UNITARY SPACE-TIME CODES AND THE CAYLEY TRANSFORM

Babak Hassibi and Yindi Jing

Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125

ABSTRACT

A recently proposed method for communicating with multiple antennas over block fading channels is *unitary space-time modulation* (USTM), so-called because the transmitted signals form a matrix with orthonormal columns. Since channel knowledge is not required at the receiver, USTM schemes are suitable for use on wireless links where channel tracking is undesirable or infeasible. Recent results have shown that, if suitably designed, USTM schemes can achieve full channel capacity at high SNR. While all this is well recognized, what is not clear is how to generate good performing constellations of (non-square) unitary matrices, that lend themselves to efficient encoding/decoding. The schemes proposed so far either exhibit poor performace, especially at high rates, or have no efficient decoding algorithms. In this paper, we propose to use the *Cayley transform* to design USTM constellations. This work is a generalization, to the non-square case, of the Cayley codes that have been proposed for *differential* USTM. The codes are designed based on an information-theoretic criterion, and lend themselves to polynomial-time (often cubic) near-maximum-likelihood decoding using a sphere decoding algorithm.

1. INTRODUCTION AND MODEL

It is well known that multiple transmit and/or receive antennas promise high data rates on scattering-rich wireless channels [1, 2]. Most of the proposed schemes that achieve these high rates require the propagation environment or channel to be known to the receiver (see, e.g., [1, 3, 4, 5] and the references therein). However, it is not always feasible to assume that the receiver knows the channel, especially when many antennas are used or either end of the link is moving so fast that the channel is changing very rapidly [6, 7].

Hence, there is much interest in space-time transmission schemes that do not require either the transmitter or receiver to know the channel. Information-theoretic calculations with a multi-antenna channel that changes in a block-fading manner first appeared in [8]. Based on these calculations, a new transmission scheme, referred to as *unitary space-time modulation* (USTM), in which the transmitted signals, viewed as matrices with spatial and tem-

poral dimensions, form a unitary matrix, was proposed in [9]. Further information-theoretic calculations in [10] and [11] show that, at high SNR, USTM schemes are capable of achieving full channel capacity. Furthermore, in [12], it is shown that all this can be done over a single coherence interval, provided the coherence interval and number of transmit antennas are sufficiently large.

While all this is well recognized, it is not clear how to design a constellation of non-square USTM matrices, that deliver on the above information-theoretic results and lend themselves to efficient encoding/decoding. The technique of [13] exhibits poor performance at high rates, whereas the constellation of [14] has no tractable decoding algorithm.

In this paper we propose to use the Cayley transform to design USTM constellations. This is an extension, to the non-square case, of earlier work on Cayley codes for *differential* USTM [15]. Although this extension is non-trivial, the codes designed here inherit many of the properties of Cayley differential codes. In particular, they:

1. Are very simple to encode: the data is broken into substreams used to parametrize the unitary matrix.

2. Apply to any antenna configuration.

3. Can be decoded in a variety of ways including simple polynomial-time linear-algebraic techniques such as successive nulling and cancelling (V-BLAST [16, 17]) or sphere decoding [18, 19].

4. Satisfy a probabilistic criterion: they maximize an expected distance between matrix pairs

1.1. Unitary Space-Time Modulation

We consider a wireless communication system with M transmit antennas and N receive antennas. We use a block-fading channel with coherence interval T [8, 9]:

$$X = \sqrt{\frac{\rho T}{M}} SH + V \qquad (1)$$

Here, $S : T \times M$ is the transmitted signal with S_{tm} the signal sent by the $m th$ transmit antenna at time t. $H : M \times N$ is

the complex-valued propagation matrix which remains constant during the coherent period. h_{mn} is the propogation coefficient between the mth transmit antenna and the nth receive antenna and has a $\mathcal{CN}(0,1)$ distribution independent of all other entries of H. The channel H is unknown to both the transmiter and the receiver. $V : T \times N$ is additive Gaussian noise with independent $\mathcal{CN}(0,1)$ entries. $X : T \times N$ is the received signal matrix with x_{tn} the received value by the nth receive antenna at time t. The transmit power constraint is $\sum_{t=1}^{T} |s_{tm}|^2 = 1$, $m = 1, 2, ..., M$, so ρ represents the expected SNR at each receiver antenna.

