] to deal with the expectation in the objective, with the principles of optimization over the Stiefel manifold [26]–[28] to deal with the orthogonality constraints. In the examples below, we have chosen to constrain the coding matrices to the manifold in (33b) by parameterizing A_q via Givens rotations [29]. We then applied stochastic quasi-gradient methods [25] to solve (33). Note that by parameterizing A_q via Givens rotations, the optimization problem (33) becomes unconstrained, and hence more straightforward to solve. Furthermore, the gradient (and the Hessian matrix) of the stochastic optimization objective with respect to the Givens rotations can be found analytically. For convenience, in Appendix IV we have provided the expression for the gradient along with a brief description of the Givens parameterization.
2) For the case where $T = M$, the restriction of our search to the group of unitary matrices reduces the number of parameters to be optimized from $4QM^2$ real parameters to $2QM^2$ real parameters. This is a significant reduction in complexity. Further reduction can be achieved by restricting our attention to real orthonormal matrices. In that case, the number of parameters is $QM(M-1)$.

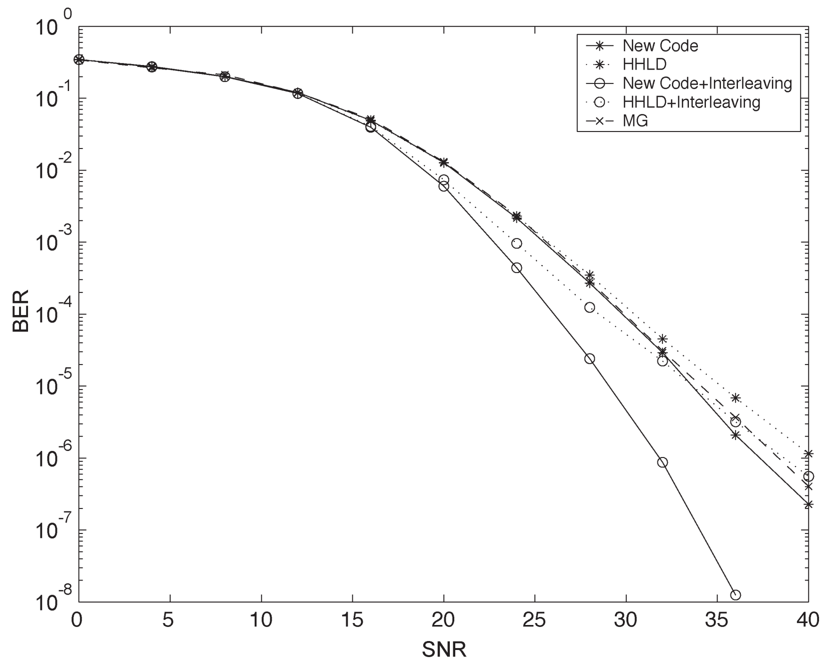

Fig. 1. Comparison between an LD code designed using (33), and the corresponding HHLD [6] and MG [8] codes when $M = N = T = 2$, $Q = NT = 4$, and 16-QAM symbols are transmitted.

3) When $T \neq M$, we embed the matrices A_q in square unitary matrices of dimension $\max(T, M)$ and parameterize the square matrices. The resulting coding matrices will be a subset of the columns or rows of the square matrices. In the case when we choose $M > T$, this parameterization guarantees that condition (18) holds, and in the case where $M < T \leq 2MQ$, we can structure the embedding so that the unitarity of the square matrix implies that (16) holds.
4) In each of the considered examples, row interleaving was implemented using the permutation matrix P given in Section V.
5) The coding matrices used in the following examples were obtained using the procedure outlined in Section IV.[8]

Example 1: We begin with a comparison with the HHLD codes [6] and the full-rate full-diversity layered code of Ma and Giannakis [8] (denoted MG) in the simple case in which $M = N = T = 2$, $Q = NT = 4$, and the symbols are drawn from a Gray-coded 16-quadrature amplitude modulation (QAM) constellation. The resulting transmission rate is $R = 8$ bits/channel use. Fig. 1 shows the bit-error-rate (BER) performance for the HHLD code [6], the MG [8] code, and a code we designed using (33). When no row interleaving is used, our code provides a slight improvement in performance over the HHLD and the MG codes. When row interleaving is employed, our code provides substantially better performance than the HHLD code. For example, the SNR gain of our code over the HHLD code is about 8 dB at a BER of around 10^{-6}. ■

Example 2: In this example, we choose $M = 3$, $N = 1$, $T = 3$, and $T = 6$. When $T = 3$, we provide comparisons with the corresponding TAST code [7], and when $T = 6$, we provide comparisons with the corresponding HHLD code [6]. In all cases, the number of symbols per block Q was chosen to be $Q = NT$. Therefore, the data rate is maintained the same for all cases.

