High-Rate Codes that are Linear in Space and Time

BABAK HASSIBI* BERTRAND M. HOCHWALD**

*Department of Electrical Engineering
California Institute of Technology, Pasadena, CA 91125
hassibi@caltech.edu

**Bell Laboratories, Lucent Technologies
600 Mountain Avenue, Murray Hill, NJ 07974
hochwald@bell-labs.com

April 24, 2001

Abstract

Multiple-antenna systems that operate at high rates require simple yet effective space-time transmission schemes to handle the large traffic volume in real time. At rates of tens of bits/sec/Hz, V-BLAST, where every antenna transmits its own independent substream of data, has been shown to have good performance and simple encoding and decoding. Yet V-BLAST suffers from its inability to work with fewer receive antennas than transmit antennas—this deficiency is especially important for modern cellular systems where a basestation typically has more antennas than the mobile handsets. Furthermore, because V-BLAST transmits independent data streams on its antennas there is no built-in spatial coding to guard against deep fades from any given transmit antenna. On the other hand, there are many previously-proposed space-time codes that have good fading resistance and simple decoding, but these codes generally have poor performance at high data rates or with many antennas.

We propose a high-rate coding scheme that can handle any configuration of transmit and receive antennas and that subsumes both V-BLAST and many proposed space-time block codes as special cases. The scheme transmits substreams of data in linear combinations over space and time. The codes are designed to optimize the mutual information between the transmitted and received signals. Because of their linear structure, the codes retain the decoding simplicity of V-BLAST, and because of their information-theoretic optimality, they possess many coding advantages. We give examples of the codes and show that their performance is generally superior to earlier proposed methods over a wide range of rates and SNR's.

Index Terms—Wireless communications, BLAST, multiple antennas, space-time codes, transmit diversity, receive diversity, fading channels

1 Introduction and Model

It is widely acknowledged that reliable fixed and mobile wireless transmission of video, data, and speech at high rates will be an important part of future telecommunications systems. One way to get high rates on a scattering-rich wireless channel is to use multiple transmit and/or receive antennas. In [1, 2], theoretical and experimental evidence demonstrates that channel capacity grows linearly as the number of transmit and receive antennas grow simultaneously.

Early uses of multiple transmit antennas in a scattering environment achieve reliability through "diversity," where redundant information is sent or received on two or more antennas in the hope that at least one path from the transmitter reaches the receiver [3, 4, 5, 6]. To keep the transmitter and receiver complexity low, linear processing is often preferred [3]. To achieve the high data rates promised in [2], however, new approaches for space-time transmission are needed. One such approach is presented in [7, 8] where a practical scheme, called V-BLAST (Vertical Bell Labs Layered Space-Time), encodes and decodes rates of tens of bits/sec/Hz with eight transmit and twelve receive antennas. The V-BLAST architecture breaks the original data stream into substreams that are transmitted on the individual antennas. The receiver decodes the substreams using a sequence of nulling and cancelling steps.

Since then there has been considerable work on a variety of space-time transmission schemes and performance measures [9] such as the space-time trellis codes of [10] and the space-time block codes of [11, 12] for the known channel and [13, 14, 15, 16, 17] for the unknown channel.

At very high rates and with a large number of antennas, many of these space-time codes suffer from complexity or performance difficulties. The number of states in the trellis codes of [10] grows exponentially with either the rate or the number of transmit antennas. The block codes of [11, 12] suffer from rate and performance loss as the number of antennas grow, and the codes of [14, 15, 16] suffer from decoding complexity if the rate is too high. Although V-BLAST can handle high data rates with reasonable complexity, the decoding scheme presented in [7] does not work with fewer receive than transmit antennas.

We present a space-time transmission scheme that has many of the coding and diversity advantages of previously designed codes, but also has the decoding simplicity of V-BLAST at high data rates. The codes work with arbitrary numbers of transmit and receive antennas.

The codes break the data stream into substreams that are dispersed in linear combinations over space and time. We refer to them simply as LD codes (linear dispersion codes). The LD codes:

1. Subsume, as special cases, both V-BLAST [7] and the block codes of [12]
2. Generally outperform both
3. Can be used for any number of transmit and receive antennas
4. Are very simple to encode
5. Can be decoded in a variety of ways including simple linear-algebraic techniques such as:
   (a) Successive nulling and cancelling (V-BLAST [7], square-root V-BLAST [18])
   (b) Sphere decoding [19, 20]
6. Are designed with the numbers of both the transmit *and* receive antennas in mind
7. Satisfy an information-theoretic optimality criterion
   - the codes are designed to maximize the mutual information between the transmit and receive signals

We briefly summarize the general structure of the LD codes. Suppose that there are M transmit antennas, N receive antennas, and an interval of T symbols available to us during which the propagation channel is constant and known to the receiver. The transmitted signal can then be written as a $T \times M$ matrix S that governs the transmission over the M antennas during the interval. We assume that the data sequence has been broken into Q substreams (for the moment we do not specify Q) and that $s_1, \ldots, s_Q$ are the complex symbols chosen from an arbitrary, say r-PSK or r-QAM, constellation. We call a rate $R = (Q/T)\log_2 r$ linear dispersion code one for which S obeys

$$S = \sum_{q=1}^{Q} \left(\alpha_q A_q + j\beta_q B_q\right), \tag{1}$$

where the real scalars $\{\alpha_q, \beta_q\}$ are determined by

$$s_q = \alpha_q + j\beta_q, \quad q = 1, \ldots Q.$$

The design of LD codes depends crucially on the choices of the parameters T, Q and the dispersion matrices $\{A_q, B_q\}$. To choose the $\{A_q, B_q\}$ we propose to optimize a nonlinear information-theoretic criterion: namely, the mutual information between the transmitted signals $\{\alpha_q, \beta_q\}$ and the received signal. We argue that this criterion is very important for achieving high spectral efficiency with multiple antennas. We also show how the information-theoretic optimization has implications for performance measures such as pairwise error probability. Section 4 has several examples of LD codes and performance comparisons with existing schemes.

We now present the multiple-antenna model considered in this paper.

1.1 The Multiple-Antenna Model

In a narrow-band, flat-fading, multi-antenna communication system with M transmit and N receive antennas, the transmitted and received signals are related by

$$x = \sqrt{\frac{\rho}{M}} H s + v, \tag{2}$$

where $x \in \mathcal{C}^N$ denotes the vector of complex received signals during any given channel use, $s \in \mathcal{C}^M$ denotes the vector of complex transmitted signals, $H \in \mathcal{C}^{N \times M}$ denotes the channel matrix, and the additive noise $v \in \mathcal{C}^N$ is $\mathcal{CN}(0,1)$ (zero-mean, unit-variance, complex-Gaussian) distributed that is spatially and temporally white. The channel matrix H and transmitted vector s are assumed to have unit variance entries, implying that

$$\mathrm{E}\,\mathrm{tr}\, HH^* = MN \quad \text{and} \quad \mathrm{E}\, s^* s = M.$$

Since the random quantities H, s, and v are independent, the normalization $\sqrt{\frac{\rho}{M}}$ in (2) ensures that ρ is the SNR at each receive antenna, independently of M. We often (but not always) assume that the channel matrix H also has independent $\mathcal{CN}(0,1)$ entries.

The entries of the channel matrix are assumed to be known to the receiver but not to the transmitter. This assumption is reasonable if training or pilot signals are sent to learn the channel, which is then constant for some coherence interval. The coherence interval of the channel should be large compared to M [21]. When the channel is known at the receiver, the resulting channel capacity (often referred to as the *perfect-knowledge* capacity) is [2, 1]

$$C(\rho, M, N) = \max_{R_s \geq 0, \mathrm{tr}\, R_s = M} \mathrm{E}\, \log\det\left(I_N + \frac{\rho}{M} H R_s H^*\right), \tag{3}$$

where the expectation is taken over the distribution of the random matrix H.[1] The capacity achieving s is a zero-mean complex Gaussian vector with covariance matrix $\mathrm{E}\, ss^* = R_{s,opt}$, where $R_{s,opt}$ is the maximizing covariance matrix in (3). When the distribution on H is rotationally invariant, i.e., when $p(H) = p(\Theta H) = p(H\Phi)$ for any unitary matrices Θ and Φ (as is the case when H has independent $\mathcal{CN}(0,1)$ entries), the optimizing covariance is $R_{s,opt} = I_M$, and (3) becomes

$$C(\rho, M, N) = \mathrm{E}\, \log\det\left(I_N + \frac{\rho}{M} HH^*\right). \tag{4}$$

This expectation can sometimes be computed in closed form (see, for example, [22]).

When the channel is constant for at least T channel uses we may write

$$x_\tau = \sqrt{\frac{\rho}{M}} H s_\tau + v_\tau, \quad \tau = 1, \ldots, T,$$

so that defining $X = \begin{bmatrix} x_1 & x_2 & \ldots & x_T \end{bmatrix}^t$, $S = \begin{bmatrix} s_1 & s_2 & \ldots & s_T \end{bmatrix}^t$ and $V = \begin{bmatrix} v_1 & v_2 & \ldots & v_T \end{bmatrix}^t$ (where the superscript ^{t} denotes "transpose"), we obtain

$$X^t = \sqrt{\frac{\rho}{M}} H S^t + V^t.$$

It is generally more convenient to write this equation in its transposed form

$$X = \sqrt{\frac{\rho}{M}} SH + V, \tag{5}$$

where we have omitted the transpose notation from H and simply redefined this matrix to have dimension $M \times N$. The matrix $X \in \mathcal{C}^{T\times N}$ is the received signal, $S \in \mathcal{C}^{T\times M}$ is the transmitted signal, and $V \in \mathcal{C}^{T\times N}$

[1] Equation (3) actually slightly generalizes [2, 1], which assume that H has independent $\mathcal{CN}(0,1)$ entries.

is the additive $\mathcal{CN}(0,1)$ noise. In X, S, and V, time runs vertically and space runs horizontally. We are concerned with designing the signal matrix S obeying the power constraint $\mathrm{E}\,\mathrm{tr}\,SS^* = TM$.

We note that, in general, the number of $T \times M$ matrices S needed in a codebook can be quite large. If the rate in bits per channel use is denoted R, then the number of matrices is 2^{RT}. For example, with $M = 4$ transmit and $N = 2$ receive antennas the channel capacity at $\rho = 20$ dB (with $\mathcal{CN}(0,1)$ distributed H) is more than 12 bits per channel use. Even with a relatively small block size of $T = 4$, we need $2^{48} \approx 10^{14}$ matrices at rate $R = 12$. The huge size of the constellations generally rules out the possibility of decoding at the receiver using exhaustive search.

The LD codes that we present easily generate the very large constellations that are needed. Moreover, because of their structure, they also allow efficient real-time decoding. In the next section, we briefly describe and analyze some existing space-time codes so that we may motivate the LD codes and explain how they are different.

2 Information-Theoretic Analysis of Some Space-Time Codes

Since the capacity of the multiple-antenna channel can easily be calculated, we may ask how well a space-time code performs by comparing the maximum mutual information that it can support to the actual channel capacity. The distribution for the $T \times M$ matrix S that achieves (4) is independent $\mathcal{CN}(0,1)$ entries, but we cannot easily use this by itself as a guideline for constructing and decoding a (random) constellation of $T \times M$ matrices because of the sheer number of matrices involved. Therefore, a constellation of matrices that has sufficient structure for efficient encoding and decoding is generally needed. One such structure is that of an *orthogonal design*, originally proposed in [11] and later generalized in [12].

2.1 Mutual information attainable with orthogonal designs

An orthogonal design is introduced by Alamouti in [11] for $T = M = 2$ and has the structure

$$S = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix}. \tag{6}$$

The complex scalars s_1 and s_2 are drawn from a particular (r-PSK or r-QAM) constellation, but we may simply assume that they are random variables such that $\mathrm{E}\left(|s_1|^2 + |s_2|^2\right) = 2$. We show that this particular structure can be used to achieve capacity when there is one receive antenna but *not* when there is more than one.

2.1.1 One receive antenna ($N = 1$)

With $N = 1$, equation (5) becomes

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \sqrt{\frac{\rho}{2}} \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

This can be rewritten as

$$\underbrace{\begin{bmatrix} x_1 \\ x_2^* \end{bmatrix}}_{\triangleq x} = \sqrt{\frac{\rho}{2}} \underbrace{\begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix}}_{\triangleq \mathcal{H}} \underbrace{\begin{bmatrix} s_1 \\ s_2 \end{bmatrix}}_{\triangleq s} + \begin{bmatrix} v_1 \\ v_2^* \end{bmatrix}. \tag{7}$$

It readily follows that

$$\mathcal{H}\mathcal{H}^* = \mathcal{H}^*\mathcal{H} = \left(|h_1|^2 + |h_2|^2\right) I_2. \tag{8}$$

We effectively have an equivalent matrix channel $\mathcal{H}$ in (7) that is a scaled unitary matrix. Maximum likelihood decoding of s_1 and s_2 is therefore decoupled [11].

We may ask how much mutual information the orthogonal design structure (6) can attain? To answer this question we need to compute the mutual information between the transmitted and received vectors s and x in the equivalent channel model (7) and compare it with the capacity of a $M = 2$, $N = 1$ multi-antenna system.

Since s has the power constraint $\mathrm{E}\, s^* s = 2$, the maximum mutual information in (7) is simply the channel capacity that is obtained with the structured channel matrix $\mathcal{H}$. If we denote this maximum mutual information by $C_{orth}(\rho)$, using (3) we obtain

$$\begin{aligned} C_{orth}(\rho) &= \max_{R_s \geq 0, \mathrm{tr}\, R_s = 2} \frac{1}{2} \mathrm{E} \log\det \left(I_2 + \frac{\rho}{2} \mathcal{H} R_s \mathcal{H}^* \right) \\ &= \max_{R_s \geq 0, \mathrm{tr}\, R_s = 2} \frac{1}{2} \mathrm{E} \log\det \left(I_2 + \frac{\rho}{2} \mathcal{H}^* \mathcal{H} R_s \right) \\ &= \max_{R_s \geq 0, \mathrm{tr}\, R_s = 2} \frac{1}{2} \mathrm{E} \log\det \left(I_2 + \frac{\rho}{2} (|h_1|^2 + |h_2|^2) R_s \right) \end{aligned}$$

where the factor $\frac{1}{2}$ in front of the expectation normalizes for the two channel uses spanned by the orthogonal design. Since, subject to a trace constraint, the determinant of any positive-definite matrix is maximized when its eigenvalues are all equal, it is easy to see that the maximizing covariance matrix is $R_{s,opt} = I_2$, so that we obtain

$$C_{orth}(\rho) = \mathrm{E} \log \left(1 + \frac{\rho}{2} (|h_1|^2 + |h_2|^2) \right) = C(\rho, M = 2, N = 1). \tag{9}$$

The expression on the right symbolically denotes the capacity attained by a system with $M = 2$ transmit antennas and $N = 1$ receive antennas at SNR ρ. This equation implies that the orthogonal design (6) can achieve the full channel capacity of the $M = 2$, $N = 1$ system, and there is no loss incurred by assuming the structure (6) as opposed to a general transmit matrix S.