Theorem 1 (Structure of Capacity-achieving Signal) *[8]*
A capacity-achieving random signal matrix for (1) may be constructed as a product $S = \Phi V$, where Φ is an isotropically distributed $T \times M$ matrix whose columns are orthonormal, and V is an independent $M \times M$ real, nonnegative, diagonal matrix. Furthermore, for either $T \gg M$, or high SNR and $T > M$, $V = I_M$ achieves capicity.

Motivated by this theorem, [9] proposed to use the transmitted signal matrix S as $S = \Phi \begin{bmatrix} I_M \\ 0 \end{bmatrix}$ with Φ a $T \times T$ unitary matrix. This is called unitary space-time modulation (USTM), and such an S is called a $T \times M$ unitary matrix. In the USTM scheme, the transmitted signals are chosen from a constellation $\mathcal{S} = \{S_1, ..., S_L\}$ of $L = 2^{RT}$ (where R is the transmission rate) $T \times M$ unitary matrices. The ML decoder is given by:

$$\hat{\ell} = \arg \max_{\ell=1,...,L} \|X^* S_\ell\|_F^2 = \arg \min_{\ell=1,...,L} \|X^* S_\ell^\perp\|_F^2, \quad (2)$$

where $S_\ell^\perp$ denotes the orthogonal complement of S_ℓ.

The pairwise probability of error (of transmitting S_ℓ and erroneously decoding $S_{\ell'}$) behaves as $|\det(S_\ell^* S_{\ell'})|^{-2N}$ [9]. Therefore, most design schemes attempt to find a constellation that maximizes $\min_{\ell \neq \ell'} |\det(S_\ell^* S_{\ell'})|$. Since L can be quite large, this calls into question the feasibility of computing and using this performance criterion. The large number of signals also rules out the possibility of decoding via an exhaustive search. To design constellations that are huge, effective, and yet still simple, so that they can be decoded in real-time, we need to introduce some structure.

2. CAYLEY USTM CODES

The space of $T \times M$ complex unitary matrices is referred to as the *Stiefel* manifold and can be parameterized by $2TM - M^2$ real free parameters. We focus on parametrization through the Cayley transform.

2.1. The Cayley transform

The Cayley transform of a complex $T \times T$ matrix Y is $(I + Y)^{-1}(I - Y)$ [20, 21]. With $Y = iA$ is skew-Hermitian,

$$V = (I + iA)^{-1}(I - iA), \quad (3)$$

is $T \times T$ unitary. Thus, the Cayley transform expresses a unitary matrix as a function of a skew-Hermitian matrix, which is described by T^2 real parameters. This parameterization is promising because it is one-to-one: $iA = (I + V)^{-1}(I - V)$. For our application, we are interested in only the first M columns of V, which has only $2TM - M^2$ free real degrees of freedom. Therefore, we will be only interested in $2TM - M^2 = T^2 - (T - M)^2$ degrees of freedom of the total T^2 degrees of freedom in A.

Because the Cayley transform maps the nonlinear Stiefel manifold to the linear space (over the reals) of skew-Hermitian matrices (and vice-versa) it is convenient to encode data onto a skew-Hermitian matrix and then apply the Cayley transform to get a unitary matrix. We call a *Cayley USTM (CUSTM) code* one for which each $T \times M$ unitary matrix is $V = (I + iA)^{-1}(I - iA) \begin{bmatrix} I_M \\ 0 \end{bmatrix}$, where the Hermitian matrix A is given by

$$A = \sum_{q=1}^{Q} \alpha_q A_q, \quad (4)$$

where $\alpha_1, ..., \alpha_Q$ are real scalars (chosen from a set $\mathcal{A}_r$ with r possible values) and where A_q are *fixed* $T \times T$ complex Hermitian matrices. The code is completely determined by the set of Hermitian basis matrices $A_1, ..., A_Q$. Each individual codeword, on the other hand, is determined by our choice of the scalars $\alpha_1, ..., \alpha_Q$. The transmission rate is clearly $R = (Q/T) \log_2 r$.

