Theorem 7: If p is odd, then $\mathcal{S}$ is a family of $q^e - 1$ GF(p)-ary sequences of period $q^e - 1$ all of whose cross correlations and shifted autocorellations are bounded by

$$\left| C_{S_f^{(1,B)}, S_f^{(1,B')}}(\tau) \right| \le q^{e/2}(q^2 - 1) + 1. \tag{11}$$

We have a similar family when p is even.

Theorem 8: If $p = 2$, then $\mathcal{S}$ is a family of $q^e - 1$ binary sequences of period $q^e - 1$ all of whose cross correlations and shifted autocorelations are bounded by

$$\left| C_{S_f^{(1,B)}, S_f^{(1,B')}}(\tau) \right| \le \begin{cases} 2q^{e/2}(q^2-1)+1, & \text{if } e \text{ is even} \\ \frac{4}{3}\, q^{e/2}(q^2-1)+1, & \text{if } e \text{ is odd and } n \text{ is even} \\ q^{e/2}(q^2-1)+1, & \text{if } e \text{ and } n \text{ are odd.} \end{cases}$$

In particular, if f is as in Section VII, then the linear complexity of every sequence in $\mathcal{S}$ is large as well.

When $n = 1$, the sequences studied here reduce to one case of Gold's sequences. However, in this case, our estimates of the cross correlations are too high. For example, when $p = 2$ and e is odd, they are too high by a factor of three. We conjecture, therefore, that our estimates are too high in general. In the case when p is odd and $en = 2m + 1$, some improvement would come if we could choose f so that $\psi(Cu + Dv)F(u)F(v)$ is small for every $C, D \in$ GF(q^e). However, the greatest improvement would come from sharper bounds on W_3. The situation is similar when en is even and when $p = 2$.

REFERENCES

[1] L. Carlitz, "Explicit evaluation of certain exponential sums," *Math. Scand.*, vol. 44, pp. 5–16, 1979.

[2] ——, "Evaluation of some exponential sums over finite fields," *Math. Nachrichtentech.*, pp. 319–339, 1980.

[3] A. H. Chan and R. Games, "On the linear span of binary sequences from finite geometries, q odd," *IEEE Trans. Inform. Theory*, vol. 36, pp. 548–552, May 1990.

[4] R. Gold, "Optimal binary sequences for spread spectrum multiplexing," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 619–620, Oct. 1967.

[5] T. Kasami, "Weight distribution formula for some classes of cyclic codes," Coordinated Sci. Lab., Univ. Illinois, Urbana-Champaign, Tech. Rep. R–285 (AD632574), 1966.

[6] ——, "Weight distribution of Bose–Chaudhuri–Hocquenghem codes," in *Combinatorial Mathematics and its Applications*. Chapel Hill, NC: Univ. North Carolina Press, 1969.

[7] E. L. Key, "An analysis of the structure and complexity of nonlinear binary sequence generators," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 732–736, Nov. 1976.

[8] A. Klapper, "Cross-correlations of geometric sequences in characteristic two," *Des., Codes, Cryptogr.*, vol. 3, pp. 347–377, 1993.

[9] ——, "d-form sequences: Families of sequences with low correlation values and large linear span," *IEEE Trans. Inform. Theory*, vol. 41, pp. 423–431, Mar. 1995.

[10] ——, "Large families of sequences with low correlations and large linear span," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1241–1248, July 1996.

[11] A. Klapper, A. H. Chan, and M. Goresky, "Cross-correlations of linearly and quadratically related geometric sequences and GMW sequences," *Discr. Appl. Math.*, vol. 46, pp. 1–20, 1993.

[12] ——, "Cascaded GMW sequences," *IEEE Trans. Inform. Theory*, vol. 39, pp. 177–183, Jan. 1993.

[13] P. V. Kumar and R. A. Scholtz, "Bounds on the linear span of bent sequences," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 854–862, Nov. 1983.

[14] A. Lempel and M. Cohn, "Maximal families of bent sequences," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 865–868, Nov. 1982.