2.1.2 Two or more receive antennas ($N \geq 2$)

With $N = 2$ receive antennas, equation (5) becomes

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = \sqrt{\frac{\rho}{2}} \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix} \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} + \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix},$$

which can be reorganized as

$$\underbrace{\begin{bmatrix} x_{11} \\ x_{21}^* \\ x_{12} \\ x_{22}^* \end{bmatrix}}_{\triangleq x} = \sqrt{\frac{\rho}{2}} \underbrace{\begin{bmatrix} h_{11} & h_{21} \\ h_{21}^* & -h_{11}^* \\ h_{12} & h_{22} \\ h_{22}^* & -h_{12}^* \end{bmatrix}}_{\triangleq \mathcal{H}} \underbrace{\begin{bmatrix} s_1 \\ s_2 \end{bmatrix}}_{\triangleq s} + \begin{bmatrix} v_{11} \\ v_{21}^* \\ v_{12} \\ v_{22}^* \end{bmatrix}. \tag{10}$$

We now readily see

$$\mathcal{H}^*\mathcal{H} = \left(|h_{11}|^2 + |h_{21}|^2 + |h_{12}|^2 + |h_{22}|^2\right) I_2. \tag{11}$$

As with $N = 1$, maximum-likelihood estimation of s_1 and s_2 is decoupled. However, unlike with $N = 1$, the orthogonal design structure prohibits us from achieving channel capacity. To see this, we compare the maximum mutual information between s and x in (10) with $C(\rho, M = 2, N = 2)$, the actual channel capacity for the system.

As before, the maximum mutual information in (10) is simply the channel capacity for the structured channel matrix $\mathcal{H}$. Denoting this maximum mutual information by $C_{orth}(\rho)$, we obtain

$$\begin{aligned} C_{orth}(\rho) &= \max_{R_s \geq 0, \mathrm{tr}\, R_s = 2} \frac{1}{2} \mathrm{E} \log\det \left(I_2 + \frac{\rho}{2}\mathcal{H}^*\mathcal{H}R_s\right) \\ &= \max_{R_s \geq 0, \mathrm{tr}\, R_s = 2} \frac{1}{2} \mathrm{E} \log\det \left(I_2 + \frac{\rho}{2}(|h_{11}|^2 + |h_{21}|^2 + |h_{12}|^2 + |h_{22}|^2)R_s\right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\mathrm{E}\,\log\det\left(I_2 + \frac{\rho}{2}(|h_{11}|^2 + |h_{21}|^2 + |h_{12}|^2 + |h_{22}|^2)I_2\right)\\
&= \mathrm{E}\,\log\left(1 + \frac{2\rho}{4}(|h_{11}|^2 + |h_{21}|^2 + |h_{12}|^2 + |h_{22}|^2)\right)\\
&= C(2\rho, M = 4, N = 1)\\
&< C(\rho, M = 2, N = 2). \qquad (12)
\end{aligned}
$$

The last equation implies that the orthogonal design (6) is restrictive and *does not* allow us to achieve the full channel capacity of the $M = 2$, $N = 2$ system, but rather the capacity of a $M = 4$, $N = 1$ system at twice the SNR. Thus, when $N = 2$ we take a loss with the structure (6). The amount of this loss is substantial at high SNR and is depicted in Figure 1 which shows the actual channel capacity compared to the maximum mutual information obtained by the orthogonal design (6).

For $N > 2$ receive antennas, the analysis is similar and is omitted. We simply state that for $M = 2$ transmit antennas and N receive antennas the 2×2 orthogonal design allows us to attain only $C(N\rho, M = 2N, N = 1)$, rather than the full $C(\rho, M = 2, N)$.

2.1.3 Other Orthogonal Designs

The preceding section focuses on the $M = 2$ orthogonal design but there also are orthogonal designs for $M > 2$. The complex orthogonal designs for $M > 2$ are no longer “full-rate” (see [12]) and therefore generally perform poorly in the maximum mutual information they can achieve, even when $N = 1$. We give an example of these non-square orthogonal designs [12, 23].

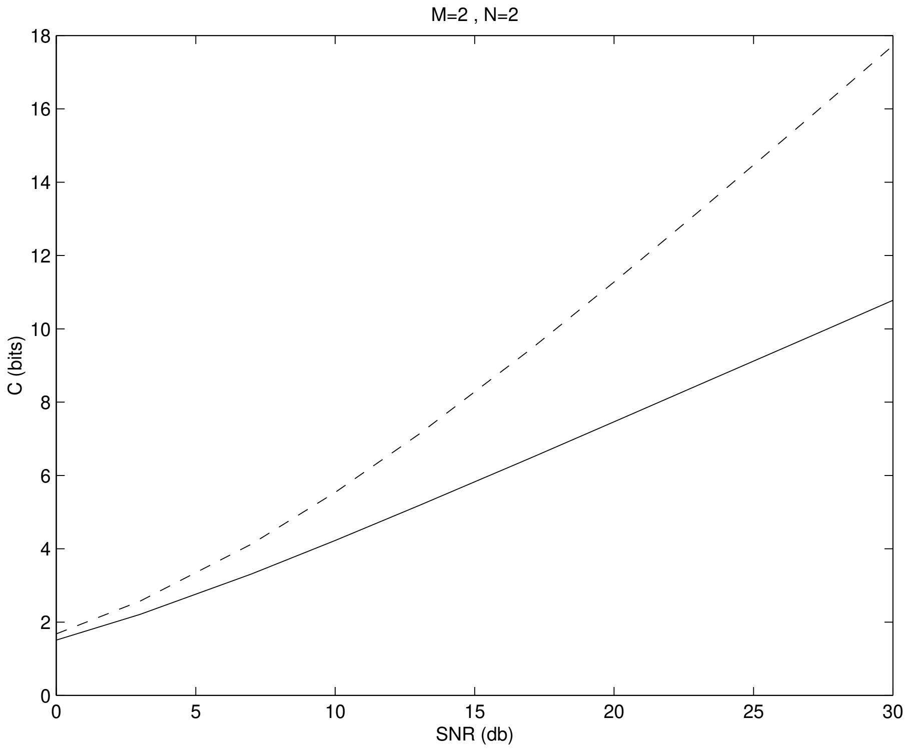

Figure 1: Maximum mutual information achieved by 2×2 orthogonal design (6) compared to actual channel capacity for the $M = 2$, $N = 2$ system. Solid line: maximum mutual information for 2×2 orthogonal design. Dashed line: capacity of $M = 2$, $N = 2$ system.

For $M = 3$, we have, for example, the rate $\frac{3}{4}$ orthogonal design

$$S = \sqrt{\frac{4}{3}} \begin{bmatrix} s_1 & s_2 & s_3 \\ -s_2^* & s_1^* & 0 \\ -s_3^* & 0 & s_1^* \\ 0 & -s_3^* & s_2^* \end{bmatrix}. \tag{13}$$

The factor $\sqrt{4/3}$ ensures that $\mathrm{E}\,\mathrm{tr}\,SS^* = T \cdot M = 12$. It can be shown that maximum-likelihood estimation of the variables s_1, s_2, s_3 is decoupled. Again using an argument similar to the one presented for $M = 2$, it is straightforward to show that the maximum mutual information attainable with (13) with N receive antennas is $\frac{3}{4}C(N\rho, M = 3N, N = 1)$ which is (much) less than the true channel capacity $C(\rho, M, N)$. We omit the proof and refer instead to Figure 2 which shows the actual channel capacity compared to the maximum mutual information obtained by the orthogonal design (13).

2.2 Mutual information attainable with V-BLAST

We show in Section 2.1 that, even though orthogonal designs allow efficient maximum-likelihood decoding and yield "full-diversity" (the appearance of the sum of the $|h_{ij}|^2$ in the mutual information formulas attests to this), orthogonal designs generally cannot achieve high spectral efficiencies in a multiple-antenna system, no matter how much coding is added to the transmitted signal constellation. This is especially true when the system has more than one receive antenna. An examination of the model (10) (and its counterparts for other orthogonal designs) reveals that the orthogonal design does not allow enough "degrees of freedom"—there are only two unknowns in (10) but four equations.

We can conclude that orthogonal designs are not suitable for very high-rate communications. On the other hand, a scheme that is proven to be suitable for high spectral efficiencies is V-BLAST [7]. In V-

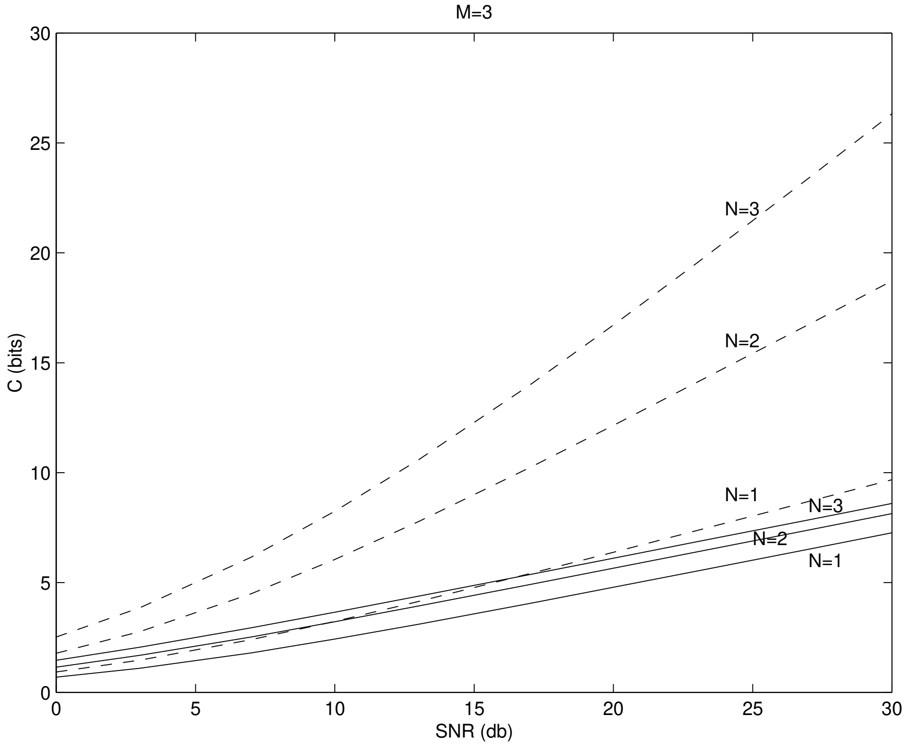

Figure 2: Maximum mutual information achieved by 4×3 orthogonal design (13) compared to actual channel capacity. Solid lines: maximum mutual information of 4×3 orthogonal design for $N = 1, 2, 3$ receive antennas. Dashed lines: capacity of $M = 3$, $N = 1, 2, 3$ systems.

BLAST each transmit antenna during each channel use sends an independent signal (often referred to as a substream). Thus, over a block of T channel uses, the $T \times M$ transmit matrix takes on the form

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1M} \\ s_{21} & s_{22} & \cdots & s_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ s_{T1} & s_{T2} & \cdots & s_{TM} \end{bmatrix}, \tag{14}$$

where each s_{ij} is an independent symbol drawn from a complex constellation. Since the transmitted symbols are not dispersed in time, T is arbitrary. (We could, for example, take $T = 1$.)

When $N \geq M$, (the number of receive antennas is at least as large as the number of transmit antennas), there exist efficient schemes for decoding the V-BLAST matrices. These are based on "successive nulling and cancelling" [7], and its more efficient variants [18], as well as more recently on sphere decoding [19]. All these decoding schemes essentially solve a well-conditioned system of linear equations. Successive nulling and cancelling provides a fast approximate solution to the maximum-likelihood decoding problem with the benefit of cubic complexity in the number of transmit antennas (M^3). Sphere decoding, on the other hand, finds the exact maximum likelihood estimate and benefits from avoiding an explicit exhaustive search. Recent work [20] has shown analytically that for a wide range of SNR's the expected computational complexity of sphere decoding is also roughly cubic in the number of transmit antennas.

Because there is no restriction on the transmitted matrix S in (14), the maximum mutual information that can be achieved by transmitting V-BLAST-like matrices is indeed the full multi-antenna channel capacity. Nevertheless, V-BLAST suffers from two deficiencies. First, nulling and cancelling fails when there are fewer receive antennas than transmit antennas, since the decoder is confronted with an *underdetermined* system of linear equations. Although sphere decoding can still be used to find the maximum likelihood estimate, the computational complexity is exponential in $M - N$. Second, because V-BLAST transmits independent data streams on its antennas there is no built-in spatial or temporal coding, and hence none of the error resilience associated with such coding. We seek to remedy these deficiencies in the next section.

3 Linear-Dispersion Space-Time Codes

In this section we propose a high-rate coding scheme that retains the decoding simplicity of V-BLAST, handles any configuration of transmit and receive antennas, and has many of the coding advantages of schemes such as the orthogonal designs without suffering the loss of mutual information.

We call a *linear-dispersion (LD) code* one for which

$$S = \sum_{q=1}^{Q} \left(s_q C_q + s_q^* D_q \right), \tag{15}$$

where $s_1, \ldots, s_Q$ are complex scalars (typically chosen from a r-PSK or r-QAM constellation) and where the C_q and D_q are *fixed* $T \times M$ complex matrices. The code is completely determined by the set of *dispersion* matrices $\{C_q, D_q\}$, whereas each individual codeword is determined by our choice of the scalars $s_1, \ldots, s_Q$.

We often find it more convenient to decompose the s_q into their real and imaginary parts,

$$s_q = \alpha_q + j\beta_q, \quad q = 1, \ldots Q$$

and to write

$$S = \sum_{q=1}^{Q} \left(\alpha_q A_q + j\beta_q B_q \right), \tag{16}$$

where $A_q = C_q + D_q$ and $B_q = C_q - D_q$. The dispersion matrices $\{A_q, B_q\}$ also specify the code.[2] The integer Q and the dispersion matrices are, for the moment, unspecified.