The mutual information achieved by each code is illustrated in Fig. 2, along with the ergodic channel capacity and the bound derived in Section III-B [cf., (24)]. We observe that for $T = 3$, our code provides larger mutual information than the TAST code (the SNR gap to the ergodic capacity is around 1.2 dB for our code, and around 1.6 dB for the TAST code). However, unlike TAST codes, LD codes naturally allow the exploitation of channel coherence intervals that are longer than the number of transmit antennas. In particular, when $T = 6$, the mutual information achieved by our code and the HHLD code is substantially closer to the channel capacity (the SNR gap has been reduced to 0.4 dB). This reduction in the SNR gap is in agreement with the result in Proposition 1, where it was shown that increasing the size of the (unitary) coding matrices reduces the statistical dependence between the entries. Moreover, since the number of symbols Q is chosen to be equal to NT, by increasing T, we allow more averaging over the coding matrices [cf., (23)], which also contributes to the reduction of the SNR gap between the ergodic channel capacity and the mutual information achieved by the code.

To examine the performance of these systems, we consider the case in which the symbols are drawn from the QPSK constellation, and hence, the rate is $R = 2$ bits/channel use. For the case when $T = 3$, our design provides a considerable BER performance gain over the TAST code, as can be seen from Fig. 3. For the case of $T = 6$, our design provides appreciable BER performance gain over the HHLD design at high SNR in the absence of row interleaving, and substantial performance gain when row interleaving is used (an SNR gain of around 5 dB is achieved at a BER of 10^{-6}). There is a detectable

[8] Available at http://www.ece.mcmaster.ca/~davidson.

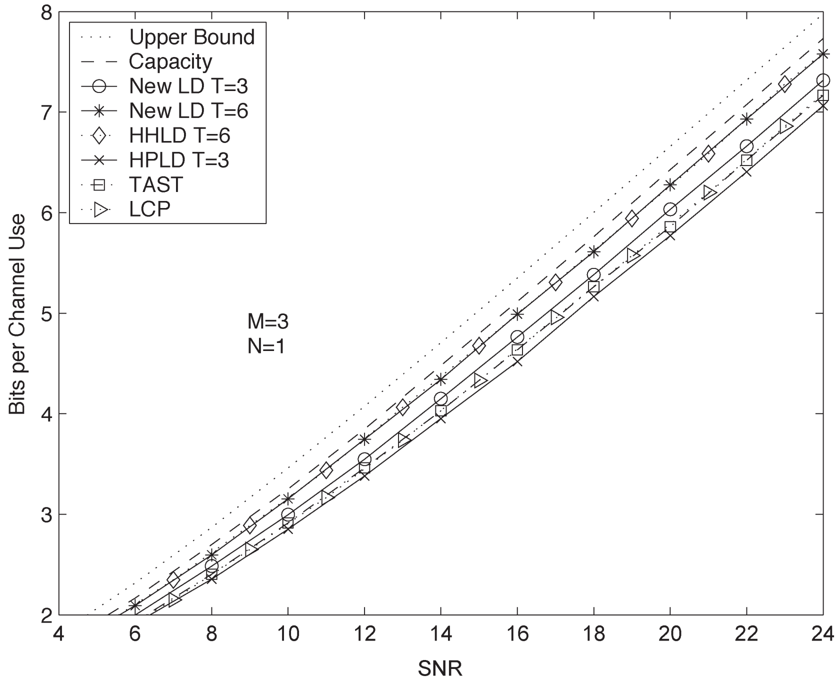

Fig. 2. Actual channel capacity, the upper bound in (24) and the information rate achieved by our LD code, the HHLD code [6], and the TAST code [7] in Example 2, the HPLD code [9] in Example 3, and the LCP code [30] in Example 4.

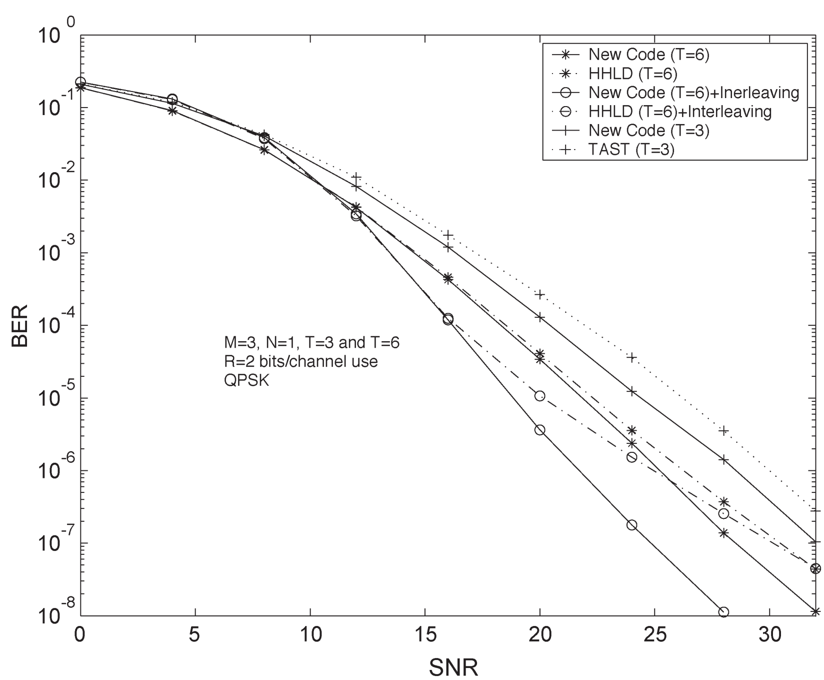

Fig. 3. Comparison between an LD code designed using (33) and codes presented in [6] (HHLD) and [7] (TAST) when $M = 3$, $N = 1$, $T = 3$ and $T = 6$, $Q = NT = 6$, and QPSK symbols are transmitted.