[15] R. Lidl and H. Niederreiter, "Finite fields," in *Encyclopedia of Mathematics and Its Applications*. Reading, MA: Addision Wesley, 1983, vol. 20.

[16] J. No, "A new family of binary pseudorandom sequences having optimal periodic correlation properties and large linear span," Ph.D. dissertation, Univ. Southern Calif., Los Angeles, 1988.

[17] J. No and P. V. Kumar, "A new family of binary pseudorandom sequences having optimal periodic correlation properties and large linear span," *IEEE Trans. Inform. Theory*, vol. 35, pp. 371–379, Mar. 1989.

[18] J. D. Olsen, R. A. Scholtz, and L. R. Welch, "Bent-function sequences," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 858–864, Nov. 1982.

[19] O. Rothaus, "On bent functions," *J. Comb. Theory Ser. A*, vol. 20, pp. 300–305, 1976.

[20] W. Sun and Y. X. Yang, "Correlation functions of a family of generalized geometric sequences," *Discr. Appl. Math.*, vol. 80, pp. 193–201, 1997.

[21] ——, "Correlations of pseudo-generalized geometric sequences," in *Proc. IEEE Int. Symp. Information Theory*, Ulm, Germany, 1997, p. 44.

Multiple-Antenna Signal Constellations for Fading Channels

Dakshi Agrawal, Thomas J. Richardson, and Rüdiger L. Urbanke

***Abstract*—In this correspondence, we show that the problem of designing efficient multiple-antenna signal constellations for fading channels can be related to the problem of finding packings with large minimum distance in the complex Grassmannian space. We describe a numerical optimization procedure for finding good packings in the complex Grassmannian space and report the best signal constellations found by this procedure. These constellations improve significantly upon previously known results.**

***Index Terms*—Multiple antennas, packings in Grassmannian space, Rayleigh flat fading.**

I. INTRODUCTION

Consider a communication system operating in a Rayleigh flat-fading environment. It is well know that if the receiver knows the channel, then the capacity of the communication system increases linearly with the minimum of the number of transmit and receive antennas [13], [2]. In the past few years, several coding schemes have been proposed to exploit this potential increase in the capacity [12]. However, if the fading environment changes rapidly or if a large number of transmit and receive antennas are employed, then estimating the channel may not be efficient or even feasible. In the case of channel estimation errors, the capacity has been shown to be bounded in signal-to-noise ratio (SNR) [7]. In such scenarios, it is more realistic to assume that the receiver does not know the channel.

Marzetta and Hochwald [8] analyzed the capacity of multiple-antenna communication systems under the assumption that the channel is unknown at the receiver. They also showed that in this case the channel

Manuscript received December 21, 1999; revised December 15, 2000. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Sorrento, Italy, June 2000.

D. Agrawal was with Bell Labs, Lucent Technology, Murray Hill, NJ 07974 USA. He is now with T. J. Watson Research Center, IBM Corporation, Hawthorne, NY 10538 USA (e-mail: agrawal@us.ibm.com).

T. J. Richardson was with Bell Labs, Lucent Technology, Murray Hill, NJ 07974 USA. He is now with Flarion Technologies, Bedminister, NJ 07921 USA (e-mail: richardson@flarion.com).

R. L. Urbanke was with Bell Labs, Lucent Technology, Murray Hill, NJ 07974 USA. He is now with EPFL, LTHC-DSC, CCH-1015 Lausanne, Switzerland (e-mail: rudiger.urbanke@epfl.ch).

Communicated by M. L. Honig, Associate Editor for Communications.

Publisher Item Identifier S 0018-9448(01)06477-X.

capacity increases with the number of transmit and receive antennas, although the increase in capacity is somewhat smaller. Nevertheless, the potential increase in capacity is still significant, and it yields a strong motivation for studying signaling schemes which can achieve a large fraction of the promised capacity at a reasonable complexity. More recently, based on various theoretical and intuitive arguments, Hochwald and Marzetta [5] have advocated the use of *unitary space–time constellations* over such channels. They showed that under some mild conditions, systems using unitary space–time constellations in conjunction with coding can achieve channel capacity.