Without loss of generality, we assume that $\alpha_1, \ldots, \alpha_Q$ and $\beta_1, \ldots, \beta_Q$ have variance $\frac{1}{2}$ and are uncorrelated. Otherwise, we can always replace them with appropriate linear combinations that have this property—this simply leads to a redefinition of the A_q's and B_q's. Thus, $s_1, \ldots, s_Q$ are unit-variance and uncorrelated. Recall from our model in Section 1.1 that the transmit signal S is normalized such that $\mathrm{E}\,\mathrm{tr}\, SS^* = TM$. This induces the following normalization on the matrices $\{A_q, B_q\}$:

$$\sum_{q=1}^{Q} \left(\mathrm{tr}\, A_q^* A_q + \mathrm{tr}\, B_q^* B_q\right) = 2TM. \tag{18}$$

The dispersion codes (16) subsume as special cases both orthogonal designs and V-BLAST. For example, the 2×2 orthogonal design (6) corresponds to $T = M = Q = 2$ and

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{19}$$

whereas V-BLAST corresponds to $Q = TM$ and

$$A_{M(\tau-1)+m} = B_{M(\tau-1)+m} = \zeta_\tau \eta_m^t, \qquad \tau = 1, \ldots, T,\ m = 1, \ldots, M \tag{20}$$

where ζ_τ and η_m are T-dimensional and M-dimensional column vectors with one in the τth and mth entries, respectively, and zeros elsewhere.

[2] We remark that it is also possible to define $A_{Q+q} = jB_q$ and $\alpha_{Q+q} = \beta_q$, for $q = 1, \ldots, Q$, so that the LD codes become

$$S = \sum_{q=1}^{2Q} \alpha_q A_q, \tag{17}$$

where the scalars α_q are real.

Note that in V-BLAST each signal $\{\alpha_q, \beta_q\}$ is transmitted from only one antenna during only one channel use. With the LD codes, however, the dispersion matrices potentially transmit some combination of *each* symbol from *each* antenna at *every* channel use. This can lead to desirable coding properties. Before we specify good choices of the dispersion matrices, we discuss decoding.

3.1 Decoding

An important property of the LD codes (16) is their linearity in the variables $\{\alpha_q, \beta_q\}$, leading to efficient V-BLAST-like decoding schemes. To see this, it is useful to write the block equation

$$X = \sqrt{\frac{\rho}{M}} SH + V = \sqrt{\frac{\rho}{M}} \sum_{q=1}^{Q} \left(\alpha_q A_q + j\beta_q B_q\right) H + V \tag{21}$$

in a more convenient form. We decompose the matrices in (21) into their real and imaginary parts to obtain

$$X_R + jX_I = \sqrt{\frac{\rho}{M}} \sum_{q=1}^{Q} \left[\alpha_q \left(A_{R,q} + jA_{I,q}\right) + j\beta_q \left(B_{R,q} + jB_{I,q}\right)\right] \left(H_R + jH_I\right) + V_R + jV_I,$$

where $H_R = \mathrm{Re}\,(H)$ and $H_I = \mathrm{Im}\,(H)$. Equivalently,

$$\begin{aligned} X_R &= \sqrt{\frac{\rho}{M}} \sum_{q=1}^{Q} \left[\left(A_{R,q}H_R - A_{I,q}H_I\right)\alpha_q + \left(-B_{I,q}H_R - B_{R,q}H_I\right)\beta_q\right] + V_R \\ X_I &= \sqrt{\frac{\rho}{M}} \sum_{q=1}^{Q} \left[\left(A_{I,q}H_R + A_{R,q}H_I\right)\alpha_q + \left(B_{R,q}H_R - B_{I,q}H_I\right)\beta_q\right] + V_I. \end{aligned}$$

We denote the columns of X_R, X_I, H_R, H_I, V_R, and V_I by $x_{R,n}$, $x_{I,n}$, $h_{R,n}$, $h_{I,n}$, $v_{R,n}$, and $v_{I,n}$, and define

$$\mathcal{A}_q = \begin{bmatrix} A_{R,q} & -A_{I,q} \\ A_{I,q} & A_{R,q} \end{bmatrix}, \quad \mathcal{B}_q = \begin{bmatrix} -B_{I,q} & -B_{R,q} \\ B_{R,q} & -B_{I,q} \end{bmatrix}, \quad \underline{h}_n = \begin{bmatrix} h_{R,n} \\ h_{I,n} \end{bmatrix}, \tag{22}$$

where $n = 1, \ldots, N$. We then gather the equations in X_R and X_I to form the single real system of equations

$$\underbrace{\begin{bmatrix} x_{R,1} \\ x_{I,1} \\ \vdots \\ x_{R,N} \\ x_{I,N} \end{bmatrix}}_{\triangleq x} = \sqrt{\frac{\rho}{M}} \mathcal{H} \underbrace{\begin{bmatrix} \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_Q \\ \beta_Q \end{bmatrix}}_{\triangleq s} + \underbrace{\begin{bmatrix} v_{R,1} \\ v_{I,1} \\ \vdots \\ v_{R,N} \\ v_{I,N} \end{bmatrix}}_{\triangleq v}, \tag{23}$$

where the equivalent $2NT \times 2Q$ real channel matrix is given by

$$\mathcal{H} = \begin{bmatrix} \mathcal{A}_1 \underline{h}_1 & \mathcal{B}_1 \underline{h}_1 & \cdots & \mathcal{A}_Q \underline{h}_1 & \mathcal{B}_Q \underline{h}_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{A}_1 \underline{h}_N & \mathcal{B}_1 \underline{h}_N & \cdots & \mathcal{A}_Q \underline{h}_N & \mathcal{B}_Q \underline{h}_N \end{bmatrix}. \tag{24}$$

We have a linear relation between the input and output vectors s and x

$$x = \sqrt{\frac{\rho}{M}} \mathcal{H} s + v, \tag{25}$$

where the equivalent channel $\mathcal{H}$ is known to the receiver because the original channel H, and the dispersion matrices $\{A_q, B_q\}$ are all known to the receiver. The receiver simply uses (24) to find the equivalent channel. The system of equations between transmitter and receiver is not underdetermined as long as

$$Q \leq NT. \tag{26}$$

We may therefore use any decoding technique already in place for V-BLAST, such as successive nulling and cancelling [7], its efficient square-root implementation [18], or sphere decoding [19]. The most efficient implementations of these schemes generally require $O(Q^3)$ computations, and have roughly constant complexity in the size of the signal constellation r [20].

3.2 Design of the Dispersion Codes

Although we have introduced the LD structure

$$S = \sum_{q=1}^{Q} \left(\alpha_q A_q + j\beta_q B_q \right),$$

we have not yet specified Q or the dispersion matrices $A_1, \ldots, A_Q$ and $B_1, \ldots, B_Q$. We have the inequality $Q \leq NT$. Intuitively, the larger Q is, the higher the maximum mutual information between s and v is since the matrix signal S has more degrees of freedom. (Recall that orthogonal designs generally have low mutual information because they do not have enough degrees of freedom.) On the other hand, the smaller Q is, the more of a coding effect we obtain since the equivalent matrix $\mathcal{H}$ becomes "skinnier" and the system of equations in (23) becomes more overdetermined. As a general practice, we find it useful to take $Q = \min(M,N) \cdot T$ since this tends to maximize the mutual information between s and x while still having the benefit of coding gain.[3]

We are left with the question of how to design the dispersion matrices. We may first examine how sensitive the performance of the LD codes is to the choice of the dispersion matrices. Experiments with choosing random dispersion matrices subject to the normalization constraint (18), or the more stringent constraints $\mathrm{tr}\, A_q^* A_q = \mathrm{tr}\, B_q^* B_q = \frac{TM}{Q}$, for $q = 1, \ldots, Q$, suggest that the performance for "average" $\{A_q, B_q\}$ is not generally very good. Figure 3 shows the bit error rate of an $M = 3$, $N = 1$ antenna system

[3] At high SNR, the capacity of the multiple-antenna system grows as $\min(M,N) \log \rho$, suggesting that we need $K = \min(M,N)$ degrees of freedom per channel use.

with randomly chosen versus optimized (according to a criterion we specify shortly) dispersion matrices. The difference is dramatic; it is important to choose the dispersion matrices wisely.

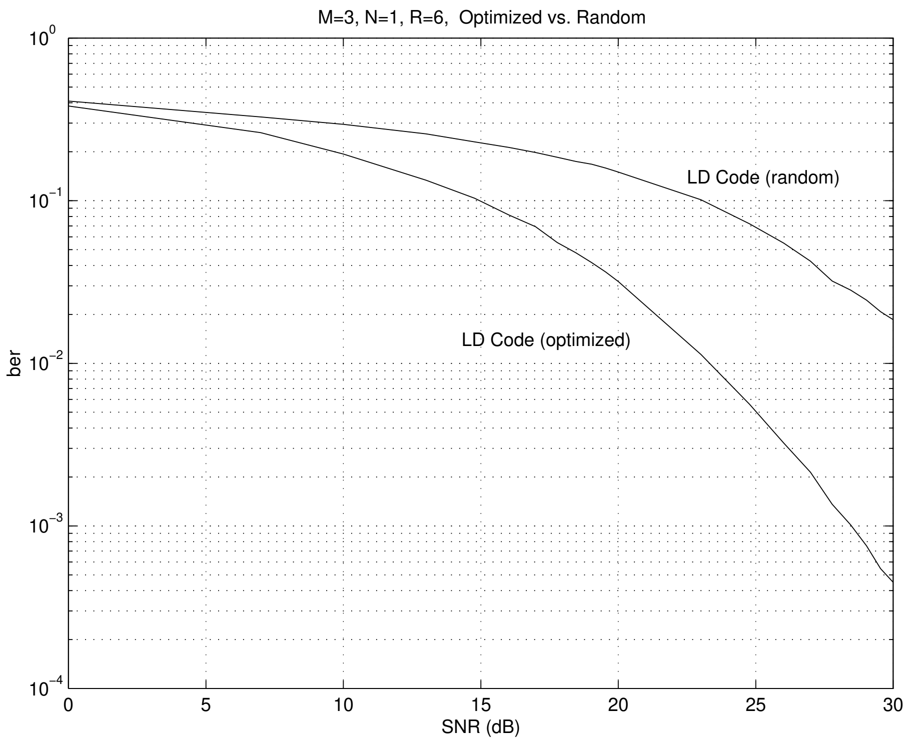

Figure 3: Bit error performance comparison for a random ($\{A_q, B_q\}$ drawn from a complex Gaussian distribution and normalized) and an optimized LD code for $M = 3$ transmit and $N = 1$ receive antenna for $T = Q = 4$, and rate $R = 6$ bits/channel use (obtained by transmitting 64-QAM on $s_1, \ldots, s_4$).

One possible way of designing the spreading matrices is to study the pairwise probability of error of two different LD codewords, say,

$$S = \sum_{q=1}^{Q} (\alpha_q A_q + j\beta_q B_q) \quad \text{and} \quad S' = \sum_{q=1}^{Q} \left(\alpha_q' A_q + j\beta_q' B_q\right).$$

The worst-case pairwise error is generally obtained when $\{\alpha_1, \ldots, \alpha_Q, \beta_1, \ldots, \beta_Q\}$ and $\{\alpha_1', \ldots, \alpha_Q', \beta_1', \ldots, \beta_Q'\}$ differ in only one element. We can then seek to choose the dispersion matrices that minimizes the probability of this error. The main drawback of this strategy is that it leads to a criterion on the individual columns of the matrix $\mathcal{H}$, rather than on the matrix in its entirety. Therefore it is conceivable that designs based on this criterion could lead to a (near) singular $\mathcal{H}$, leading to other forms of errors. Finally, it is not clear what effect minimizing pairwise error probability has on the overall error probability, especially for a high-rate system. Therefore, this strategy for choosing the dispersion matrices does not appear to be promising.

We can also study the *average* pairwise error probability, obtained by choosing Gaussian s in (25) and averaging the pairwise error obtained between an independent s and s'. We show in Appendix B that the average pairwise error has upper bound

$$P_e(\text{pairwise}) \leq \mathrm{E}\,\frac{1}{2}\det(I + \frac{\rho}{2M}\mathcal{H}^t\mathcal{H})^{-1/2}. \tag{27}$$

We can then seek to minimize the upper bound with an appropriate choice of $\{A_1, \ldots, A_Q\}$ and $\{B_1, \ldots, B_Q\}$. Even though (27) is a simple formula, suggesting that it can possibly be minimized, we do not attempt to do so here. The main reason is the following. Since multiple antennas are used for very high rates, the pairwise error probability for any two signals is extremely small. In Section 1.1 we argue that even for the small test-case of $M = 4$ transmit and $N = 2$ receive antennas, we could theoretically have a constellation size of as many as 2^{48} signal-matrices at $\rho = 20$ dB. It is therefore conceivable that the pairwise error probability between any two could be roughly $2^{-48} \approx 10^{-14}$. Trying to minimize a quantity such as (27) that is already so small can be numerically quite difficult.

Fortunately, information theory suggests a natural alternative that is connected with minimizing (27) but is more fundamental. Recall from Section 2.1 that orthogonal designs are deficient in the maximum mutual information they support for $M > 2$ or $N > 1$. We therefore choose $\{A_q, B_q\}$ to maximize the mutual information between s and x in (23). This guarantees that we are taking the smallest possible mutual information penalty within the LD structure (16). We propose to design codes that are "blessed" by the "logdet" formula (3).

We formalize the design criterion as follows.

The Design Method

1. Choose $Q \le NT$ (typically, $Q = \min(M, N) \cdot T$).

2. Choose $\{A_q, B_q\}$ that solve the optimization problem

$$C_{\mathrm{LD}}(\rho, T, M, N) = \max_{A_q, B_q, q=1,\ldots Q} \frac{1}{2T} \mathrm{E}\, \log\det\left(I_{2NT} + \frac{\rho}{M}\mathcal{H}\mathcal{H}^t\right) \tag{28}$$

   for an SNR ρ of interest, subject to one of the following constraints

   (i) $\sum_{q=1}^{Q} \left(\mathrm{tr}\, A_q^* A_q + \mathrm{tr}\, B_q^* B_q\right) = 2TM$

   (ii) $\mathrm{tr}\, A_q^* A_q = \mathrm{tr}\, B_q^* B_q = \frac{TM}{Q}$, $q = 1, \ldots Q$

   (iii) $A_q^* A_q = B_q^* B_q = \frac{T}{Q} I_M$, $q = 1, \ldots Q$

   where $\mathcal{H}$ is given by (24) with the $\underline{h}_n$ having independent $\mathcal{N}(0, \frac{1}{2})$ entries.