2.2. Decoding the CUSTM code

If we partition the matrix A as $\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, then some algebra shows that

$$S = \begin{bmatrix} 2\Delta_1^{-1} - I_M \\ -2(I + iA_{22})^{-1}iA_{12}^*\Delta_1^{-1} \end{bmatrix}, S^\perp = \begin{bmatrix} -2(I + iA_{11})^{-1}iA_{12}\Delta_2^{-1} \\ 2\Delta_1^{-1} - I_{T-M} \end{bmatrix},$$

where $\Delta_1 = I + jA_{11} + A_{12}(I + jA_{22})^{-1}A_{12}^*$ and $\Delta_2 = I + jA_{22} + A_{12}^*(I + jA_{11})^{-1}A_{12}$. This implies that the second form of the ML decoder in (2) reduces to

$$\arg \min_{\alpha_q} \left\| \left(-2iX_1^*(I + iA_{11})^{-1}A_{12} + X_2^*(2 - \Delta_2) \right) \Delta_2^{-1} \right\|_F^2,$$

where we have partitioned $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$. This decoder is not quadratic in $\{\alpha_q\}$ and so may be difficult to solve. If we adopt the approach of [15] and ignore the covariance of the additive noise term $\Delta_2^{-1}\Delta_2^{-*}$, we obtain

$$\arg \min_{\alpha_q} \left\| -2iX_1^*(I + iA_{11})^{-1}A_{12} + X_2^*(2 - \Delta_2) \right\|_F^2,$$

which, however, is still not quadratic in the $\{\alpha_q\}$. Fortunately, in our problem, we have excess degrees of freedom in A: T^2, rather than $2TM - M^2 = T^2 - (T - M)^2$.

Therefore we may impose constraints on A to simplify the problem. Indeed, if we let

$$(I + jA_{11})^{-1}A_{12} = B, \qquad (5)$$

for some fixed $M \times (T - M)$ matrix B, some algebra shows that the above decoder reduces to

$$\hat{\alpha}_{\text{lin}} = \arg\min_{\{\alpha_q\}} \|X_2^* - X_2^* B^* B - 2iX_1^* B - iX_2^* A_{22} + iX_2^* B^* A_{11} B\|_F^2, \qquad (6)$$

which is now quadratic in $\{\alpha_q\}$. We call (6) the "linearized" decoder because the system of equations obtained in solving for unconstrained $\{\alpha_q\}$ is linear. For a wide range of rates and SNR (6) can be solved exactly in roughly $O(Q^3)$ computations using sphere decoding [18, 19]. Furthermore, simulation results show that the penalty for solving (6) is small, especially when weighed against the complexity of exact ML.

To keep the complexity of the sphere decoding algorithm polynomial, it is important that the number of linear equations resulting from (6) be at least as large as the number of unknowns. Looking at (6) suggests that we have $2M(T - M)$ real equations and Q real unknowns. However, due to the Hermitian constraints not all $2M(T - M)$ equations are independent. A careful analysis yields the following result.

Theorem 2 (Number of equations) *The number of independent equations obtained from (6) is* $2N(T - M) - N^2$, *when* $T - M \geq N$ *and* $(T - M)^2$ *when* $T - M < N$. *Therefore, we require* $Q \leq 2N(T - M) - N^2$ *when* $T - M \geq N$ *and* $Q \leq (T - M)^2$ *when* $T - M < N$.