deterioration in the performance of the interleaved scheme at low SNRs. This is due to the fact that the proposed interleaving scheme does not control the conditioning of $\tilde{\mathcal{H}}\tilde{\mathcal{H}}^{\mathrm{T}}$, and the weaker dimensions are more vulnerable to noise. However, the interleaved codes perform better than the noninterleaved codes at typical operating SNRs. ■

Example 3: In this example, we compare our codes with the $M = T = 3$, $N = 1$, and $Q = NT = 3$ code presented in [9], which we will denote by HPLD. The mutual information achieved by this code is shown in Fig. 2. At a data rate of 3 bits/channel use, the SNR gap to the ergodic capacity of this code is about 1.9 dB, which is larger than the SNR gaps of both the proposed LD code and the TAST code. In order to examine the performance of our codes in comparison with that of the HPLD code, we consider the case in which the transmitted symbols are drawn from a 16-QAM constellation, yielding a transmission rate of $R = 4$ bits/channel use. We have designed two LD codes in this case; one each for the schemes with and

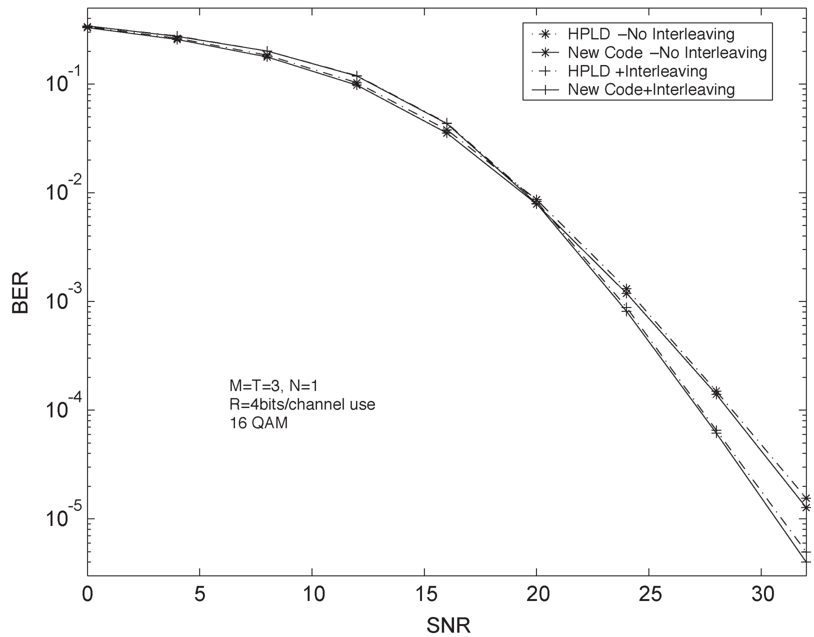

Fig. 4. Comparison between the new LD codes and the HPLD code [9] when $M = T = 3$, $N = 1$, $Q = NT = 3$, and 16-QAM symbols are transmitted.

without row interleaving. For the case when row interleaving is employed, we restricted our attention, as in [9], to coding matrices that satisfy $A_{2q-1} = A_{2q}$ for $q = 1, \ldots, Q$. Notice that this choice of A_q results in the $(2q-1)$th and the $(2q)$th columns of $\mathcal{H}$ to be orthogonal to each other. We observe from Fig. 4 that, at high SNR, row interleaving provides a significant improvement in the performance for both codes. However, our codes perform slightly better than those designed in [9], and are much easier to design. ■