Note that equation (28) is effectively equation (3) with $R_s = I_{2Q}$; as mentioned in Section 3, we may take the entries of s (the α_q's and β_q's) to be uncorrelated with variance $\frac{1}{2}$. Moreover, because the real and imaginary parts of the noise vector v in (23) also have variance $\frac{1}{2}$, the SNR remains ρ. We also note that equation (28) differs from (3) by the outside factor $\frac{1}{2T}$ because the effective channel is real-valued and the LD code spans T channel uses.

We now make some remarks:

1. Clearly, $C_{\mathrm{LD}}(\rho, T, M, N) \le C(\rho, M, N)$.

2. The problem (28) can be solved subject to any of the constraints (i)–(iii). Constraint (i) is simply the power constraint (18) that ensures $\mathrm{E}\,\mathrm{tr}\, SS^* = TM$. Constraint (ii) is more restrictive and ensures that each of the transmitted signals α_q and β_q are transmitted with the same overall power from the M antennas during the T channel uses. Finally, constraint (iii) is the most stringent, since it forces the symbols α_q and β_q to be dispersed with equal energy in all spatial and temporal directions.

3. Since constraints (i)–(iii) are successively more restrictive, the corresponding maximum mutual informations obtained in (28) are necessarily successively smaller. Nevertheless, we have found that constraint (iii) generally imposes only a small information-theoretic penalty while having the advantage of better coding (or diversity) gains. Using symmetry arguments one may conjecture that the optimal solution to the problem with constraint (i) should automatically satisfy the constraint (ii). But we have not experimented sufficiently with constraint (i) to confirm this; we instead usually restrict our attention to constraints (ii) and (iii). We have empirically found that of two codes with equal mutual informations, the one satisfying the more stringent constraint gives lower error rates. Examples of this phenomenon appear in Section 4.

4. The solution to (28) subject to any of the constraints (i)–(iii) is highly nonunique: simply reordering the $\{A_q, B_q\}$ gives another solution, as does pre- or post-multiplying all the matrices by the same unitary matrix. However, there is also another source of non-uniqueness which is more subtle. Equation (23) shows that we can always pre-multiply the transmit vector $s = \begin{bmatrix} \alpha_1 & \beta_1 & \dots & \alpha_Q & \beta_Q \end{bmatrix}^t$ by a $2Q \times 2Q$ orthogonal matrix Φ^t to obtain a new vector $s' = \Phi^t s = \begin{bmatrix} \alpha_1' & \beta_1' & \dots & \alpha_Q' & \beta_Q' \end{bmatrix}^t$ with entries that are still independent and $\mathcal{N}(0, \frac{1}{2})$-distributed. Thus, we may rewrite (23) as

$$x = \sqrt{\frac{\rho}{M}}\mathcal{H}\Phi\Phi^t s = \sqrt{\frac{\rho}{M}}\mathcal{H}\Phi s'.$$

Defining $\mathcal{A}_q$, $\mathcal{B}_q$, and $\underline{h}_n$ as in (22). allows us to write the new equivalent channel $\mathcal{H}' = \mathcal{H}\Phi$ as

$$\mathcal{H}' = \begin{bmatrix} \mathcal{A}_1\underline{h}_1 & \dots & \mathcal{B}_Q\underline{h}_1 \\ \vdots & \ddots & \vdots \\ \mathcal{A}_1\underline{h}_N & \dots & \mathcal{B}_Q\underline{h}_N \end{bmatrix} \begin{bmatrix} \phi_{1,1} & \dots & \phi_{1,2Q} \\ \vdots & \ddots & \vdots \\ \phi_{2Q,1} & \dots & \phi_{2Q,2Q} \end{bmatrix} = \\ \begin{bmatrix} \sum_{p=1}^{Q}\left(\mathcal{A}_p\phi_{2p-1,1} + \mathcal{B}_p\phi_{2p,1}\right)\underline{h}_1 & \dots & \sum_{p=1}^{Q}\left(\mathcal{A}_p\phi_{2p-1,2Q} + \mathcal{B}_p\phi_{2p,2Q}\right)\underline{h}_1 \\ \vdots & \ddots & \vdots \\ \sum_{p=1}^{Q}\left(\mathcal{A}_p\phi_{2p-1,1} + \mathcal{B}_p\phi_{2p,1}\right)\underline{h}_N & \dots & \sum_{p=1}^{Q}\left(\mathcal{A}_p\phi_{2p-1,2Q} + \mathcal{B}_p\phi_{2p,2Q}\right)\underline{h}_N \end{bmatrix}. \tag{29}$$

Since the entries of s and s' have the same joint distribution, the maximum mutual information obtained from the equivalent channels $\mathcal{H}$ and $\mathcal{H}'$ are the same. This implies that the transformation from the dispersion matrices $\{A_q, B_q\}$ to $\{A'_q, B'_q\}$

$$\begin{aligned} A'_q &= \textstyle\sum_{p=1}^{Q}\left(A_p\phi_{2p-1,2q-1} + B_p\phi_{2p,2q-1}\right) \\ B'_q &= \textstyle\sum_{p=1}^{Q}\left(A_p\phi_{2p-1,2q} + B_p\phi_{2p,2q}\right) \end{aligned} \qquad q = 1, \dots Q \tag{30}$$

where $\Phi = \{\phi_{i,j}\}_{i,j=1}^{2Q}$ is an orthogonal matrix, preserves the mutual information. Thus, the transformation (30) is another source of nonuniqueness to the solution of (28).

This nonuniqueness can be used to our advantage because a judicious choice of the orthogonal matrix Φ allows us to change the dispersion code through the transformation (30) to satisfy other criteria (such as space-time diversity) without sacrificing mutual information. Examples of this appear in Remark 7, where we construct unitary $\{A_q, B_q\}$ from the rank-one V-BLAST dispersion matrices (20), and in Section 4 in some of the two and three-antenna LD code constructions.

5. The constraints (i)–(iii) are convex in the dispersion matrices $\{A_q, B_q\}$ since they can be rewritten as

(i′) $\sum_{q=1}^{Q}\left(\operatorname{tr} A_q^* A_q + \operatorname{tr} B_q^* B_q\right) \leq 2TM$

(ii′) $\operatorname{tr} A_q^* A_q \leq \frac{TM}{Q}$, $\operatorname{tr} B_q^* B_q \leq \frac{TM}{Q}, \quad q = 1, \ldots Q$

(iii′) $A_q^* A_q \leq \frac{T}{Q} I_M$, $B_q^* B_q \leq \frac{T}{Q} I_M, \quad q = 1, \ldots Q$

all of which are convex. However, the cost function $\frac{1}{2T}\mathrm{E}\ \log\det\left(I_{2NT} + \frac{\rho}{M}\mathcal{H}\mathcal{H}^t\right)$ is neither concave nor convex in the variables $\{A_q, B_q\}$. Therefore, it is possible that (28) has local maxima. Nevertheless, we have been able to solve (28) with relative ease using gradient-based methods and it does not appear that local maxima pose a great problem. Table 1 gathers the maximum mutual informations obtained via gradient-ascent for a variety of different T, M, and N. The results show that maximum mutual informations obtained are quite close to the Shannon capacity (which is clearly an upper bound on what can be achieved) and so they suggest that the values obtained, if not the global maxima, are quite close to them. (For convenience, we include the gradient of the cost function (28) in Appendix A.)

6. We know that for $T = 2$, $M = 2$, $N = 1$, one solution to (28), for any of the constraints (i)–(iii), is the orthogonal design (19). This holds simply because the mutual information of this particular orthogonal design achieves the actual channel capacity $C(\rho, M = 2, N = 1)$. We note that there are also many other solutions that work equally well.

7. When $N \geq M$ and $Q = MT$ one solution to (28), subject to either constraints (i) or (ii), is given by the V-BLAST matrices (20) since these achieve the full capacity of the multiple antenna link. The V-BLAST matrices, however, are rank-one and therefore do not satisfy constraint (iii). But it is also possible to obtain an explicit solution to (28) subject to (iii). For $T = M$, one such set of matrices is

given by

$$A'_{M(k-1)+l} = B'_{M(k-1)+l} = \frac{1}{\sqrt{M}} D^{k-1}\Pi^{l-1}, \quad k = 1, \ldots, M, \ l = 1, \ldots, M \tag{31}$$

where

$$D = \begin{bmatrix} 1 & 0 & \cdots & & 0 \\ 0 & e^{j\frac{2\pi}{M}} & 0 & & \cdots \\ \vdots & & \ddots & & \\ 0 & & & & e^{j\frac{2\pi(M-1)}{M}} \end{bmatrix}, \qquad \Pi = \begin{bmatrix} 0 & \cdots & & 0 & 1 \\ 1 & 0 & \cdots & & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}.$$

The above code can be constructed by starting with the V-BLAST matrices (20) and applying the transformation (30) with a suitable Φ. We do not give the full Φ here, and only mention that, for $T = M = 2$, the transformation is

$$A'_1 = \frac{A_1 + A_4}{\sqrt{2}} \ , \ A'_2 = \frac{A_2 + A_3}{\sqrt{2}} \ , \ A'_3 = \frac{A_1 - A_4}{\sqrt{2}} \ , \ A'_4 = \frac{A_2 - A_3}{\sqrt{2}}$$

with similar expressions for the $\{B_q\}$. It can be readily checked that the matrix Φ constructed from the coefficients relating $\{A'_q, B'_q\}$ to $\{A_q, B_q\}$ is orthogonal. Figure 4 in Section 4 presents a performance comparison of the LD code (31) with V-BLAST.

8. The block length T is essentially also a design variable. Although it must be chosen shorter than the coherence time of the channel, it can be varied to help the optimization (28). We have found that choosing $M \leq T \leq 2M$ often yields good performance.

9. Although the SNR ρ is a design variable, we have found that the optimization (28) is not sensitive to its value for large ρ ($\rho \approx 20$ dB). Once the optimization is performed, the resulting LD code generally

works well over a wide range of SNR's.

10. It does not appear that (28) has a simple closed-form solution for general T, M, N, although we see in Section 4 that, in some non-trivial cases, it can lead to solutions with simple structure. We have found that the solution to (28) often yields an equivalent channel matrix $\mathcal{H}$ that is "as orthogonal as possible." Although complete orthogonality appears not always to be possible, our experience with optimizing (28) shows that the difference $C(\rho, M, N) - C_{\mathrm{LD}}(\rho, T, M, N)$ can be made quite small with a proper choice of T and Q (see Table 1 in Section 4.1). Thus, there appears to be very little capacity penalty in assuming the LD structure (16).

11. When the equivalent channel matrix $\mathcal{H}$ is orthogonal, maximum-likelihood decoding and the V-BLAST-like nulling/cancelling [7] perform equally well because the estimation errors of $s_1, \ldots, s_Q$ are decoupled.

12. The design criterion (28) *depends explicitly on the number of receive antennas* N, both through the choice of Q and through the matrix $\mathcal{H}$ in (24). Hence, the optimal codes, for a given ρ, T, and M, are different for different N.

    Nevertheless, a code designed for N receive antennas can also easily be decoded using nulling/cancelling or sphere decoding with $N' > N$ antennas. Hence, if we wish to broadcast data to more than one user, we may use a code designed for the user with the fewest receive antennas, with a rate supported by all the users.

13. The ultimate rate of the code depends on the number of signals sent Q, the block length of the code T, and the size of the constellation from which $s_1, \ldots, s_Q$ are chosen. We assume that the constellation

is r-PSK or r-QAM. Then the rate in bits/channel use is easily seen to be

$$R = \frac{Q}{T} \log_2 r. \tag{32}$$

14. A standard gray-code assignment of bits to the symbols of the r-PSK or r-QAM constellation may be used.

15. We see that the average pairwise error probability (27) and the design criterion (28) have a similar expression. By interchanging the expectation and log in (28), we see that maximizing (28) has some connections to minimizing (27).

    On the other hand, our design criterion is not directly connected with the diversity design criterion given in [9] and [10], which is concerned with maximizing

$$\min_{S \neq S'} |\det(S - S')^*(S - S')|. \tag{33}$$

    A constellation attains full diversity if (33) is nonzero. This criterion depends only on matrix pairs, and therefore does not exclude matrix designs with low spectral efficiencies.

    At high spectral efficiencies, the number of signals in the constellation of possible S matrices is roughly exponential in the channel capacity at a given SNR. This number can be very large—in Section 4 we present a code for $T = M = 8$ and $R = 16$ that effectively has $2^{RT} = 2^{128} \approx 3.4 \times 10^{38}$ matrices. The relation between the diversity criterion and the performance of such a large constellation is very tenuous. Even if $\det(S - S')^*(S - S') = 0$ for many pairs of $S \neq S'$, our probability of encountering one of these matrices may still be exceedingly small, and the constellation performance may still be excellent. For such a large constellation it is probably more important for the matrices in this constellation to be distributed in the space of matrices according to the distribution that attains capacity; the mutual information criterion attempts to achieve this distribution.

As the SNR is allowed to increase, the performance of some given space-time code with some given rate becomes more dependent on the diversity criterion since making a decoding error to a "nearest neighbor" becomes relatively more important. Chernoff bound computations in [10] show that the pairwise error falls off as $1/\rho^{\nu}$, where ν is the rank of $S - S'$. However, by increasing the SNR and keeping the code (and hence rate) fixed, we are effectively reducing the relative spectral efficiency of the code as compared with the channel capacity. We are therefore led again to the conclusion that diversity plays a secondary role at high spectral efficiencies. In Section 4, we present a comparison of codes that satisfy various combinations of the mutual information and diversity criteria. The code that satisfies both criteria performs best, followed by the mutual information criterion only, followed by the diversity criterion only.

16. Although the dispersion matrices $\{A_q, B_q\}$ can, in general, be complex, we have found that constraining them to be real imposes little, if any, penalty in the optimized mutual information.

17. Our mutual information calculations and design examples assume that the channel matrix H has independent $\mathcal{CN}(0,1)$ entries, but designs for other channel distributions using the mutual information criterion are also possible.

