MULTI-ANTENNA CAYLEY DIFFERENTIAL CODES

Babak Hassibi

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

Bertrand M. Hochwald

Mathematics of Communications Research
Bell Laboratories, Lucent Technologies
Murray Hill, NJ 07974

ABSTRACT

Multiple antenna differential modulation using unitary matrices requires no channel knowledge at the receiver, and so is ideal for use on wireless links where channel tracking is undesirable or infeasible, either because of rapid changes in the channel characteristics or because of limited system resources. Although this basic principle is well understood, it is not known how to generate good-performing constellations of unitary matrices, for any number of transmit and receive antennas and especially at high rates. We propose a class of *Cayley codes* that works with any number of antennas, and allows for polynomial-time near-maximum-likelihood decoding based on either successive nulling/cancelling or sphere decoding. The codes use the Cayley transform, which maps the highly nonlinear Stiefel manifold of unitary matrices to the linear space of skew-Hermitian matrices. This leads to a simple linear constellation structure in the Cayley transform domain and to an information-theoretic design criterion based on emulating a Cauchy random matrix. Simulations show that Cayley codes allow efficient and effective high-rate data transmission in multi-antenna communication systems without knowing the channel.

1. INTRODUCTION AND MODEL

Multiple transmit and/or receive antennas promise high data rates on scattering-rich wireless channels [1, 2]. Many 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 channel is known to the receiver, 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. A standard method that avoids this in single-antenna wireless channels is differential phase-shift keying (DPSK) [8]. Recently, these differential techniques have been generalized to the multi-antenna setting [9, 10, 11], where the transmitted signals form unitary matrices. Although this basic principle is well understood, it is not known how to generate good-performing constellations of unitary matrices that lend themselves to efficient encoding and decoding, for any number of transmit and receive antennas and especially at high rates. The main difficulty stems from the highly non-linear search space of unitary matrices. A design method which uses elements of group theory has appeared in [12].

In this paper we propose a family of Cayley codes for differential unitary space-time modulation. The Cayley codes break the data stream into substreams which are used to parameterize the unitary matrices that are transmitted. The codes work with any antenna configuration and:

1. Are very simple to encode
2. Can be used for any number of transmit and receive antennas
3. Can be decoded in a variety of ways including simple polynomial-time linear-algebraic techniques such as successive nulling and cancelling (V-BLAST [13, 14]) or sphere decoding [15, 16].
4. Satisfy a probabilistic criterion: they maximize an expected distance between matrix pairs

1.1. Differential Unitary Space-Time Modulation

A narrow-band, flat-fading, multi-antenna communication system with M transmit and N receive antennas, which is (roughly) constant for M channel uses, can be written as

$$X_\tau = \sqrt{\rho} S_\tau H + W_\tau, \tag{1}$$

where W_τ and H are $M \times N$ matrices of independent $\mathcal{CN}(0,1)$ random variables, unknown to the receiver, and X_τ is the $M \times N$ received complex signal matrix. In differential unitary space-time modulation [9, 10] the transmitted matrix, S_τ, at block τ satisfies the following so-called fundamental transmission equation $S_\tau = V_{z_\tau} S_{\tau-1}$, where $z_\tau \in \{0, \ldots, L-1\}$ is the data to be transmitted. Since the channel is used M times, the corresponding transmission rate is

$R = (1/M)\log_2 L$. If we further assume that the propagation environment is approximately constant for $2M$ consecutive channel uses, then we may write

$$X_\tau = \sqrt{\rho} S_\tau H + W_\tau = \sqrt{\rho} V_{z_\tau} S_{\tau-1} H + W_\tau = V_{z_\tau} X_{\tau-1} + \underbrace{W_\tau - V_{z_\tau} W_{\tau-1}}_{W'_\tau}.$$

Note that H does not appear in the above equation. From $S_\tau = V_{z_\tau} S_{\tau-1}$ it is apparent that the matrices V_ℓ should be unitary, otherwise the product $S_\tau = V_{z_\tau} V_{z_{\tau-1}} \cdots V_{z_1}$ can go to zero or diverge. Moreover, when V_{z_τ} is unitary the additive noise term W'_τ is statistically independent of V_{z_τ} and so the maximum-likelihood (ML) decoder of z_τ is

$$\hat{z}_\tau = \arg \max_{\ell=0,\ldots,L-1} \|X_\tau - V_\ell X_{\tau-1}\|. \tag{2}$$

In [9, 10] it is shown that the pairwise block probability of error (of transmitting V_ℓ and erroneously decoding $V_{\ell'}$) behaves as $|\det(V_\ell - V_{\ell'})|^{-2N}$. Therefore, most design schemes [9, 10, 12, 17] have focused on finding a constellation $\mathcal{V} = \{V_0, \ldots, V_{L-1}\}$ of $L = 2^{MR}$ unitary $M \times M$ matrices that maximizes $\min_{\ell \neq \ell'} |\det(V_\ell - V_{\ell'})|$. In general, L can be quite large, which 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 must introduce some structure.