Example 4: As a final example, we compare our designs with the full-rate full-diversity layered code of Ma and Giannakis [8] (denoted MG) and the linear constellation precoding scheme developed in [30] (denoted LCP). We will also provide comparisons with the HPLD codes [9] and TAST [7] codes. We choose $T = M = 3$ and $N = 1$. For our codes, the number of transmitted symbols Q is chosen to be $Q = NT = 3$, which is the same as the value of Q chosen for the corresponding LCP, HPLD, and TAST codes. The MG codes are a full symbol rate design, and hence, $Q_{\mathrm{MG}} = MT = 9$. One of the advantages of the MG codes in this scenario is that they are "information lossless." That is, the mutual information achieved by the MG codes coincides with the dashed line representing the ergodic capacity in Fig. 2. In addition, the MG, LCP, and TAST codes generically achieve full diversity, and the HPLD code for this scenario also achieves full diversity. Therefore, there is considerable interest in comparing the block-error-rate performance of these codes with that of our code (recall that diversity is defined in terms of the rate of decay of the block error rate with $\log(\mathrm{SNR})$ at high SNR). Since $Q_{\mathrm{MG}} = 9$ and $T = 3$, the smallest uncoded data rate that we can consider is 3 bits/channel use, which corresponds to using BPSK symbols in the MG code. Since $Q = 3$ for the other codes, they must employ an octal constellation to achieve the same data rate, and we will choose Gray-coded 8-PSK. The block error rate curves for these systems are provided in Fig. 5. At high SNRs, the slopes of these curves are (essentially) the same, and hence all five schemes (essentially) achieve full diversity. However, our code has a significant performance advantage over the HPLD, LCP, and TAST codes (approximately 1, 1.2, and 2 dB, respectively, at a block error rate of 10^{-5}), and a substantial advantage over the MG codes (approximately 6 dB at a block error rate of 10^{-5}). The performance advantage of our codes is due, in part, to the fact that our design (implicitly) considers the overall average PEP (see Sections III and IV), rather than just the diversity component. The slow rate of decay of the block error rate of the MG code at moderate SNRs is worthy of some discussion. We suspect that this is due to the fact that in this example the MG code results in a "fat" equivalent channel matrix (of dimension[9] 6×9), whereas the equivalent channel matrices of the other codes are square (6×6). In addition to this performance issue, the "fat" equivalent channel matrix of the MG code requires the use of computationally expensive detection techniques [31]. These difficulties are alleviated by the flexibility of the LD framework, which allows the designer to manipulate the signaling strategy so that the equivalent channel matrix is square (or tall).

VII. Conclusion

In this paper, we demonstrated that coding matrices drawn at random from the group of unitary matrices are asymptotically optimal from several design perspectives, including MSE, mutual information, and average PEP. We then provided a systematic, versatile, and efficient constellation-independent design technique for finding "good" unitary coding matrices for practical systems. In our examples, these codes performed better than those currently available. The properties of a

[9]The equivalent channel matrix $\mathcal{H}$ is nominally $2NT \times 2Q = 6 \times 18$, but since BPSK signaling is employed, only Q columns of $\mathcal{H}$ are active.

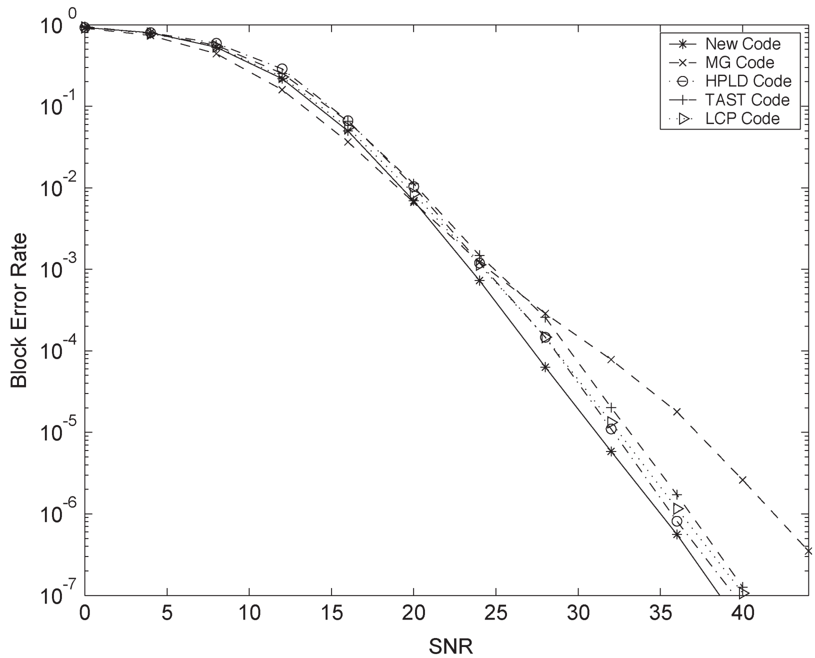

Fig. 5. Comparison between the new LD code and the codes presented in [8] (MG), [30] (LCP), [9] (HPLD), and [7] (TAST) when $M = T = 3$, $N = 1$. For the MG code, $Q = 9$ and the underlying constellation is BPSK, and for the other codes, $Q = 3$ and the underlying constellation is 8-PSK.

unitary coded system prompted us to propose a row interleaving scheme for LD codes that was shown to significantly improve the system performance at high SNR without increasing the complexity of the detector.

Appendix I
A Bound on the Ergodic Channel Capacity

Let $C(M,N,\rho)$ denote the ergodic channel capacity of a system with M transmit and N receive antennas. It is well known [1] that $C(M,N,\rho) = \mathrm{E}_H\{\log\det(I_N + (\rho/M)H^{\mathrm{H}}H)\}$. In order to show that the right hand side of (24) is an upper bound on $C(M,N,\rho)$, we point out that since $C(M,N,\rho)$ is a monotonically increasing function of M, then