The objective of this correspondence is to find unitary space–time constellations which minimize uncoded bit error rate to the extent possible. We show that the problem of designing such unitary space–time constellations can be related to the problem of finding packing with large minimum distance in complex Grassmannian space. Conway, Harding, and Sloan [1] have considered an analogous packing problem in real Grassmannian space. We extend their methodology to the complex Grassmannian space and design unitary space–time constellations for one, two, and three transmit antennas for coherence intervals of length up to ten. These packings improve significantly upon the previously known results [4].

The rest of this correspondence is organized as follows. In Section II, we describe the channel model and consider the pairwise probability of error between two points of a unitary space–time constellation. The Chernoff bound on the pairwise probability of error leads to a figure of merit for unitary constellations. In Section III, we show that the proposed figure of merit connects the problem of finding good unitary space–time constellations to the problem of finding good packings in the complex Grassmannian space.

Motivated by this connection, we propose to find good packings in complex Grassmannian space. In Section IV, we describe our optimization technique in detail which uses the *relaxation method* in conjunction with gradient search algorithms. In particular, we describe a suitable parameterization of the tangent space of certain matrices, which is used extensively to compute gradients in an efficient manner. The results obtained by our numerical search are presented in Section V. For a few values of $T,\ M,$ and $L,$ where optimal packings in complex Grassmannian space are known, we show that in most cases, our search program succeeds in finding nearly optimal packings. For other values of $T,\ M,$, and $L,$ the packings found by our search program improve significantly upon previously known packings. We provide simulation results to demonstrate the capability of the discovered packings to reduce the bit error rate when employed as unitary space–time constellations. Finally, in the Appendix, we give in detail the parameterization of certain matrices and spaces that are used in our search programs.

We use the following notation in the correspondence. Let A be a square complex matrix. The element in ith row and jth column of A is denoted by A_{ij}. A^* denotes the conjugate transpose of A. $\partial A/\partial\theta$ denotes the matrix obtained by taking partial derivative of the elements of A with respect to θ, i.e., $(\partial A/\partial\theta)_{ij} = \partial A_{ij}/\partial\theta$.

II. The Channel Model and the Figure of Merit

Consider a multiple-antenna communication system equipped with M transmit antennas and N receive antennas that operates in a Rayleigh flat-fading environment. We assume that the fading processes between distinct pairs of transmit and receive antennas are statistically independent. Further, we assume that the fading is *quasi-static*, i.e., the fading coefficients among different pairs of transmit and receive antennas remain constant for a *coherence interval* of T symbol periods, and then change simultaneously to independent realizations after every T symbol periods. For such channels, using the complex baseband representation, the channel input–output relationship over one coherence interval can be modeled as

$$X_{T\times N} = \sqrt{\frac{\rho}{M}}\, S_{T\times M} H_{M\times N} + W_{T\times N}$$

where X is the $T\times N$ matrix of received signals, S is the $T\times M$ matrix of transmitted signals, H is the $M\times N$ matrix of the Rayleigh fading coefficients, and W is the $T\times N$ matrix of the additive receiver noise. Furthermore, ρ is the expected SNR at each receive antenna. In this notation, the jth column of S represents the signal transmitted over the jth transmit antenna as a function of time. It is assumed that the entries of W are independent samples of a zero-mean circularly symmetric complex Gaussian random variable. Note that we only need to consider the case when $M \leqslant T$, since Marzetta and Hochwald [8] have shown that the capacity obtained with $M > T$ transmit antennas equals the capacity obtained with $M = T$ transmit antennas.