2. CAYLEY DIFFERENTIAL CODES

The space of $M \times M$ complex unitary matrices is referred to as the *Stiefel* manifold and can be parameterized by M^2 real free parameters. Possible parametrizations can be given in terms of products of Givens rotations, Householder reflections [18], or through the exponential map [19]. We focus on parametrization through the Cayley transform.

2.1. The Cayley transform

The Cayley transform of a complex $M \times M$ matrix Y is $(I+Y)^{-1}(I-Y)$ [18, 20]. With $Y = iA$ skew-Hermitian,

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

is unitary. Thus, the Cayley transform expresses a unitary matrix as a function of a skew-Hermitian matrix, which is described by M^2 real parameters. This parameterization is promising since it is one-to-one: $iA = (I+V)^{-1}(I-V)$.

2.2. Cayley codes

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. It is most straightforward to encode the data linearly. We call a *Cayley Differential (CD) code* one for which each unitary matrix is

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

where the Hermitian matrix A is given by

$$A = \sum_{q=1}^{Q} \alpha_q A_q, \tag{4}$$

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

2.3. Decoding the CD codes

An important property of the CD codes is the ease with which the receiver may form a system of linear equations in the variables $\{\alpha_q\}$. To see this, we write the ML decoder (2) using the Cayley transform,

$$\hat{\alpha}_{\rm ml} = \arg \min_{\{\alpha_q\}} \left\| (I + iA)^{-1} \left(X_\tau - X_{\tau-1} - \frac{1}{i} A (X_\tau + X_{\tau-1}) \right) \right\|^2.$$

This decoder is not quadratic in $\{\alpha_q\}$ and so may be difficult to solve. However, if we ignore the covariance of the additive noise term $(I + A^2)^{-1}$, then we obtain the "linearized ML" decoder

$$\hat{\alpha}_{\rm lin} = \arg \min_{\{\alpha_q\}} \left\| X_\tau - X_{\tau-1} - \frac{1}{i} \sum_{q=1}^{Q} \alpha_q A_q (X_\tau + X_{\tau-1}) \right\|^2, \tag{5}$$

so called because the system of equations obtained in solving (5) for unconstrained $\{\alpha_q\}$ is linear. This implies that a simple approximate solution for $\{\alpha_q\}$ can use nulling and cancelling. An exact solution can use sphere decoding [15, 16]. For reasonable rates and SNR, both methods require $O(Q^3)$ computations. Simulation results show that the penalty for solving (5) is small, especially when weighed against the complexity of exact ML.

Nulling and cancelling explicitly requires that the number of equations be at least as large as the number of unknowns and sphere decoding benefits from this, since it reduces the computations. Looking at the linear equation obtained from (5) suggests that we have $2MN$ real equations and Q real unknowns. However, due to the Hermitian constraints not all $2MN$ equations are independent. A careful analysis yields the following result.

Theorem 1 (Number of equations) *The number of independent equations obtained from (5) is $K(2M - K)$, where $K = \min(M, N)$. Therefore, we require $Q \leq K(2M - K)$.*

2.4. Design of the CD Codes

We introduced the CD structure (4) and showed how to choose Q according to Thm.1. What remains is to design the Hermitian basis matrices $A_1, \dots, A_Q$ and choose the discrete set $\mathcal{A}_r$ from which the α_q are drawn.

If the rates being considered are reasonably small then maximizing $|\det(V_\ell - V_{\ell'})|$ 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 $|\det(V_\ell - V_{\ell'})|$, but rather by how well the matrices are distributed over the space of unitary matrices. The optimal distribution is given below.

Theorem 2 *The mutual information between the unitary matrix V and $(X_{\tau-1}, X_\tau)$ in the differential scheme*

$$\begin{bmatrix} X_{\tau-1} \\ X_\tau \end{bmatrix} = \sqrt{\rho} \begin{bmatrix} S_{\tau-1} \\ VS_{\tau-1} \end{bmatrix} H + \begin{bmatrix} W_{\tau-1} \\ W_\tau \end{bmatrix}, \quad (6)$$

where $S_{\tau-1} = V_{z_{\tau-1}} V_{z_{\tau-2}} \dots$, and where H, $W_{\tau-1}$, and W_τ are $M \times N$ matrices with independent $\mathcal{CN}(0,1)$ entries, is maximized when V is isotropically distributed.

An *isotropically-distributed* unitary matrix is one whose probability density function is invariant to pre- and post-multiplication by an arbitrary unitary matrix [21, 22]. In the Cayley transform domain Thm. 2 translates to the following.

Theorem 3 *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*

$$p(A) = \frac{2^{M^2-M}(M-1)! \cdots 1!}{\pi^{M(M+1)/2}} \frac{1}{\det(I + A^2)^M}. \quad (7)$$

(7) 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 CD code constellation should resemble samples from a Cauchy random matrix distribution. Drawing upon implications from the scalar case $M = 1$, 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 $\theta \in \{\pi/r, 3\pi/r, 5\pi/r, \dots, (2r-1)\pi/r\}$. Thus, $\mathcal{A}_2 = \{-1, 1\}$, $\mathcal{A}_4 == \{-2.41, -0.41, 0.41, 2.41\}$, $\mathcal{A}_8 = \{-5.03, -1.50, -0.67, -0.20, 0.20, 0.67, 1.50, 5.03\}$, and so on. Note the points rapidly spread themselves out as r increases, reflecting the long tail of the Cauchy distribution.