$$C(M,N,\rho) \le \lim_{\tilde{M}\to\infty} C(\tilde{M},N,\rho).$$

Furthermore,

$$\lim_{\tilde{M}\to\infty} C(\tilde{M},N,\rho)$$
$$= \mathrm{E}_H\left\{\lim_{\tilde{M}\to\infty} \log\det\left(I_N + \frac{\rho}{\tilde{M}}\begin{bmatrix} H_1^{\mathrm{H}}H_1 & \cdots & H_1^{\mathrm{H}}H_N \\ \vdots & \ddots & \vdots \\ H_N^{\mathrm{H}}H_1 & \cdots & H_N^{\mathrm{H}}H_N\end{bmatrix}\right)\right\} \tag{37}$$
$$= \mathrm{E}_H\left\{\log\det(I_N + \rho I_N)\right\} \tag{38}$$
$$= N\log(1+\rho) \tag{39}$$

where, in (37), we have denoted the jth column of H as H_j, and in (38), we have used the fact that for a richly scattered environment, the entries of the channel matrix H are i.i.d. complex Gaussian random variables $\mathcal{CN}(0,1)$, and hence, for any two columns in H, H_i and H_j, we have $\lim_{\tilde{M}\to\infty}(1/\tilde{M})H_i^{\mathrm{H}}H_j = \delta_{ij}$, where δ_{ij} is the Kronecker delta function.

Appendix II
Proof of Proposition 1

Before proceeding with the proof of Proposition 1, we first point out that in Lemma 2, the assumption of diagonal C and D matrices having nondecreasing entries is without loss of generality. To show that, let $E = UCV^{\mathrm{T}}$ be the singular value decomposition of an arbitrary real matrix E. Here, U and V are orthonormal and $C = \mathrm{Diag}(c_1,\ldots,c_n)$ and $(c_1 \ge c_2 \ge \ldots \ge c_n \ge 0)$. Now

$$\begin{aligned}\mathrm{Tr}(E\Gamma) &= \mathrm{Tr}\left(U^{\mathrm{T}}(EVV^{\mathrm{T}}\Gamma)U\right)\\ &= \mathrm{Tr}\left((U^{\mathrm{T}}EV)(V^{\mathrm{T}}\Gamma U)\right) = \mathrm{Tr}(C\Xi)\end{aligned}$$

where $\Xi = V^{\mathrm{T}}\Gamma U$. Notice that by appropriately choosing E, one can choose any entry from Γ, implying that each entry of Γ converges in distribution to a standard normal random variable. Hence, if Γ converges in distribution to the isotropic standard normal distribution, then so does Ξ (the same argument holds for D and Λ).

As mentioned in Proposition 1, we assume that for some $N \ge 1$, $Q \ge M$, with T/M held constant, and the coding matrices are drawn at random from the group of unitary matrices. Notice that for the actual finite-sized coding matrices, in order to ensure the existence of coding matrices A_q that satisfy (16) and (23), the block size T must be restricted to be no greater than $2MQ$. However, the result in Proposition 1 applies for any

T that grows in proportion to M. For the sake of brevity, we will only consider square coding matrices A_q, i.e., $T/M = 1$.

We begin by observing that the (i, j)th $2T \times 2T$ block of $(1/M)\mathcal{H}\mathcal{H}^{\mathrm{T}}$ can be written as

$$Z_{ij} = \frac{1}{M}\sum_q \mathcal{A}_q h_i h_j^{\mathrm{T}} \mathcal{A}_q^{\mathrm{T}}. \tag{40}$$

Now, the (k, ℓ)th element of this block can be written as

$$\begin{aligned}[Z_{ij}]_{k\ell} &= \frac{1}{M}\sum_q \mathcal{A}_q(k,:) h_i h_j^{\mathrm{T}} \mathcal{A}_q(\ell,:)^{\mathrm{T}} \\ &= \frac{1}{M}\left(h_j^{\mathrm{T}}\left[\sum_q \mathcal{A}_q(\ell,:)^{\mathrm{T}}\mathcal{A}_q(k,:)\right] h_i\right)\end{aligned} \tag{41}$$

where $\mathcal{A}_q(\ell,:)$ denotes the ℓth row of $\mathcal{A}_q$. Consider the following scenario, in which we pick $2Q$ i.i.d. random unitary coding matrices A_q from a uniform distribution with respect to the Haar measure of unitary matrices, and construct the corresponding $2M \times 2M$ matrices[10] $\mathcal{A}_q$ with the symplectic structure given in (6), such that $\mathcal{A}_q\mathcal{A}_q^{\mathrm{T}} = (M/Q)I_{2M}$.

We note that, while the matrices A_q are uniformly distributed with respect to the Haar measure of unitary matrices, the matrices $\mathcal{A}_q$, due to their symplectic structure, are not uniformly distributed with respect to the Haar measure of orthogonal matrices. However, our treatment needs only the circular symmetry of A_q. By invoking the result in Lemma 2, we can examine the asymptotic behavior of $[Z_{ij}]_{k\ell}$ in (41) as M grows. In particular, Lemma 2 enables us to assume independence between the entries of unitary coding matrices of sufficiently large dimensions. Using this result, we first consider the case when $|k-\ell| \neq T$ in (41). For that case, one can write

$$\begin{aligned}&\frac{1}{2Q}\sum_{q=1}^{2Q}\mathcal{A}_q(\ell,:)^{\mathrm{T}}\mathcal{A}_q(k,:) \\ &= \frac{1}{2Q}\sum_{q=1}^{2Q}\begin{bmatrix}\mathcal{A}_q(\ell,1)\mathcal{A}_q(k,1) & \cdots & \mathcal{A}_q(\ell,1)\mathcal{A}_q(k,2M) \\ \vdots & \ddots & \vdots \\ \mathcal{A}_q(\ell,2T)\mathcal{A}_q(k,1) & \cdots & \mathcal{A}_q(\ell,2M)\mathcal{A}_q(k,2M)\end{bmatrix}.\end{aligned} \tag{42}$$