Based on various theoretical and intuitive arguments, Hochwald and Marzetta [8], [5] advocate the use of unitary space–time constellations over such channels. A unitary space–time constellation of L signals consists of $T\times M$ complex matrices $\Phi_1,\ \Phi_2,\ \ldots,\ \Phi_L$, each with orthonormal columns. The first step in designing such signal constellations is to obtain a tractable design criterion. In general, the criterion of minimizing the uncoded bit error rate is intractable and one looks at other simplified performance measures, such as the pairwise probability of error, to design signal constellations. Such a simplified criterion is usually justified by the observation that for large enough SNR, the union bound, which depends only on the pairwise probability of error, is a sufficiently accurate approximation to the bit error rate.

Under maximum-likelihood decoding, the pairwise probability of error between two unitary signal points Φ_1 and Φ_2 is an intricate function of the singular values $\lambda_1,\ \ldots,\ \lambda_M$ of $\Phi_1^*\Phi_2$ [5]. In order to obtain a tractable design criterion, we look for a *single* figure of merit that can be used to closely approximate this function. Clues for this figure of merit are provided by the Chernoff bound on the pairwise probability of error. As discussed in [5], the Chernoff bound on the pairwise probability of error is given by

$$P_e \leqslant \frac{1}{2}\prod_{j=1}^{M}\left[\frac{1}{1+\frac{(\rho T/M)^2(1-\lambda_j^2)}{4(1+\rho T/M)}}\right]^N. \tag{1}$$

For one transmit antenna, the Chernoff bound is an increasing function of λ_1^2. While for two transmit antennas, it increases with

$$(\lambda_1^2+\lambda_2^2) - \left(\frac{1}{1+\frac{2M}{\rho T}}\right)^2 \lambda_1^2\lambda_2^2. \tag{2}$$

Either for small values of λ_1 and λ_2 or for a low SNR, $\rho,\ (\lambda_1^2+\lambda_2^2)$ becomes the dominating term in (2). Fig. 1 shows this dominance. In each of the three graphs, a particular SNR is fixed, and the pairwise probability of error versus $(\lambda_1^2/2+\lambda_2^2/2)^{1/2}$ for $\lambda_1/\lambda_2 = 0.001, 0.5,$ and 1.0 is plotted. Clearly, for small values of $(\lambda_1^2/2+\lambda_2^2/2)^{1/2}$, the pairwise probability of error is relatively insensitive to the ratio λ_1/λ_2. It turns out that the values of $(\lambda_1^2/2+\lambda_2^2/2)^{1/2}$ considered in this correspondence are small enough to make $(\lambda_1^2/2+\lambda_2^2/2)^{1/2}$ an accurate measure of the pairwise probability of error. Thus, for two transmit antennas in the region of our interest, the Chernoff bound on the pairwise probability of error is an (approximately) increasing function of $(\lambda_1^2/2+\lambda_2^2/2)^{1/2}$.

The above observation motivates the use of $\lambda_1^2+\lambda_2^2+\cdots+\lambda_M^2$ as the quantity that governs the pairwise probability of error between two signal points Φ_1 and Φ_2, where $\lambda_1,\ \ldots,\ \lambda_M$ are the singular values of $\Phi_1^*\Phi_2$. We refer to the sum of squared singular values of $\Phi_1^*\Phi_2$ as

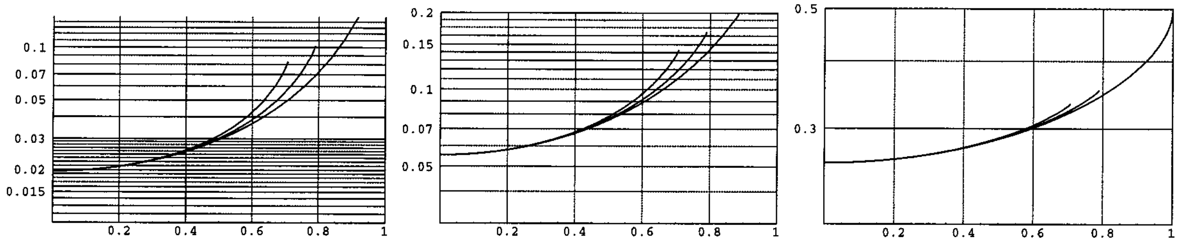

Fig. 1. Pairwise probability of error versus $(\lambda_1^2/2 + \lambda_2^2/2)^{1/2}$ for $\rho = 10$ (left graph), $\rho = 5$ (middle graph), and $\rho = 1$ (right graph). Each graph has three curves corresponding to $\lambda_1/\lambda_2 = 0.001$ (leftmost curve), $.50$, and 1 (rightmost curve), respectively.