4 Examples of LD Codes and Performance

In this section we present simulations that compare the performance of LD codes to V-BLAST and orthogonal designs over a wide range of SNR's and various combinations of receive and transmit antennas. All the LD codes are designed for a target SNR of $\rho = 20$ dB (see Remark 9 in Section 3.2).

LD vs. V-BLAST: $M = 2$, $N = 2$, $R = 4$

We look first at an $M = 2$, $N = 2$ system at rate $R = 4$ and compare V-BLAST with an LD code. In V-BLAST, $Q = 4$, and the matrices are given by (20). To design an LD code we also choose $Q = 4$ but use the matrices given by (31) that satisfy constraint (iii) in (28). To achieve $R = 4$, we transmit QPSK on $s_1, \ldots, s_4$. The results can be seen in Figure 4 where the bit errors are compared. Even though both V-BLAST and the LD code support the full channel capacity, which is 11.28 bits/channel use at SNR=20 dB, the LD code has better performance; this can probably be attributed to the spatial and temporal dispersion of the symbols that V-BLAST lacks.

Since we are transmitting at rate $R = 4$ our spectral efficiency is low relative to the channel capacity, and we may therefore anticipate significant coding advantages from also satisfying the diversity criterion (33)—see Remark 15 in Section 3.2 for an explanation of the relative importance of diversity at low spectral efficiencies. The LD code (31) may be modified as in Remark 4 in Section 3.2, without changing its mutual information, by premultiplying the transmitted signal vector by an orthogonal matrix Φ^t. In [24] it is shown that full diversity is achieved for this two-antenna code transmitting QPSK if we choose Φ to be the block diagonal matrix

$$\Phi = \mathrm{diag}\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} \phi_R & \phi_I \\ -\phi_I & \phi_R \end{bmatrix}, \begin{bmatrix} \theta_R & \theta_I \\ -\theta_I & \theta_R \end{bmatrix}, \begin{bmatrix} (\phi\theta)_R & (\phi\theta)_I \\ -(\phi\theta)_I & (\phi\theta)_R \end{bmatrix} \right), \tag{34}$$

(where the subscript "R" denotes real part, and "I" denotes imaginary part) where $\theta^2 = \phi$, and $\phi = e^{i/2}$. The result is a code that satisfies both the mutual information criterion and diversity criterion; it is also displayed in Figure 4 and has the best performance. Although the codes in the figure all satisfy the mutual information condition, the importance of also satisfying the diversity criterion at relatively low spectral efficiencies is underscored. The next example shows that satisfying the mutual information condition is most important at

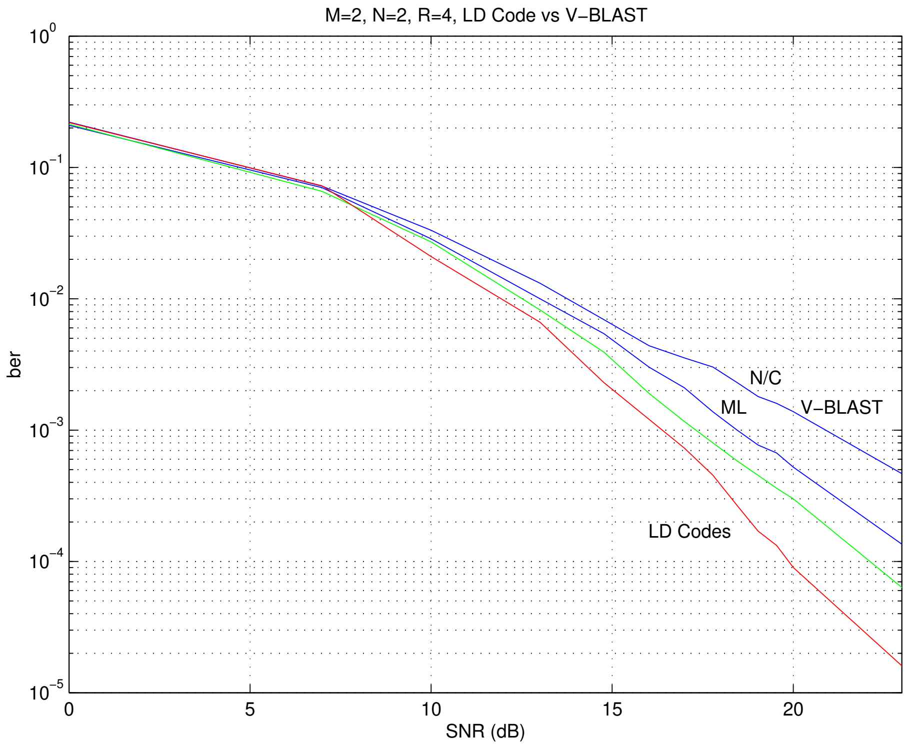

Figure 4: The upper two curves are the bit error performance of V-BLAST (20) with nulling/cancelling (upper), and with maximum likelihood decoding (lower). The lower two curves are the LD code given by (31) for $M = N = T = 2$ and $Q = 4$ (upper) and its modification (34) with $\phi = e^{i/2}$ (lower). For both these codes sphere decoding is used to find the maximum likelihood estimates. The rate is $R = 4$, and is obtained by transmitting QPSK on $s_1, \ldots, s_4$.

higher spectral efficiencies.

LD vs. OD: $M = N = 2$**,** $R = 8$

We show in Section 2.1.2 that the 2×2 orthogonal design is deficient in mutual information when $N > 1$. This deficiency should be reflected in its performance with $N = 2$. We test its performance when $M = N = 2$ at $R = 8$ versus the LD code given by (31) for $T = 2$ and $Q = 4$. The result can be seen in Figure 5 which clearly shows the better performance of the LD code over a wide range of SNR's. To achieve $R = 8$, we see from (32) that the orthogonal design needs to choose s_1 and s_2 from a 256-QAM constellation, while the LD code can choose from a 16-QAM constellation because it has four symbols $s_1, \ldots, s_4$. We note that the orthogonal design has good diversity (33) [12] but achieves only 7.47 bits/channel use mutual information at $\rho = 20$ dB, while the LD code achieves the full channel capacity of 11.28 bits/channel use. The orthogonal design and LD code are maximum-likelihood decoded (using the sphere decoder in the case of the LD code).

As in the previous example, we get an even better LD code by using (34) to improve its diversity without changing its mutual information. We can, however, anticipate only limited gains at this higher spectral efficiency. As shown in [24], setting $\phi = e^{0.488i}$ is a good choice when transmitting 16-QAM. The performance of this constellation is also shown in Figure 5. Its performance is better than the unmodified LD code at high SNR. Clearly, the best code satisfies both the mutual information and diversity criteria, if possible.

LD vs. OD: $M = 3$**,** $N = 1$**,** $R = 6$

We present a code for $M = 3$ transmit antennas and $N = 1$ receive antennas and compare it with the orthogonal design presented in Section 2.1.3 with blocklength $T = 4$. The orthogonal design (13) is written

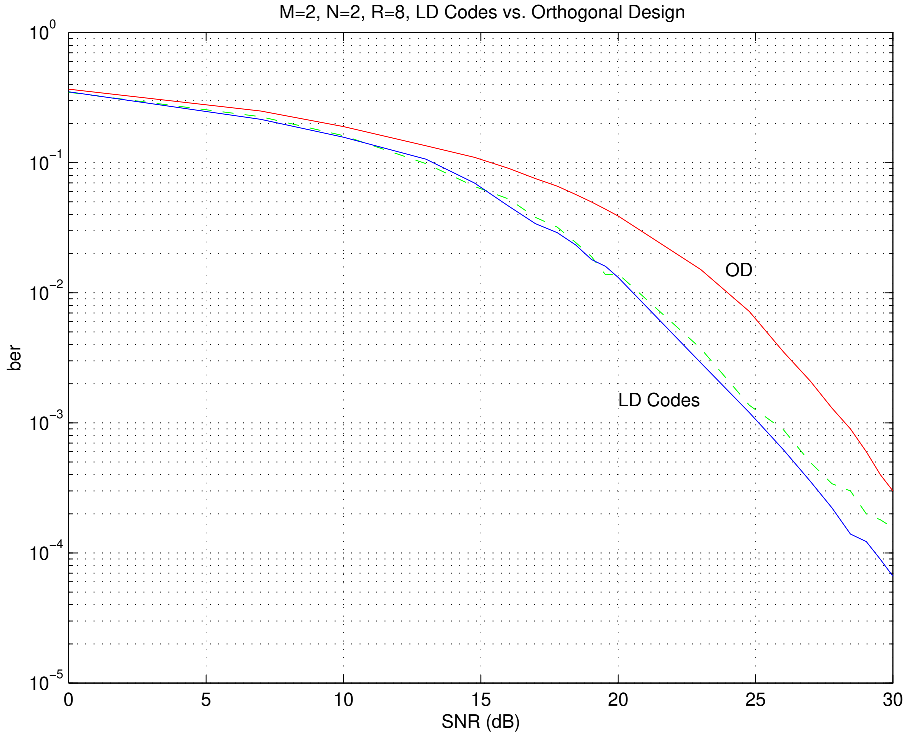

Figure 5: Bit error performance of the 2×2 orthogonal design (19) and the LD codes given by (31) for $M = N = T = 2$ and $Q = 4$ (dashed line) and its modification (34) with $\phi = e^{0.488i}$ (lower solid line). The rate is $R = 8$ implying that the orthogonal design transmits 256-QAM whereas the LD codes transmit 16-QAM. The decoding is maximum likelihood in both cases.

in terms of $\{\alpha_q\}$ and $\{\beta_q\}$ as

$$S = \sqrt{\frac{4}{3}} \begin{bmatrix} \alpha_1 + j\beta_1 & \alpha_2 + j\beta_2 & \alpha_3 + j\beta_3 \\ -\alpha_2 + j\beta_2 & \alpha_1 - j\beta_1 & 0 \\ -\alpha_3 + j\beta_3 & 0 & \alpha_1 - i\beta_1 \\ 0 & -\alpha_3 + j\beta_3 & \alpha_2 - i\beta_2 \end{bmatrix}. \tag{35}$$

It turns out that this orthogonal design is a local maximum to (28) for $T = 4$ and $Q = 3$. It achieves a mutual information of 5.13 bits/channel use at $\rho = 20$ dB, whereas the channel capacity is 6.41 bits/channel use.

To find an LD code with the same blocklength, we first observe that Q must obey the constraint $Q \leq N \cdot T$, with $N = 1$ and $T = 4$. Therefore $Q \leq 4$, and to obtain the highest possible mutual information we choose $Q = 4$. After optimizing (28) using a gradient-based search (Appendix A) and converging to a local maximum at $\rho = 20$ dB, we find

$$S = \begin{bmatrix} \alpha_1+\alpha_3+j[\frac{\beta_2+\beta_3}{\sqrt{2}}+\beta_4] & \frac{\alpha_2-\alpha_4}{\sqrt{2}}-j[\frac{\beta_1}{\sqrt{2}}+\frac{\beta_2-\beta_3}{2}] & 0 \\ \frac{-\alpha_2+\alpha_4}{\sqrt{2}}-j[\frac{\beta_1}{\sqrt{2}}+\frac{\beta_2-\beta_3}{2}] & \alpha_1-j\frac{\beta_2+\beta_3}{\sqrt{2}} & -\frac{\alpha_2+\alpha_4}{\sqrt{2}}+j[\frac{\beta_1}{\sqrt{2}}-\frac{\beta_2-\beta_3}{2}] \\ 0 & \frac{\alpha_2+\alpha_4}{\sqrt{2}}+j[\frac{\beta_1}{\sqrt{2}}-\frac{\beta_2-\beta_3}{2}] & \alpha_1-\alpha_3+j[\frac{\beta_2+\beta_3}{\sqrt{2}}-\beta_4] \\ \frac{\alpha_2-\alpha_4}{\sqrt{2}}+j[\frac{\beta_1}{\sqrt{2}}+\frac{\beta_2-\beta_3}{2}] & -\alpha_3+j\beta_4 & -\frac{\alpha_2+\alpha_4}{\sqrt{2}}+j[\frac{\beta_1}{\sqrt{2}}-\frac{\beta_2-\beta_3}{2}] \end{bmatrix}. \tag{36}$$

This code has a mutual information of 6.25 bits/channel use at $\rho = 20$ dB, which is most of the channel capacity. The matrix S has some interesting features: first, it has orthogonal (but not orthonormal) columns; second, its corresponding $\mathcal{H}\mathcal{H}^t$ matrix is nonzero in only 12 of its 56 off-diagonal entries.

Figure 6 compares the performance of the orthogonal design (35) with the LD code (36) at rate $R = 6$. (From (32), the rate of either code is $(Q/T)\log_2 r$; we achieve $R = 6$ by having the orthogonal design send 256-QAM, and the LD code send 64-QAM.) The decoding in both cases is maximum likelihood, which in the case of the LD code is accomplished with the sphere decoder. We also compare decoding with nulling/cancelling, which appears to be only slightly worse than maximum likelihood (this is perhaps because the columns of the LD code are orthogonal—see Remark 11 in Section 3.2). We see from Figure 6 that the LD code performs uniformly better.

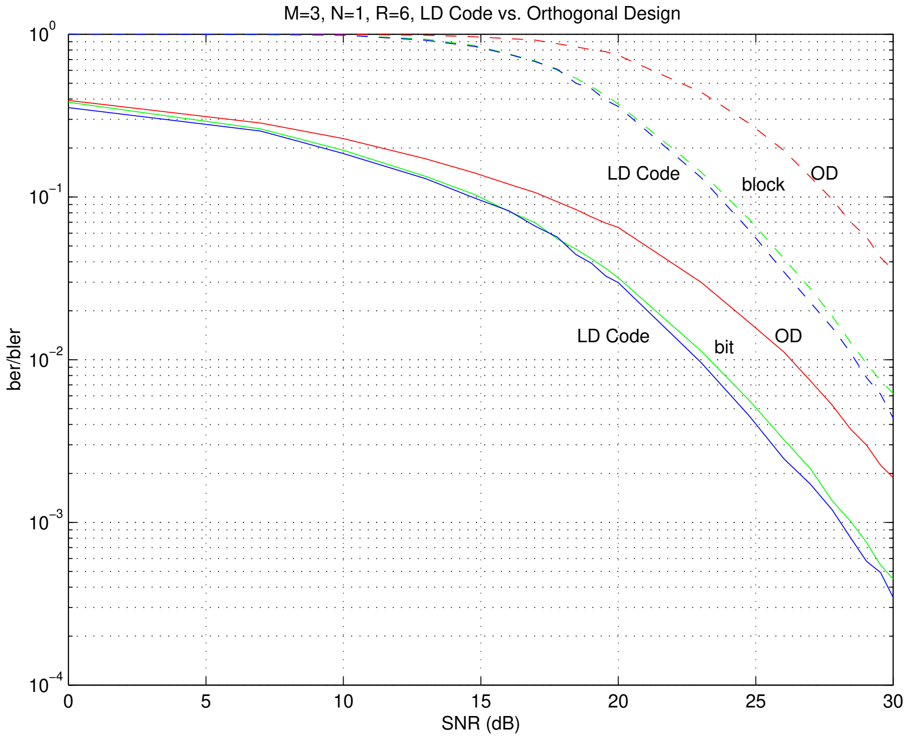

Figure 6: Block (dashed) and bit (solid) error performance of orthogonal design (35) and the LD code (36) for $M = 3$ antennas. The rate is $R = 6$ bits/channel use, obtained in the orthogonal design by transmitting 256-QAM on $s_1, \ldots, s_3$ and obtained in the LD code by transmitting 64-QAM on $s_1, \ldots, s_4$. The uppermost block and bit curves are the orthogonal design, decoded with maximum likelihood. The lower two block and bit curves (very close to one another) are the LD code decoded with nulling/cancelling (upper) and maximum likelihood (lower). The comparison of block error is meaningful here because the block size in all cases is $T = 4$.