2.3. CUSTM code revisited

With the choice (5), we can explicitly write the CUSTM code as

$$S = \begin{bmatrix} I_M + 2B\Delta_2^{-1}B^* \\ -2i\Delta_2^{-1}B^* \end{bmatrix}, \qquad (7)$$

where

$$\Delta_2 = I_{T-M} + B^*B - \sum_{q_1=1}^{Q_1} \alpha_{q_1} B^* A_{11,q_1} B + i\sum_{q_2=1}^{Q_2} \beta_{q_2} A_{22,q_2}, \qquad (8)$$

and $Q_1 + Q_2 = Q$.

2.4. Design of the CUSTM Codes

We introduced the CUSTM structure (7) and showed how to choose Q according to Thm.2. What remains is to design the Hermitian basis matrices $\{A_{11,q_1}\}$ and $\{A_{22,q_2}\}$ and choose the discrete set $\mathcal{A}$ from which the α_q are drawn.

If the rates being considered are reasonably small then the criterion given in [9] of maximizing $|\det(S_\ell^* S_{\ell'})|$ for all $\ell' \neq \ell$ may be tractable. At high rates, however, the criterion

becomes intractable because of the number of matrices involved, and the performance of the constellation may not be governed so much by its worst-case pairwise $|\det(S_\ell^* S_{\ell'})|$, but rather by how well the matrices are distributed over the space of unitary matrices. In fact, from the information-theoretic considerations in [10] and [11] it follows that the optimal distribution is an *isotropically-random* unitary matrix, i.e., one whose probability density function is invariant to pre- and post-multiplication by an arbitrary unitary matrix. In the Cayley transform domain this translates to the following.

Theorem 3 (Optimal distribution on A) *The unitary matrix* $V = (I + iA)^{-1}(I - iA)$ *is isotropically-distributed if and only if the Hermitian matrix A has the matrix Cauchy distribution*

$$p(A) = \frac{2^{M^2-M}(M - 1)! \cdots 1!}{\pi^{M(M+1)/2}} \frac{1}{\det(I + A^2)^M}. \qquad (9)$$

The probability density function (9) is the matrix generalization of the familiar scalar Cauchy distribution, which can be regarded as the random variable $\tan(\theta/2)$, with θ uniform on $[0, 2\pi]$. Theorem 3 implies that, at high rates, our CUSTM constellation $A = \sum_{q=1}^{Q} A_q \alpha_q$ should resemble samples from a Cauchy random matrix distribution. Drawing upon the implications from the case of one transmit antenna $M = 1$, we propose to choose the set $\mathcal{A}_r$ as the r-point discretization of a scalar Cauchy random variable. In other words, we choose $\mathcal{A}_r$ as the image of the function $\tan(\theta/2)$ applied to the set $\theta \in \{\pi/r, 3\pi/r, 5\pi/r, \ldots, (2r-1)\pi/r\}$. Thus, we have $\mathcal{A}_2 = \{-1, 1\}$, $\mathcal{A}_4 = \{-1 - \sqrt{2}, 1 - \sqrt{2}, -1 + \sqrt{2}, 1 + \sqrt{2}\} = \{-2.41, -0.41, 0.41, 2.41\}$, $\mathcal{A}_8 = \{-5.03, -1.50, -0.67, -0.20, 0.20, 0.67, 1.50, 5.03\}$, and so on. Note that the points rapidly spread themselves out as r increases, thus reflecting the long tail of the Cauchy distribution. Note also that for any choice of r the rate of the Cayley code is $R = Q/T \log_2 r$.

Finally, we propose to choose the basis matrices $\{A_{11,q_1}\}$ and $\{A_{22,q_2}\}$, as well as the fixed matrix B, such that the resulting $A = \sum_{q=1}^{Q} \alpha_q A_q$ emulates samples of a Cauchy random matrix, or equivalently, $V = (I + iA)^{-1}(I - iA)$ emulates samples of an isotropically unitary matrix. For this we generalize the approach of [15] and use the criterion

$$\max_{\{A_{11,q_1}\},\{A_{22,q_2}\},B} \mathrm{E} \log\det(S^\perp - S^{\perp'})^*(S^\perp - S^{\perp'}),$$

where the expectation is over $\alpha_1, \ldots, \alpha_{Q_1}$ and $\beta_1, \ldots, \beta_{Q_2}$ chosen independently from a Cauchy distribution. Some algebra simplifies the above optimization problem to the following:

$$\max_{\{A_{11,q_1}\},\{A_{22,q_2}\},B} \mathrm{E} \log|\det(B^* A_{11} B + A_{22})|^2 - 2\mathrm{E} \log|\det \Delta_2|^2. \qquad (10)$$