Finally, we propose to choose the basis matrices such that the resulting A emulates samples of a Cauchy random matrix, or equivalently V emulates samples of an isotropically unitary matrix. For this we use the following result.

Theorem 4 *Let V and V' be independent $M \times M$ random unitary matrices with distribution $p_V(\cdot)$. Then*

$$\frac{1}{M} \mathrm{E} \log\det(V - V')(V - V')^* \leq 0, \quad (8)$$

with equality when $p_V(\cdot)$ is the isotropic distribution.

This implies that the cost function in (8) is maximized by the isotropically random distribution. We therefore propose to choose $A_1, \dots, A_Q$ to maximize this cost function. After some algebra, this leads to

$$\arg \max_{A_q = A_q^*, q=1,\dots,Q} \left[-\frac{2}{M} \mathrm{E} \log\det(I + A^2) + \frac{1}{M} \mathrm{E} \log\det A^2 \right], \quad (9)$$

where $A = \sum_{q=1}^{Q} A_q \alpha_q$ and the expectation is over $\alpha_1, \dots, \alpha_Q$ chosen independently from a Cauchy distribution. This optimization may be performed numerically using gradient-ascent methods along with Monte Carlo simulation.

3. SIMULATION RESULTS

Fig. 1 compares the BER performance of a $M = N = 2$ CD code at rate $R = 6$, decoded via sphere decoding, versus a rate $R = 6$ orthogonal design from [11]. The superior performance of the CD code is clear. Fig. 2 compares the BER and BLER performances of sphere decoding versus ML decoding for a $M = 4$, $N = 2$, $R = 4$ CD code. The performance loss is not significant and well worth the computational savings over an exhaustive ML search. Finally, Fig. 3 shows the drastic improvement of sphere decoding over nulling and cancelling for an $M = N = 4$, $R = 8$ CD code. In this example the computational complexity of both methods was comparable.

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] J. G. Proakis, *Digital Communications*. McGraw-Hill, 2000.

[9] 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`.

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

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

[12] A. Shokrollahi, B. Hassibi, B. Hochwald, and W. Sweldens, "Representation theory for high-rate multiple-antenna code design," *IEEE Trans. Info. Theory*, vol. 47, no. 6, pp. 2335–2367, 2000. Download available at `http://mars.bell-labs.com`.

[13] 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.

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

[15] 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.

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

[17] B. Hassibi and M. Khorrami, "Fully-diverse multi-antenna constellations and fixed-point-free Lie groups," *submitted to IEEE Trans. Info. Theory*, 2000. Download available at `http://mars.bell-labs.com`.

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

[19] A. A. Sagle and R. E. Walde, *Introduction to Lie Groups and Lie Algebras*. Academic Press, 1986.

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

[21] 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.

[22] 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.

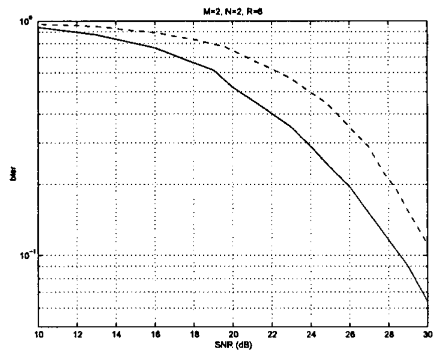

Fig. 1. $M = N = 2$, $R = 6$. Solid: CD code with sphere decoding. Dashed: orthogonal design with ML decoding.

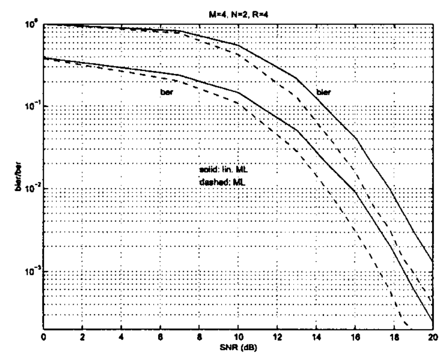

Fig. 2. $M = 4$, $N = 2$, $R = 4$. Solid: sphere decoding. Dashed: ML decoding.

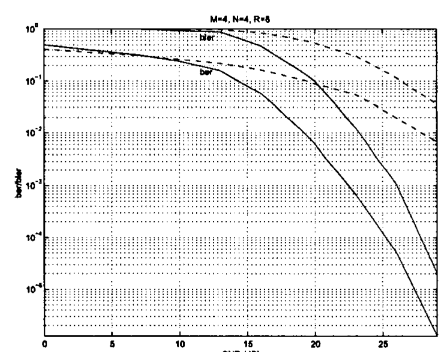

Fig. 3. $M = N = 4$, $R = 8$. Solid: sphere decoding. Dashed: Nulling and cancelling.