Notice that for any $k \neq \ell$ or $n \neq m$, $\{\zeta_q | \zeta_q = \mathcal{A}_q(\ell,m)\mathcal{A}_q(k,n)\}_{q=1,\ldots,2Q}$ is a set of i.i.d. zero-mean random variables. Thus, for $k \neq \ell$, by the strong law of large numbers, $(1/2Q)\sum_{q=1}^{2Q}\mathcal{A}_q(\ell,:)^{\mathrm{T}}\mathcal{A}_q(k,:)$ converges to $0_{2M\times 2M}$ almost surely.

For the case when $k = \ell$, we have for $Q \geq M$

$$\frac{1}{2Q}\sum_{q=1}^{2Q}\mathcal{A}_q(\ell,:)^{\mathrm{T}}\mathcal{A}_q(\ell,:) \xrightarrow{\text{a.s.}} \rho_0 I_{2M} \tag{43}$$

where $\rho_0 = \mathrm{E}_H\{|\mathcal{A}_q(\ell,k)|^2\} = 1/2Q$. Next, we consider the case when $|k-\ell| = T$. In this case, we notice that because of the structure in (6), we can no longer assume independence between the entries. In fact, for this case [cf., (6)]

$$\begin{aligned}&\frac{1}{2Q}\sum_{q=1}^{2Q}\mathcal{A}_q(\ell,:)^{\mathrm{T}}\mathcal{A}_q(\ell,:) \\ &= \frac{1}{2Q}\sum_q\begin{bmatrix}A_{R,q}(\ell,:)^{\mathrm{T}}A_{I,q}(\ell,:) & A_{R,q}(\ell,:)^{\mathrm{T}}A_{R,q}(\ell,:) \\ -A_{I,q}(\ell,:)^{\mathrm{T}}A_{I,q}(\ell,:) & A_{I,q}(\ell,:)^{\mathrm{T}}A_{R,q}(\ell,:)\end{bmatrix}.\end{aligned} \tag{44}$$

Now, using the independence assumption on the real and imaginary components of the entries of A_q as per Lemma 2, one can argue that for $Q \geq M$

$$\frac{1}{2Q}\sum_{q=1}^{2Q}\mathcal{A}_q(\ell,:)^{\mathrm{T}}\mathcal{A}_q(\ell,:) \xrightarrow{\text{a.s.}} \frac{1}{2Q}\begin{bmatrix}0 & I_M \\ -I_M & 0\end{bmatrix}. \tag{45}$$

Substituting (43) and (45) in (41), one obtains

$$[Z_{ij}]_{k\ell} \xrightarrow{\text{a.s.}} \frac{1}{M}\left(h_j^{\mathrm{T}} h_i \delta_{k\ell} + h_j^{\mathrm{T}}\begin{bmatrix}0 & I_M \\ -I_M & 0\end{bmatrix} h_i \delta_{|k-\ell|,T}\right) \tag{46}$$

where δ_{ij} denotes the Kronecker delta. Given our assumption on the distribution of the elements of H, it is observed that for $i \neq j$, the entries of h_i and h_j are independent. Therefore, by the strong law of large numbers, $(1/M)h_j^{\mathrm{T}}h_i \xrightarrow{\text{a.s.}} 0$, and hence, $[Z_{ij}]_{k\ell} \xrightarrow{\text{a.s.}} 0$. For $i = j$, the second term on the right-hand side of (46) vanishes, and the strong law of large numbers implies that $(1/M)h_i^{\mathrm{T}}h_i \xrightarrow{\text{a.s.}} 1$. Therefore

$$[Z_{ij}]_{k\ell} \xrightarrow{\text{a.s.}} \delta_{ij}\delta_{k\ell}.$$

Thus, our claim that as M increases $\mathcal{H}\mathcal{H}^{\mathrm{T}}$ converges almost surely to $\mathrm{E}_H\{\mathcal{H}\mathcal{H}^{\mathrm{T}}\} = MI_{2NT}$ is asserted.