$T = 6$, $M = 3$ LD code from 2×2 OD

The $T = 4$, $M = 3$, $N = 1$ LD code (36) is obtained via a gradient search and has mutual information 6.25 bits/channel use at $\rho = 20$ dB. However, this is less than the full $M = 3$, $N = 1$ capacity of 6.41, and we would like to close the gap a little. We should be able to make an LD code with mutual information at least as large as the mutual information of the two-antenna orthogonal design (6), which is 6.28. We do not resume our gradient search since the value 6.25 appears to be a local maximum, but rather try a slightly different approach. We begin with the two-antenna orthogonal design and create a three-antenna LD code that preserves its mutual information.

One possible code is obtained by symmetrically concatenating three 2×2 orthogonal designs (normalized to obey the power constraint)

$$S = \sqrt{\frac{3}{2}} \begin{bmatrix} s_1 & s_2 & 0 \\ -s_2^* & s_1^* & 0 \\ 0 & s_3 & s_4 \\ 0 & -s_4^* & s_3^* \\ s_5 & 0 & s_6 \\ -s_6^* & 0 & s_5^* \end{bmatrix}. \tag{37}$$

When viewed as an LD code, (37) has the deficiency that A_q and B_q are only nonzero for two channel uses and not for the full six channel uses. Moreover, A_q and B_q have rank two, rather than their full possible rank of three. Consequently, constraint (iii) in Section 3.2 is not satisfied, and as we point out in Remark 3, of two codes that have the same mutual information, the one satisfying the stronger constraint generally performs better. It is clear that the code (37) is really only a two-antenna code in crude disguise and performs worse than (36), even though its mutual information is slightly higher.

To improve its performance, we seek to modify it so that constraint (iii) is satisfied without changing its mutual information. One possible modification is described in Remark 4 in Section 3.2. (See, in particular, the transformation involving Φ (30).) Let $\mathcal{F}$ denote the 3×3 DFT matrix, and choose $\tilde{\Phi}$ to be the 6×6 real orthogonal matrix obtained by replacing each element $f_{kl} = g_{kl} + jh_{kl}$ of $\mathcal{F}$ by the 2×2 real matrix $\begin{bmatrix} g_{kl} & -h_{kl} \\ h_{kl} & g_{kl} \end{bmatrix}$. The transformation of $\{A_q, B_q\}$ to new dispersion matrices $\{A'_q, B'_q\}$ is

$$A'_q = \sum_{p=1}^{6} A_p \tilde{\phi}_{p,q}, \quad B'_q = \sum_{p=1}^{6} B_p \tilde{\phi}_{p,q}, \qquad q = 1, \ldots 6. \tag{38}$$

The resulting S matrix is

$$S = \frac{1}{\sqrt{2}} \begin{bmatrix} \sum_{i=1}^{3} s_{2i-1} & \sum_{i=1}^{3} s_{2i} & 0 \\ \sum_{i=1}^{3} -s_{2i}^* & \sum_{i=1}^{3} s_{2i-1}^* & 0 \\ 0 & s_1 - \frac{s_3}{2} + \frac{\sqrt{3}s_4}{2} - \frac{s_5}{2} - \frac{\sqrt{3}s_6}{2} & s_2 - \frac{\sqrt{3}s_3}{2} - \frac{s_4}{2} + \frac{\sqrt{3}s_5}{2} - \frac{s_6}{2} \\ 0 & -s_2^* + \frac{\sqrt{3}s_3^*}{2} + \frac{s_4^*}{2} - \frac{\sqrt{3}s_5^*}{2} + \frac{s_6^*}{2} & s_1^* - \frac{s_3^*}{2} + \frac{\sqrt{3}s_4^*}{2} - \frac{s_5^*}{2} - \frac{\sqrt{3}s_6^*}{2} \\ s_1 - \frac{s_3}{2} - \frac{\sqrt{3}s_4}{2} - \frac{s_5}{2} + \frac{\sqrt{3}s_6}{2} & 0 & s_2 + \frac{\sqrt{3}s_3}{2} - \frac{s_4}{2} - \frac{\sqrt{3}s_5}{2} - \frac{s_6}{2} \\ -s_2^* - \frac{\sqrt{3}s_3^*}{2} + \frac{s_4^*}{2} + \frac{\sqrt{3}s_5^*}{2} + \frac{s_6^*}{2} & 0 & s_1^* - \frac{s_3^*}{2} - \frac{\sqrt{3}s_4^*}{2} - \frac{s_5^*}{2} + \frac{\sqrt{3}s_6^*}{2} \end{bmatrix}. \tag{39}$$

Each dispersion matrix spans all six channel uses and is unitary ($A_q^{'*} A'_q = B_q^{'*} B'_q = I_3$). Thus, constraint (iii) is satisfied. Because the transformation (38) is a special case of the transformation (30), the mutual information is still 6.28 bits/channel use ($\rho = 20$ dB).

We see in Figure 7 that this code performs very well: displayed is (37) (which has the same performance as the 2×2 orthogonal design) and the LD codes (36) and (39) for rate $R = 2$. (The symbol constellation is hence QPSK.) The code (37) has the worst performance. The LD code (36) with $T = Q = 4$ has better performance, despite its lower mutual information, because it satisfies constraint (iii). The best performer, however, is (39), because its mutual information is higher than (36) (6.28 vs. 6.25), it satisfies constraint (iii), and perhaps also because it has a longer block length ($T = 6$ vs. $T = 4$).

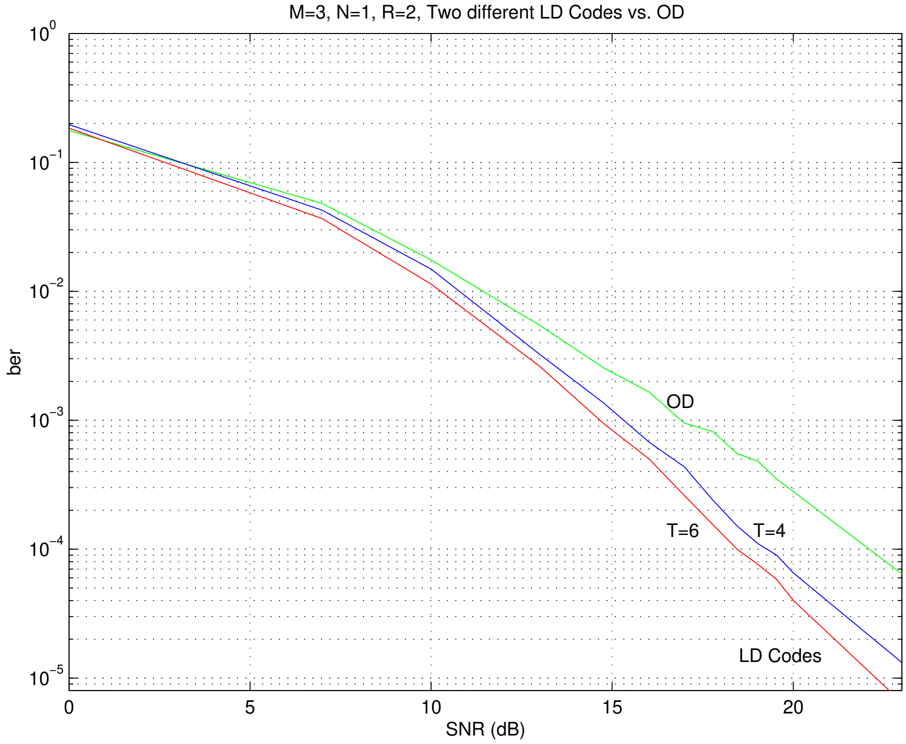

Figure 7: Bit error performance for $M = 3$, $N = 1$ and $R = 2$. The top curve is (37), whose performance is identical to a 2×2 orthogonal design. The middle curve is the $T = 4$ LD code (36), and the lower curve is the $T = 6$ LD code (39). In all cases the transmitted symbols are QPSK, decoded via maximum likelihood.

Two LD Codes: $M = 2, 4$**,** $N = 2$**,** $R = 8$

Figure 8 demonstrates the dramatic improvement of increasing the number of transmitter antennas from $M = 2$ to $M = 4$ with $N = 2$. An LD code was designed for $T = 6$, $M = 4$ and $N = 2$ that attains 11.84 bits/channel use at SNR $\rho = 20$ dB, whereas the channel capacity is 12.49 bits/channel use. We do not explicitly present the code because $Q = 12$ and there are therefore twenty-four 6×4 A and B matrices. (The reader may obtain the code by contacting the authors.) We compare this code with the best LD code we have for $M = 2$ transmit antennas (equation (31) modified with (34) where $\phi = e^{0.488i}$).

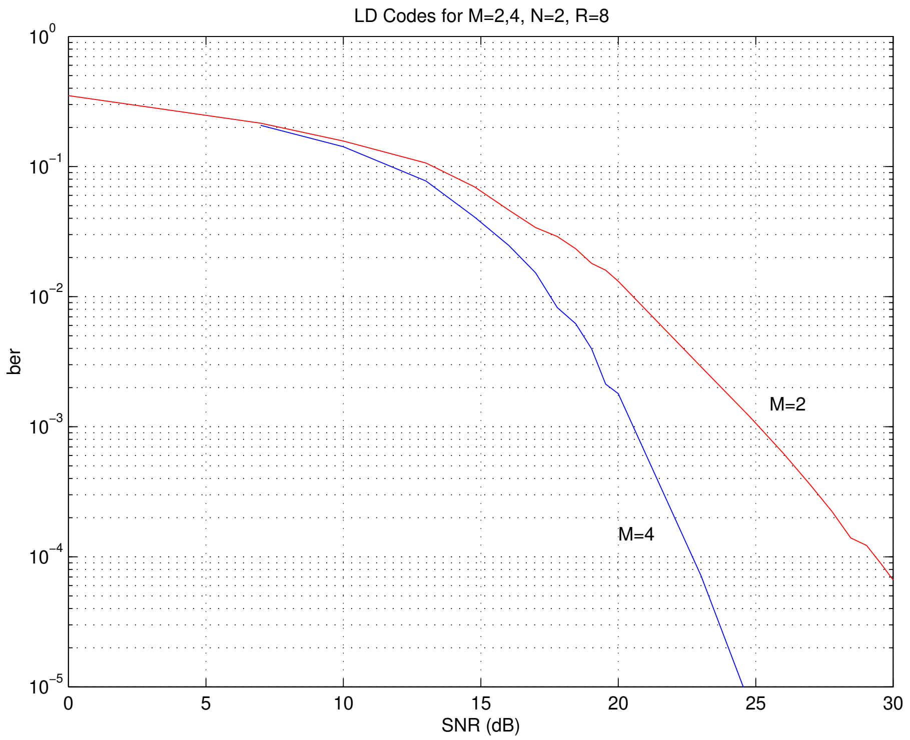

Figure 8: Bit error performance of the best LD code for $T = M = N = 2$ and $Q = 4$ (equation (31) modified with (34)) and an LD code for $T = 6$, $M = 4$, and $N = 2$. The rate is $R = 8$ and is obtained by transmitting 16-QAM on each symbol. The decoding in both cases is maximum likelihood. The LD code for $M = 4$ achieves 11.84 bits/channel use mutual information at $\rho = 20$ dB versus the channel capacity of 12.49, and benefits dramatically from the two extra transmit antennas.

LD Code: $M = 8$**,** $N = 4$**,** $R = 16$

The last example is an LD code for $M = 8$ and $N = 4$ at rate $R = 16$ bits/channel use displayed in Figure 9. It is worth noting that the capacity of an $M = 8$, $N = 4$ system at 20 dB is 24.94 bits/channel use. We therefore restrict our attention in this figure to relatively high SNR. The LD code was designed using gradient search applied to (28) until a local maximum was obtained at $\rho = 20$ dB. The code attains a mutual information of 23.10 bits/channel use, with $T = 8$ and has $Q = 32$. To obtain $R = 16$, we choose $s_1, \ldots, s_{32}$ from a 16-QAM constellation. Because of the sheer number of matrices involved, we again do not explicitly present the LD code here. We include this example to demonstrate that very high rates are well within the reach of these codes, even with maximum likelihood decoding. The figure compares the performance of nulling/cancelling versus maximum likelihood decoding with the sphere decoder, and maximum likelihood performs far better. It is remarkable that the sphere decoder succeeds at all in obtaining the maximum likelihood estimate, since a full exhaustive search would need to examine $2^{RT} = 2^{128} \approx 3.4 \times 10^{38}$ hypotheses.

4.1 Table of Mutual Informations

Table 1 summarizes the mutual informations of some LD codes that we generated, including the examples from the previous section, and the actual channel capacities at $\rho = 20$ dB. As can be observed, $C_{\mathrm{LD}}(\rho, T, M, N)$ is very close to $C(\rho, M, N)$; there is little penalty in the linear structure of the dispersion codes. When studying this table, we should bear in mind that the entries for $C_{LD}(\rho, T, M, N)$ are not necessarily the best achievable since (28) was maximized via gradient ascent. Our maxima are therefore quite possibly local. Furthermore, the values of $C_{LD}(\rho, T, M, N)$ are for codes with block lengths obeying $M \leq T \leq 2M$. Conceivably, increasing T could also yield higher values for $C_{LD}(\rho, T, M, N)$.