the *correlation* between Φ_1 and Φ_2. This correlation can be computed explicitly in terms of Φ_1 and Φ_2 by using

$$\lambda_1^2 + \lambda_2^2 + \cdots + \lambda_M^2 = \langle \Phi_1^* \Phi_2,\, \Phi_1^* \Phi_2 \rangle$$

where $\langle A,\, B \rangle$ is the standard inner product of complex matrices A and B and is equal to $\sum_{j,\,k} A_{jk} B_{jk}^*$ [6, p. 271].

For one transmit antenna, minimizing the correlation also minimizes the exact pairwise probability of error, while for more than one transmit antenna, it is only an approximation to the more exact expression. As the number of terms in the Chernoff bound (cf. (1)) grows with M, the correlation between two signal matrices is not expected to be as good an indicator of the pairwise probability of error for a large number of transmit antennas. Nevertheless, as we show in this correspondence, for a small number of transmit antennas, maximum correlation between two distinct signal points is a good *figure of merit* for unitary space–time constellations. Constellations that optimize this figure of merit also have low overall probability of error. Moreover, this figure of merit provides a connection between good unitary space–time constellations and dense packings in the complex Grassmannian space. In the next section, we describe this connection and show how the algorithms to find dense packings in the complex Grassmannian space can be used to find good unitary space–time constellations.

III. Unitary Space–Time Constellations and Packings in Complex Grassmannian Space

We start with some definitions. The complex Grassmannian space $\mathcal{G}(T,\, M,\, \mathbb{C})$ is the set of all M-dimensional subspaces of the complex T-dimensional vector space $\mathbb{C}^T$. It forms a compact Riemannian manifold of dimension $(2MT - 2M^2)$ [11, p. 530]. Let $\mathcal{U}(T \times M)$ be the set of all $T \times M$ complex matrices with orthonormal columns. Let $\Phi_1,\, \Phi_2 \in \mathcal{U}(T \times M)$ be two orthonormal matrices whose column spaces are $P_1,\, P_2 \in \mathcal{G}(T,\, M,\, \mathbb{C})$, respectively. The distance between P_1 and P_2 can be defined as

$$\begin{aligned} d(P_1,\, P_2) &:= \sqrt{T - (\lambda_1^2 + \lambda_2^2 + \cdots + \lambda_M^2)} \\ &= \sqrt{T - \langle \Phi_i^* \Phi_j,\, \Phi_i^* \Phi_j \rangle} \end{aligned}$$

where $\lambda_1,\, \ldots,\, \lambda_M$ are the singular values of $\Phi_1^* \Phi_2$. One can check that this distance function is independent of the choices of matrices Φ_1 and Φ_2, and hence it is well defined.

A finite subset S of the complex Grassmannian space $\mathcal{G}(T,\, M,\, \mathbb{C})$ is referred to as a *packing* in $\mathcal{G}(T,\, M,\, \mathbb{C})$ of cardinality $|S|$. The minimum distance $d(S)$ of a packing S is given by

$$d(S) := \min_{\substack{P_i,\, P_j \in S \\ P_i \neq P_j}} d(P_i,\, P_j). \tag{3}$$

With this definition, we can state the packing problem in complex Grassmannian space as follows.