This optimization may be performed numerically using gradient-ascent methods along with Monte Carlo simulation.

3. SIMULATION RESULTS

We present a simple $T = 4$, $M = N = 2$, $R = 2$ example using a randomly-chosen (though suitably normalized) CUSTM code. The resulting BER and BLER curves are presented in Fig. 1. The main point is that there is little penalty (only around 1db) incurred when using the linearized ML cost (6), compared to the true ML. This penalty is certainly well worth the computational savings over an exhaustive ML search.
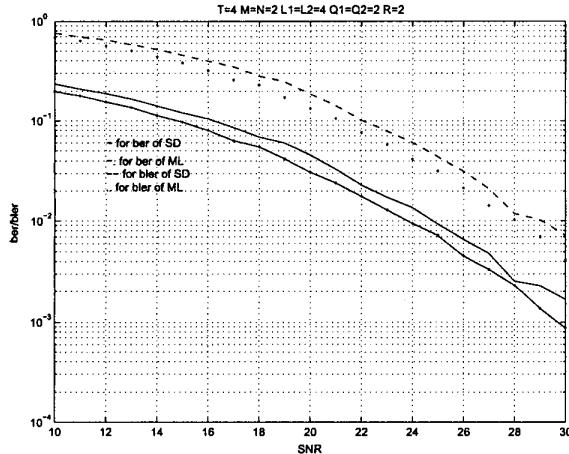

Fig. 1. $T = 4$, $M = N = 2$, $R = 2$: BER and BLER of linearized ML (6) compared to the true ML.

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

[4] 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.

[5] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," *submitted to IEEE Trans. Info. Theory*, 2000. Download available at http://mars.bell-labs.com.

[6] T. L. Marzetta, "Blast training: Estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annual Allerton Conference on Communications, Control, and Computing*, Sept. 22–24 1999.

[7] 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.

[8] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Info. Theory*, vol. 45, pp. 139–157, 1999.

[9] 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.

[10] L. Zheng and D. Tse, "Packing spheres in the Grassman manifold: a geometric approach to the noncoherent multi-antenna channel," *submitted to IEEE Trans. Info. Theory*, 2000.

[11] B. Hassibi and T. L. Marzetta, "Block-fading channels and isotropically-random unitary inputs: The received signal density in closed-form," *Submitted to IEEE Trans. on Info. Thy.*, 2001.

[12] B. Hochwald, T. Marzetta, and B. Hassibi, "Space-time autocoding," *to appear in the IEEE Trans. Info. Theory*, 2001. Download available at http://mars.bell-labs.com.

[13] B. Hochwald, T. Marzetta, T. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Info. Theory*, pp. 1962–1973, 2000.

[14] T. Marzetta, B. Hassibi, and B. Hochwald, "Structured unitary space-time constellations," *to appear in the IEEE Trans. Info. Theory*, 2002. Download available at http://mars.bell-labs.com.

[15] B. Hassibi and B. Hochwald, "Cayley differential unitary space-time codes," *to appear in the IEEE Trans. Info. Theory*, 2001. Download available at http://mars.bell-labs.com.

[16] 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.

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

[18] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Mathematics of Computation*, vol. 44, pp. 463–471, April 1985.

[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] R. Horn and C. Johnson, *Topics in Matrix Analysis*. Cambridge: University Press, 1991.

[21] L. Mirsky, *An Introduction to Linear Algebra*. Oxford: Clarendon Press, 1955.