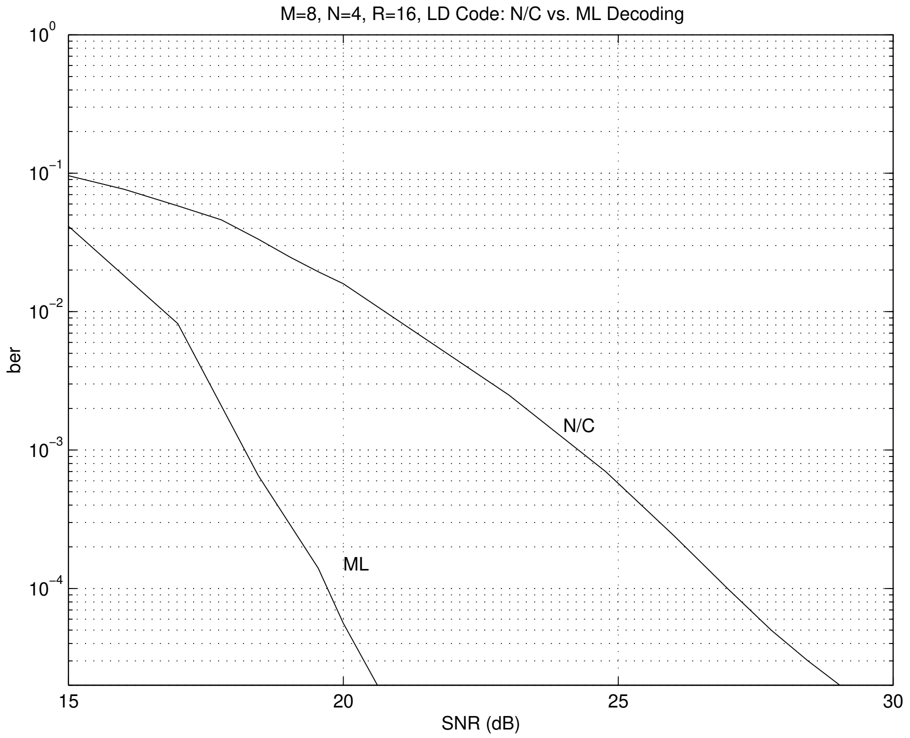

Figure 9: Bit error performance of an LD code for $M = 8$ and $N = 4$ with $Q = 32$ with nulling/cancelling (upper curve) and maximum likelihood decoding (lower curve). The rate is $R = 16$ and is obtained by transmitting 16-QAM on each symbol. The LD code achieves mutual information 23.10 bits/channel use at $\rho = 20$ dB versus the channel capacity of 24.94.

(T,M,N)	$C_{LD}(\rho, T, M, N)$	$C(\rho, M, N)$	Reference
(2,2,1)	6.28	6.28	Eq. (6)
(4,3,1)	6.25	6.41	Eq. (36), Fig. 6
(6,3,1)	6.28	6.41	Eq. (39), Fig. 7
(4,3,2)	11.63	12.14	
(4,4,1)	6.34	6.47	
(6,4,2)	11.84	12.49	Fig. 8
(8,8,4)	23.10	24.94	Fig. 9
$(T = M, M, N \geq M)$	$C(\rho, M, N)$	*	Eq. (31), Figs. 4 & 5

Table 1: Mutual information $C_{LD}(\rho, T, M, N)$ obtained via gradient-ascent optimization of the cost function (28), compared to the actual channel capacity $C(\rho, M, N)$ for different values of M and N at SNR $\rho = 20$ dB.

5 Conclusion

The linear dispersion codes we have introduced are simple to encode and decode, apply to any combination of transmit and receive antennas, subsume as special cases many earlier space-time transmission schemes, and satisfy an information-theoretic optimality property. We have argued that codes that are deficient in mutual information can never be used to attain capacity. We also have shown that information-theoretic optimality has a theoretical connection with low pairwise error probability and good performance at high spectral efficiencies. The LD codes are designed to be linear while having little (if any) penalty in mutual information, and additional channel coding across $s_1, \ldots, s_Q$ can be combined with an LD code to attain most (if not all) of the channel capacity.

We have given some specific examples of the LD codes, and presented a recipe for generating more codes within this linear structure for any combination of transmit and receive antennas. Our simulations indicate that codes generated with this recipe compare favorably with existing space-time schemes in their good performance and low complexity. We have argued that the diversity criterion commonly used to design space-time codes plays a secondary role to mutual information criterion at high spectral efficiencies. The diversity criterion alone may lead to code designs that cannot attain capacity.

In our simulations, we decoded the LD codes by either maximum likelihood or by nulling and cancelling. Because of the linear relation between s and x in (25), the maximum likelihood search could be accomplished efficiently using the sphere decoding algorithm, which, for SNR's of interest, has polynomial complexity in the number of antennas.

We have used the average capacity (across different channel realizations) as our design metric, rather than an outage capacity (for one channel realization), because channels are rarely static for very long and because the outage capacity does not have a simple closed-form expression. It is reasonable to expect that codes that work well on average should also help reduce the probability of outage events, but this remains to be explored.

It would be interesting to see if the LD codes that satisfy (28) possess any general algebraic structure. This would lead to better theoretical understanding of the codes, as well as to possibly faster and better decoding. The codes (36) and (39), for example, are local maxima of the cost function and yet have simple structure. It would also be important to characterize theoretically how close $C_{\rm LD}(\rho, T, M, N)$ can be made to $C(\rho, M, N)$—what is the penalty incurred by the LD structure? Our experience suggests that the penalty is small. Perhaps the penalty approaches zero as $T \to \infty$?

Finally, there are potentially many ways to optimize the cost function (28), and the gradient method we chose is only one of them. More sophisticated optimization techniques may also be useful.

A The Gradient Computation

In all the simulations presented in this paper the maximization of the cost function in (28), needed to design the LD codes, is performed using a simple constrained-gradient-ascent method. In this section, we compute the gradient of the cost function in (28) that this method requires. More sophisticated optimization techniques that we do not consider, such as Newton-Raphson, scoring, and interior-point methods, can also use this gradient.

To help compute this gradient, we rewrite the cost function in (28) as

$$\frac{1}{2T}\mathrm{E}\ \log\det\left(I_{2NT}+\frac{\rho}{M}\sum_{q=1}^{Q}\begin{bmatrix}\mathcal{A}_q & \dots & 0\\ \vdots & \ddots & \vdots\\ 0 & \dots & \mathcal{A}_q\end{bmatrix}\begin{bmatrix}\underline{h}_1\\ \vdots\\ \underline{h}_N\end{bmatrix}\begin{bmatrix}\underline{h}_1\\ \vdots\\ \underline{h}_N\end{bmatrix}^t\begin{bmatrix}\mathcal{A}_q & \dots & 0\\ \vdots & \ddots & \vdots\\ 0 & \dots & \mathcal{A}_q\end{bmatrix}^t+(\mathcal{B}_q\leftarrow\mathcal{A}_q)\right) \tag{A.1}$$

where $\mathcal{A}_q$, $\mathcal{B}_q$ and $\underline{h}_n$ are defined in (22) for $q=1,\dots,Q$ and $n=1,\dots,N$.

We wish to compute the gradient of the cost function in (A.1) with respect to the spreading matrices $A_{R,q}$, $A_{I,q}$, $B_{R,q}$, and $B_{I,q}$. To simplify the gradient calculation, we assume the logarithm in (A.1) is base e. The gradient with respect to $A_{R,q}$ is computed here explicitly—the remaining three gradients follow similar arguments and are given at the end of the section.

The (i,j)th entry of the gradient of a function $f(A_{R,q})$ is

$$\left[\frac{\partial f(A_{R,q})}{\partial A_{R,q}}\right]_{ij}=\lim_{\delta\to 0}\frac{f(A_{R,q}+\delta\zeta_i\eta_j^t)-f(A_{R,q})}{\delta}, \tag{A.2}$$

where ζ_i and η_j are the T-dimensional and M-dimensional unit column vectors with one in the ith and jth entries, respectively, and zeros elsewhere. Defining the matrix appearing in the $\log\det(\cdot)$ of (A.1) as

$Z(A_{R,q})$, so that $f(A_{R,q}) = \frac{1}{2T}\mathrm{E}\,\log\det Z(A_{R,q})$; we then obtain

$$Z(A_{R,q} + \delta\zeta_i\eta_j^t) = Z(A_{R,q}) +$$

$$\frac{\delta\rho}{M}\begin{bmatrix} \begin{bmatrix} \zeta_i\eta_j^t & 0 \\ 0 & \zeta_i\eta_j^t \end{bmatrix} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \begin{bmatrix} \zeta_i\eta_j^t & 0 \\ 0 & \zeta_i\eta_j^t \end{bmatrix} \end{bmatrix} \begin{bmatrix} \underline{h}_1 \\ \vdots \\ \underline{h}_N \end{bmatrix} \begin{bmatrix} \underline{h}_1 \\ \vdots \\ \underline{h}_N \end{bmatrix}^t \begin{bmatrix} \mathcal{A}_q & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathcal{A}_q \end{bmatrix}^t \tag{A.3}$$

$$+\text{transpose} + \text{higher order terms}$$

Applying $\log\det(\cdot) = \mathrm{tr}\ \log(\cdot)$ yields

$$f(A_{R,q} + \delta\zeta_i\eta_j^t) - f(A_{R,q}) = \frac{1}{2T}\mathrm{E}\,\mathrm{tr}\ \log\Bigg(I_{2NT} +$$

$$\frac{\delta\rho}{M}\begin{bmatrix} \begin{bmatrix} \zeta_i\eta_j^t & 0 \\ 0 & \zeta_i\eta_j^t \end{bmatrix} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \begin{bmatrix} \zeta_i\eta_j^t & 0 \\ 0 & \zeta_i\eta_j^t \end{bmatrix} \end{bmatrix} \begin{bmatrix} \underline{h}_1 \\ \vdots \\ \underline{h}_N \end{bmatrix} \begin{bmatrix} \underline{h}_1 \\ \vdots \\ \underline{h}_N \end{bmatrix}^t \begin{bmatrix} \mathcal{A}_q & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathcal{A}_q \end{bmatrix}^t Z^{-1}$$

$$+\text{transpose} + \text{higher order terms}\Bigg) \tag{A.4}$$

$$= \frac{1}{2T}\mathrm{E}\,\mathrm{tr}\,\frac{\delta\rho}{M}\begin{bmatrix} \begin{bmatrix} \zeta_i\eta_j^t & 0 \\ 0 & \zeta_i\eta_j^t \end{bmatrix} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \begin{bmatrix} \zeta_i\eta_j^t & 0 \\ 0 & \zeta_i\eta_j^t \end{bmatrix} \end{bmatrix} \begin{bmatrix} \underline{h}_1 \\ \vdots \\ \underline{h}_N \end{bmatrix} \begin{bmatrix} \underline{h}_1 \\ \vdots \\ \underline{h}_N \end{bmatrix}^t \begin{bmatrix} \mathcal{A}_q & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathcal{A}_q \end{bmatrix}^t Z^{-1}$$

$$+\text{transpose} + \text{higher order terms}$$

where in the last step we use $\log(I + A) = A - A^2/2 + \ldots$. This now leads to

$$\lim_{\delta\to 0} \frac{f(A_{R,q}+\delta\zeta_i\eta_j^t)-f(A_{R,q})}{\delta} =$$

$$\frac{2\rho}{TM}\mathrm{E}\,\mathrm{tr}\begin{bmatrix} \begin{bmatrix} \zeta_i\eta_j^t & 0 \\ 0 & \zeta_i\eta_j^t \end{bmatrix} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \begin{bmatrix} \zeta_i\eta_j^t & 0 \\ 0 & \zeta_i\eta_j^t \end{bmatrix} \end{bmatrix} \begin{bmatrix} \underline{h}_1 \\ \vdots \\ \underline{h}_N \end{bmatrix} \begin{bmatrix} \underline{h}_1 \\ \vdots \\ \underline{h}_N \end{bmatrix}^t \begin{bmatrix} \mathcal{A}_q & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \mathcal{A}_q \end{bmatrix}^t Z^{-1}$$

$$= \frac{2\rho}{TM}\mathrm{E}\,\mathrm{tr}\begin{bmatrix} \begin{bmatrix} \eta_j^t & 0 \\ 0 & \eta_j^t \end{bmatrix} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \begin{bmatrix} \eta_j^t & 0 \\ 0 & \eta_j^t \end{bmatrix} \end{bmatrix} \begin{bmatrix} \underline{h}_1 \\ \vdots \\ \underline{h}_N \end{bmatrix} \begin{bmatrix} \underline{h}_1 \\ \vdots \\ \underline{h}_N \end{bmatrix}^t \begin{bmatrix} \mathcal{A}_q & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \mathcal{A}_q \end{bmatrix}^t Z^{-1} \tag{A.5}$$

$$\times \begin{bmatrix} \begin{bmatrix} \zeta_i & 0 \\ 0 & \zeta_i \end{bmatrix} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \begin{bmatrix} \zeta_i & 0 \\ 0 & \zeta_i \end{bmatrix} \end{bmatrix}$$

$$= \frac{2\rho}{TM}\left(P_{q,i,j} + P_{q,i+T,j+M} + \ldots + P_{q,i+(2N-1)T,j+(2N-1)M}\right)$$

where we have defined the $2NT \times 2NM$ matrix P_q as

$$P_q = \mathrm{E}\left(Z^{-1} \begin{bmatrix} \underline{h}_1 \\ \vdots \\ \underline{h}_N \end{bmatrix} \begin{bmatrix} \underline{h}_1^t & \cdots & \underline{h}_N^t \end{bmatrix} \begin{bmatrix} \mathcal{A}_q & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathcal{A}_q \end{bmatrix} \right). \tag{A.6}$$

This concludes the gradient calculation with respect to $A_{R,q}$.