Appendix III
Insight Into Row Interleaving

The row-interleaving operation derived in Section V has the tendency to increase the diagonal dominance of $\tilde{\mathcal{H}}\tilde{\mathcal{H}}^{\mathrm{T}}$. In order to show this, we write the $2T \times 2T$ blocks of $\tilde{\mathcal{H}}\tilde{\mathcal{H}}^{\mathrm{T}}$ in an analogous way to the $2T \times 2T$ blocks of $\mathcal{H}\mathcal{H}^{\mathrm{T}}$ in (40)

$$\tilde{Z}_{ij} = \frac{1}{M}\sum_q \tilde{\mathcal{A}}_q\begin{bmatrix}h_i^{(1)} \\ \vdots \\ h_i^{(T)}\end{bmatrix}\begin{bmatrix}h_j^{(1)} \\ \vdots \\ h_j^{(T)}\end{bmatrix}^{\mathrm{T}}\tilde{\mathcal{A}}_q^{\mathrm{T}}, \qquad i,j = 1,\ldots,N. \tag{47}$$

The (k, ℓ)th element of this block can be written as

$$\begin{aligned}[\tilde{Z}_{ij}]_{k\ell} &= \frac{1}{M}\sum_{q=1}^{2Q}\mathcal{A}_q(k,:)h_i^{(\kappa(k))}h_j^{(\kappa(\ell))}\mathcal{A}_q(\ell,:)^{\mathrm{T}} \\ &= \frac{1}{M}h_i^{(\kappa(\ell))^{\mathrm{T}}}\left[\sum_{q=1}^{2Q}\mathcal{A}_q(k,:)^{\mathrm{T}}\mathcal{A}_q(\ell,:)\right]h_j^{(\kappa(k))}\end{aligned} \tag{48}$$

[10]While we have only considered square coding matrices, the generalization to rectangular matrices is straightforward.

where $\kappa(i)$ was defined in (36). We observe that the entries of the off-diagonal blocks in (48) involve inner products of vectors from different channel realizations. Hence, they tend to approach zero as M grows, even when entries of the coding matrices are not completely statistically independent. In contrast, the diagonal blocks will be averaged out solely by the statistical independence of coding matrices. In fact, each row of $\tilde{\mathcal{H}}\tilde{\mathcal{H}}^T$ has exactly two entries that involve an inner product of the same channel realization and averaging over coding matrices (one of these is a diagonal entry). All other entries involve inner products of different realizations and averaging over coding matrices.

APPENDIX IV
GRADIENT COMPUTATION

One way to parameterize an $M \times M$ unitary matrix A is through Givens rotations [28], [29], [32]. Thereby

$$A = \left(\prod_{m=1}^{\frac{M(M-1)}{2}} G_m \right) D \left(\prod_{m=\frac{M(M-1)}{2+1}}^{M(M-1)} G_m \right) \tag{49}$$

where G_m is a planar rotation matrix for coordinates i_m and k_m with angle θ_m. This matrix can be constructed from the identity matrix by replacing the (i_m, i_m)th element by a, the (i_m, k_m)th element by b, the (k_m, i_m)th element by c, and the (k_m, k_m)th element by d, where for type ℓ rotations, $\ell \in \{1, 2\}$, $a = \cos(\theta_m)$, $b = \sin(\theta_m)$, $c = (-1)^\ell \sin(\theta_m)$, and $d = (-1)^{(\ell-1)} \cos(\theta_m)$. The matrix D is a diagonal unitary matrix, i.e, $[D]_{ii} = e^{j\phi_i}$.

If the coding matrices A_q are parameterized using (49), we can express the equivalent-channel matrix [cf., (5) and (6)] as a function of the set of Givens rotations Θ that parameterize the matrices A_q and the physical channel H as

$$\mathcal{H}(\Theta, H) = [\, \breve{h}_1 \quad \breve{h}_2 \quad \cdots \quad \breve{h}_{2Q} \,]$$

where $\breve{h}_q = (I_N \otimes \mathcal{A}_q)[h_1^{\mathrm{T}} \dots h_2^{\mathrm{T}}]^{\mathrm{T}}$, and $\mathcal{A}_q$ is defined in (6). Let C denote the matrix of cofactors of $\mathcal{H}$. Then

$$\det(\mathcal{H}) = \sum_i [\breve{h}_j]_i C_{ij} = \breve{h}_j^{\mathrm{T}} \breve{C}_j$$

where $\breve{C}_j$ is the jth column of C. This expression is convenient because a given planar rotation $\theta_r \in \Theta$ affects only one coding matrix A_q. Let $q(r)$ denote the index of that matrix. Then, θ_r affects only $\mathcal{A}_{q(r)}$ and $\breve{h}_{q(r)}$. That is, θ_r affects only one column of $\mathcal{H}$. Therefore

$$\frac{\partial \det(\mathcal{H})}{\partial \theta_r} = \frac{\partial \breve{h}_{q(r)}^{\mathrm{T}}}{\partial \theta_r} \breve{C}_{q(r)}. \tag{50}$$

Since $\breve{h}_{q(r)} = (I_N \otimes \mathcal{A}_{q(r)})[h_1^{\mathrm{T}} \dots h_N^{\mathrm{T}}]^{\mathrm{T}}$, we have that

$$\frac{\partial \breve{h}_{q(r)}}{\partial \theta_r} = \left(I_N \otimes \frac{\partial \mathcal{A}_{q(r)}}{\partial \theta_r} \right) [h_1^{\mathrm{T}} \quad \dots \quad h_N^{\mathrm{T}}]^{\mathrm{T}}. \tag{51}$$

Since $Q = NT$, $\mathcal{H}$ is square, and hence, $\det(\mathcal{H}\mathcal{H}^{\mathrm{T}}) = \det(\mathcal{H})^2$. Therefore, using (50), (51), and (6), the chosen Givens parameterization of A_q, and the standard rules for the derivative of a composite function, the gradient is readily computed.

REFERENCES

[1] I. E. Telatar, "Capacity of multiantenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.
[2] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multielement antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
[3] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
[4] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.
[5] G. Foschini, G. Golden, R. Valenzuela, and P. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 11, pp. 1841–1852, Nov. 1999.
[6] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1804–1824, Jul. 2002.
[7] H. El Gamal, M. O. Damen, and N. C. Beaulieu, "Linear threaded algebraic space-time constellations," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2372–2388, Oct. 2003.
[8] X. Ma and G. B. Giannakis, "Full-diversity full-rate complex-field space-time coding," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2917–2930, Nov. 2003.
[9] R. W. Heath and A. J. Paulraj, "Linear dispersion codes for MIMO systems based on frame theory," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2429–2441, Oct. 2002.
[10] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1639–1642, Jul. 1999.
[11] L. Zheng and D. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.
[12] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search of the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2389–2402, Oct. 2003.
[13] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2201–2214, Aug. 2002.
[14] G. Ginis and J. Cioffi, "On the relation between V-BLAST and the GDFE," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 364–366, Sep. 2001.
[15] J. C. Belfiore and G. Rekaya, "Quaternionic lattices for space-time coding," in *Proc. IEEE Information Theory Workshop*, Paris, France, Apr. 2003, pp. 267–270.
[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[17] F. A. Graybill, *Matrices With Applications in Statistics*, 2nd ed. Belmont, CA: Wadsworth, 1983.
[18] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[19] A. D. Aristotile, P. Diaconis, and C. M. Newman, "Brownian motion and the classical groups," Dept. Statistics, Stanford Univ., Stanford, CA, Tech. Rep. 2002-18, Aug. 2002.
[20] L. Nachbin, *The Haar Integral*. Princeton, NJ: Van Nostrand, 1965.
[21] R. Kashaev and O. Tirkkonen, "Linear matrix modulators from group representation theory," in *Proc. IEEE Information Theory Workshop*, Paris, France, Apr. 2003, pp. 42–45.
[22] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
[23] H. El Gamal and A. Hammons, "On the design of algebraic space-time codes for MIMO block-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 151–163, Jan. 2003.
[24] A. Graham, *Kronecker Products and Matrix Calculus, with Applications*. New York: Elis Horwood Ltd., 1981.

[25] A. Gaivoronski, “Stochastic quasigradient methods and their implementation,” in *Numerical Techniques for Stochastic Optimization*. New York: Springer-Verlag, 1988, ch. 16, pp. 313–351.
[26] A. Edelman, T. Arias, and S. Smith, “The geometry of algorithms with orthogonality constraints,” *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.
[27] J. H. Manton, “Optimization algorithms exploiting unitary constraints,” *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 635–650, Mar. 2002.
[28] B. Hassibi and B. M. Hochwald, “Cayley differential unitary space-time codes,” *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1485–1503, Jun. 2002.
[29] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.
[30] Y. Xin, Z. Wang, and G. B. Giannakis, “Space-time diversity systems based on linear constellation precoding,” *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 294–309, Mar. 2003.
[31] M. O. Damen, K. Abed-Meraim, and J. C. Belfiore, “Generalised sphere decoder for asymmetrical space-time communication architecture,” *Electron. Lett.*, vol. 36, no. 2, pp. 166–167, Jan. 2000.
[32] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. New York: Cambridge Univ. Press, 1994.

Ramy H. Gohary received the B.Sc. degree in electronics and communications engineering from Assiut University, Assiut, Egypt, in May 1996. In January 2000, he received the M.Sc. degree in communications engineering from Cairo University, Cairo, Egypt. He is currently working towards the Ph.D. degree in electrical engineering at McMaster University, ON, Canada.

He is interested in the analysis and design of multiple-input multiple-output (MIMO) wireless communication systems, applications of optimization techniques in signal processing, and implementation of iterative detection schemes in multiple-antenna multiuser communication systems.

Timothy N. Davidson (M’96) received the B.Eng. (Hons. I) degree in electronic engineering from the University of Western Australia (UWA), Perth, Australia, in 1991, and the D.Phil. degree in engineering science from the University of Oxford, U.K., in 1995.

He is currently an Associate Professor in the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he holds the (Tier II) Canada Research Chair in Communication Systems. His research interests are in signal processing, communications, and control, with current activity focused on signal processing for digital communication systems. He has held research positions at the Communications Research Laboratory, McMaster University, the Adaptive Signal Processing Laboratory, UWA, and the Australian Telecommunications Research Institute, Curtin University of Technology, Perth, Western Australia.

Dr. Davidson was awarded the 1991 J. A. Wood Memorial Prize [for “the most outstanding (UWA) graduand” in the pure and applied sciences], and the 1991 Rhodes Scholarship for Western Australia.