Similar expressions can be found for the gradients with respect to $A_{I,q}$, $B_{R,q}$, and $B_{I,q}$. With

$$R_q = \mathrm{E}\left(Z^{-1} \begin{bmatrix} \underline{h}_1 \\ \vdots \\ \underline{h}_N \end{bmatrix} \begin{bmatrix} \underline{h}_1^t & \cdots & \underline{h}_N^t \end{bmatrix} \begin{bmatrix} \mathcal{B}_q & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathcal{B}_q \end{bmatrix} \right), \tag{A.7}$$

these gradients are

$$\left[\frac{\partial f(A_{R,q})}{\partial A_{R,q}}\right]_{ij} = \frac{2\rho}{TM}\sum_{n=1}^{N}\left(P_{q,i+(2n-2)T,j+(2n-2)M} + P_{q,i+(2n-1)T,j+(2n-1)M}\right) \tag{A.8}$$

$$\left[\frac{\partial f(A_{I,q})}{\partial A_{I,q}}\right]_{ij} = \frac{2\rho}{TM}\sum_{n=1}^{N}\left(P_{q,i+(2n-1)T,j+(2n-2)M} - P_{q,i+(2n-2)T,j+(2n-1)M}\right) \tag{A.9}$$

$$\left[\frac{\partial f(B_{R,q})}{\partial B_{R,q}}\right]_{ij} = \frac{2\rho}{TM}\sum_{n=1}^{N}\left(R_{q,i+(2n-1)T,j+(2n-2)M} - R_{q,i+(2n-2)T,j+(2n-1)M}\right) \tag{A.10}$$

$$\left[\frac{\partial f(B_{I,q})}{\partial B_{I,q}}\right]_{ij} = -\frac{2\rho}{TM}\sum_{n=1}^{N}\left(R_{q,i+(2n-2)T,j+(2n-2)M} + R_{q,i+(2n-1)T,j+(2n-1)M}\right). \tag{A.11}$$

B Average Pairwise Probability of Error

The signal model (25) is

$$x = \sqrt{\frac{\rho}{M}}\mathcal{H}s + v,$$

where $\mathcal{H}$ is a given real $2NT \times 2Q$ matrix. We compute the pairwise error by conditioning on $\mathcal{H}$,

$$P_e(\text{pairwise}) = \mathrm{E}\, P_e(\text{pairwise}|\mathcal{H}).$$

We assume that two independent $2Q \times 1$ signal vectors s and s' are chosen with entries from independent zero-mean real Gaussian densities with variance $1/2$. The entries of the additive $2NT \times 1$ noise vector v are also chosen from this density. We want the probability that a maximum-likelihood decoder mistakes s for s', given that s is transmitted, averaged over s and s', and conditioned on $\mathcal{H}$.

Because x has a Gaussian distribution, we equivalently want

$$\begin{aligned}
P_e(\text{pairwise}|\mathcal{H}) &= \mathrm{P}\{\|x - \sqrt{\rho/M}\mathcal{H}s'\| < \|x - \sqrt{\rho/M}\mathcal{H}s\| \mid s \text{ transmitted}\} \\
&= \mathrm{P}\{(\sqrt{\rho/M}\mathcal{H}(s-s') + v)^t(\sqrt{\rho/M}\mathcal{H}(s-s') + v) < v^t v\} \\
&= \mathrm{P}\{(\sqrt{2\rho/M}\mathcal{H}s + v)^t(\sqrt{2\rho/M}\mathcal{H}s + v) < v^t v\} && \text{(B.12)} \\
&= \mathrm{P}\{(2\rho/M)s^t\mathcal{H}^t\mathcal{H}s + \sqrt{2\rho/M}(s^t\mathcal{H}^t v + v^t\mathcal{H}s) < 0\}. && \text{(B.13)}
\end{aligned}$$

Equation (B.12) follows because s and s' are independent Gaussian vectors and we have replaced the difference of the two vectors by a single vector with twice the variance.

To compute (B.13), we look at the characteristic function of the scalar $y = (2\rho/M)s^t\mathcal{H}^t\mathcal{H}s + \sqrt{2\rho/M}(s^t\mathcal{H}^t v + v^t\mathcal{H}s)$. We can write

$$y = [s^t\ v^t]\underbrace{\begin{bmatrix} (2\rho/M)\mathcal{H}^t\mathcal{H} & \sqrt{2\rho/M}\mathcal{H}^t \\ \sqrt{2\rho/M}\mathcal{H} & 0 \end{bmatrix}}_{R}\begin{bmatrix} s \\ v \end{bmatrix}.$$

The characteristic function of y is

$$\phi(\omega) = \mathrm{E}\, e^{j\omega y} = \int ds\, dv\, \frac{1}{(2\pi)^{Q+NT}(1/2)^{Q+NT}} \exp\left\{ -[s^t v^t](I + j\omega R) \begin{bmatrix} s \\ v \end{bmatrix} \right\}.$$

We use the formula

$$\int dx\, e^{-x^t \Lambda x} = \pi^{n/2} (\det \Lambda)^{-1/2},$$

for positive real Λ ($\Lambda + \Lambda^* > 0$), where x is a real $n \times 1$ vector, to conclude that

$$\phi(\omega) = \det(I + j\omega R)^{-1/2}.$$

This implies that

$$\begin{aligned} P_e(\text{pairwise}|\mathcal{H}) &= \mathrm{P}\{y < 0\} \\ &= \int_{-\infty}^{0} dy\, p(y) \\ &= \frac{1}{2\pi} \int_{-\infty}^{0} dy \int_{-\infty}^{\infty} d\omega\, e^{j\omega y} \det(I + j\omega R)^{-1/2} \end{aligned} \tag{B.14}$$

We wish to switch the order of integration, and we have to make sure that $e^{j\omega y}$ is well-behaved as a function of ω and y. We can ensure its good behavior by shifting the path of integration to $\omega \in (-\infty - j/2, \infty - j/2)$. (The formal argument for why this does not affect the value of the integral is omitted.) We obtain

$$\begin{aligned} P_e(\text{pairwise}|\mathcal{H}) &= \frac{1}{2\pi} \int_{-\infty-j/2}^{\infty-j/2} d\omega \int_{-\infty}^{0} dy\, e^{j\omega y} \det(I + j\omega R)^{-1/2} \\ &= \frac{1}{2\pi} \int_{-\infty-j/2}^{\infty-j/2} d\omega\, \det(I + j\omega R)^{-1/2} \left. \frac{e^{j\omega y}}{j\omega} \right|_{-\infty}^{0} \\ &= \frac{1}{2\pi j} \int_{-\infty-j/2}^{\infty-j/2} d\omega\, \frac{1}{\omega} \det(I + j\omega R)^{-1/2}. \end{aligned} \tag{B.15}$$

We now compute $\det(I + j\omega R)$ by first computing the eigenvalues of R. The eigenvalues of R are the solutions to

$$\det \begin{bmatrix} \lambda I_{2Q} - (2\rho/M)\mathcal{H}^t\mathcal{H} & -\sqrt{2\rho/M}\mathcal{H}^t \\ -\sqrt{2\rho/M}\mathcal{H} & \lambda I_{2NT} \end{bmatrix} = 0$$

as a polynomial in λ. Using a standard determinant identity [25], we can write this equation as

$$\begin{aligned} 0 &= \lambda^{2NT} \det(\lambda I_{2Q} - (2\rho/M)\mathcal{H}^t\mathcal{H} - \frac{2\rho/M}{\lambda}\mathcal{H}^t\mathcal{H}) \\ &= \lambda^{2NT-2Q} \det(\lambda^2 I_{2Q} - 2(\lambda+1)(\rho/M)\mathcal{H}^t\mathcal{H}) \\ &= \lambda^{2NT-2Q} \prod_{q=1}^{2Q} (\lambda^2 - 2(\lambda+1)\sigma_q), \end{aligned}$$

where $\sigma_1, \ldots, \sigma_{2Q}$ are the eigenvalues of $(\rho/M)\mathcal{H}^t\mathcal{H}$. Solving this last equation yields $2NT - 2Q$ zero eigenvalues, with the remaining $4Q$ eigenvalues given by

$$\sigma_q \pm \sqrt{\sigma_q^2 + 2\sigma_q}, \qquad q = 1, \ldots, 2Q.$$

Therefore,

$$\begin{aligned} \det(I + j\omega R) &= \prod_{q=1}^{2Q} \left[1 + j\omega(\sigma_q + \sqrt{\sigma_q^2 + 2\sigma_q})\right] \left[1 + j\omega(\sigma_q - \sqrt{\sigma_q^2 + 2\sigma_q})\right] \\ &= \prod_{q=1}^{2Q} [1 + 2\omega\sigma_q(j + \omega)], \end{aligned}$$

and (B.15) becomes

$$\begin{aligned} P_e(\text{pairwise}|\mathcal{H}) &= \frac{1}{2\pi j} \int_{-\infty - j/2}^{\infty - j/2} d\omega \, \frac{1}{\omega} \prod_{q=1}^{2Q} [1 + 2\omega\sigma_q(j + \omega)]^{-1/2} \\ &= \frac{1}{2\pi j} \int_{-\infty}^{\infty} d\omega \frac{1}{\omega - j/2} \prod_{q=1}^{2Q} [1 + 2\sigma_q(\omega - j/2)(\omega + j/2)]^{-1/2} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2\pi j}\int_{-\infty}^{\infty} d\omega \frac{\omega + j/2}{\omega^2+1/4}\prod_{q=1}^{2Q}[1+2\sigma_q(\omega^2+1/4)]^{-1/2} \\
&= \frac{1}{4\pi}\int_{-\infty}^{\infty} \frac{1}{\omega^2+1/4}\prod_{q=1}^{2Q}[1+2\sigma_q(\omega^2+1/4)]^{-1/2}. \qquad \text{(B.16)}
\end{aligned}
$$

To get an upper bound on the pairwise error probability, we ignore the second appearance of ω^2 in (B.16) to obtain

$$
\begin{aligned}
P_e(\text{pairwise}|\mathcal{H}) &\le \frac{1}{4\pi}\int_{-\infty}^{\infty} d\omega\,\frac{1}{\omega^2+1/4}\prod_{q=1}^{2Q}[1+\sigma_q/2]^{-1/2} \\
&= \frac{1}{2}\prod_{q=1}^{2Q}[1+\sigma_q/2]^{-1/2} \\
&= \frac{1}{2}\det\left[I+\frac{\rho}{2M}\mathcal{H}^t\mathcal{H}\right]^{-1/2}.
\end{aligned}
$$

It follows that

$$
P_e(\text{pairwise}) \le \mathrm{E}\,\frac{1}{2}\det\left[I+\frac{\rho}{2M}\mathcal{H}^t\mathcal{H}\right]^{-1/2}. \qquad \text{(B.17)}
$$

Applying a union bound to this average pairwise probability of error yields an upper bound on probability of error of a signal constellation. Suppose that the transmission rate is R, so that there are 2^{RT} elements in our constellation for S; then

$$
P_e \le 2^{RT}\mathrm{E}\,\frac{1}{2}\det\left[I+\frac{\rho}{2M}\mathcal{H}^t\mathcal{H}\right]^{-1/2} = \frac{1}{2}\mathrm{E}\,e^{-(\frac{1}{2T}\log\det[I+\frac{\rho}{2M}\mathcal{H}^t\mathcal{H}]-R)T}.
$$

References

[1] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[2] I. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *Eur. Trans. Telecom.*, vol. 10, pp. 585–595, Nov. 1999.

[3] A. Wittneben, “Basestation modulation diversity for digital simulcast,” in *Proc. IEEE VTC*, pp. 848–853, 1991.

[4] N. Seshadri and J. Winters, “Two signaling schemes for improving the error performance of frequency-division-duplex (fdd) transmission systems using transmitter antenna diversity,” in *Proc. IEEE VTC*, pp. 508–511, 1993.

[5] A. Wittneben, “A new bandwidth-efficient transmit antenna modulation diversity scheme for linear digital modulation,” in *Proc. IEEE ICC*, pp. 1630–1634, 1993.

[6] J. Winters, “The diversity gain of transmit diversity in wireless systems with Rayleigh fading,” in *Proc. IEEE ICC*, pp. 1121–1125, 1994.

[7] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, “Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture,” *Electronic Letters*, vol. 35, pp. 14–16, Jan. 1999.

[8] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, “Simplified processing for high spectral efficiency wireless communication employing multi-element arrays,” *J. Sel. Area Comm.*, vol. 17, pp. 1841–1852, Nov. 1999.

[9] 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,” in *Proc. IEEE VTC*, pp. 136–140, 1996.

[10] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Info. Theory*, vol. 44, pp. 744–765, 1998.

[11] S. M. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Sel. Area Comm.*, pp. 1451–1458, Oct. 1998.

[12] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Info. Theory.*, vol. 45, pp. 1456–1467, July 1999.

[13] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communication in Rayleigh flat-fading," *IEEE Trans. Info. Theory*, vol. 46, pp. 543–564, Mar. 2000.

[14] B. Hochwald and W. Sweldens, "Differential unitary space time modulation," *IEEE Trans. Comm.*, vol. 48, pp. 2041–2052, Dec. 2000. Download available at `http://mars.bell-labs.com`.

[15] B. Hughes, "Differential space-time modulation," *IEEE Trans. Info. Theory*, pp. 2567–2578, Nov. 2000.

[16] A. Shokrollahi, B. Hassibi, B. Hochwald, and W. Sweldens, "Representation theory for high-rate multiple-antenna code design," *to appear in IEEE Trans. Info. Theory*, 2001. Download available at `http://mars.bell-labs.com`.

[17] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *J. Sel. Area Comm.*, pp. 1169–1174, July 2000.

[18] B. Hassibi, "An efficient square-root algorithm for BLAST," *submitted to IEEE Trans. Sig. Proc.*, 1999. Download available at `http://mars.bell-labs.com`.

[19] M. O. Damen, A. Chkeif, and J.-C. Belfiore, "Lattice code decoder for space-time codes," *IEEE Comm. Let.*, pp. 161–163, May 2000.

[20] B. Hassibi and H. Vikalo, "On the expected complexity of sphere decoding," *In preparation*, 2001.

[21] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *submitted to IEEE Trans. Info. Theory*, 2000. Download available at `http://mars.bell-labs.com`.

[22] B. Hassibi, B. Hochwald, and T. Marzetta, "Space-time autocoding," *submitted to IEEE Trans. Info. Theory*, 1999. Download available at `http://mars.bell-labs.com`.

[23] G. Genesan and P. Stoica, "Space-time diversity using orthogonal and amicable orthogonal designs," in *Proc. ICASSP*, pp. 335–338, 2000.

[24] M. O. Damen, A. Tewfik, and J.-C. Belfiore, "A construction of a space-time code based on the theory of numbers," *submitted to IEEE Trans. Info. Theory*, 2000.

[25] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation.* Englewood Cliffs, NJ: Prentice-Hall, 2000.