Packing Problem*:* Given the natural numbers $T,\, M,$ and $L,$ with $M \leqslant T$, design a packing S in $\mathcal{G}(T,\, M,\, \mathbb{C})$ of cardinality $|S| = L$ so that its minimum distance $d(S)$ is as large as possible. □

A packing S^* in $\mathcal{G}(T,\, M,\, \mathbb{C})$ is said to be optimal if $d(S^*) \geqslant d(S)$ for any other packing S in $\mathcal{G}(T,\, M,\, \mathbb{C})$ of the same cardinality. It is evident from the definition of the minimum distance of a packing and from the chosen figure of merit for a unitary space–time constellation that the problem of finding packings with a large minimum distance and the problem of designing a unitary space–time constellation with small maximum correlation between its points are equivalent. Indeed, using the orthonormal bases of the subspaces, the packing problem can be recast as follows.

Variation of the Packing Problem*:* Given natural numbers $T,\, M,$ and $L,$ with $M \leqslant T$, find a collection

$$S = \{\Phi_1,\, \Phi_2,\, \ldots,\, \Phi_L\} \subset \mathcal{U}(T \times M)$$

such that the maximum correlation between two distinct points of S, given by

$$\sigma^*(S) := \max_{1 \leqslant i < j \leqslant L} \sqrt{\langle \Phi_i^* \Phi_j,\, \Phi_i^* \Phi_j \rangle} \tag{4}$$

is minimized. □

Thus, the problem of designing unitary space–time constellations with a good figure of merit is the same as the problem of finding packings in complex Grassmannian space with large minimum distance. This is not too surprising since, fundamentally, a unitary space–time constellation is just a collection of M-dimensional subspaces of $\mathbb{C}^T$. Recall that in the absence of the additive noise at the receiver, the received signal at each antenna is a linear combination of the columns of the transmitted matrix. The coefficients of this linear combination are circularly symmetric random variables. Therefore, if we change a particular signal matrix Φ_i to another orthonormal matrix that has the same column span as Φ_i, it will not alter the statistics of the received signal. More concretely, the conditional distribution of the received signal X given the transmitted signal Φ_i is given by [8], [5]

$$p(X|\Phi_i) = \frac{\exp(-\mathrm{tr}\{[I_T + (\rho/M)\Phi_i \Phi_i^*]^{-1} X X^*\})}{\pi^{TN} \det^N [I_T + (\rho/M)\Phi_i \Phi_i^*]} \tag{5}$$

which remains unchanged if we replace the signal Φ_i with the signal $\Phi_i U_{M \times M}$, where $U_{M \times M}$ is a unitary matrix. Therefore, at the receiver, two signal matrices with the same column spans look identical and a unitary constellation is just a collection of M-dimensional subspaces of $\mathbb{C}^T$.

Since the complex Grassmannian space $\mathcal{G}(T,\, M,\, \mathbb{C})$ is a $(2TM - 2M^2)$-dimensional differentiable manifold, the packing problem is an optimization problem involving $(2TM - 2M^2) \times L$ parameters. On the other hand, as shown in the Appendix, any element of $\mathcal{U}(T \times M)$ can be represented by $2TM - M^2$ parameters. Therefore, the variation of the packing problem given above is an optimization problem involving

$(2TM - M^2) \times L$ parameters. It turns out that despite the economy offered by the direct parameterization of the manifold $\mathcal{G}(T, M, \mathbb{C})$, it is preferable to solve the second version of the packing problem using a parameterization of $\mathcal{U}(T \times M)$. Many important quantities, such as the value of $d(S)$, can be computed relatively faster in the second representation. Moreover, unlike the first problem, where the parameters can take unbounded values, values of parameters in the second problem are bounded (see the Appendix). This leads to a higher degree of numerical stability in the optimization algorithm. Due to these reasons, we decided to work with $\mathcal{U}(T \times M)$, and we minimized the maximum correlation $\sigma^*(S)$ for a collection of orthonormal matrices S. In the next section, we will provide the details of our optimization technique.

IV. Optimization Technique

The parameters involved in the parameterization of $\mathcal{U}(T \times M)$ lie in a compact differential manifold. Moreover, $\sigma^*(S)$ is a continuous function of its parameters. Therefore, it achieves at least one global minimum, and one can consider the direct minimization of $\sigma^*(S)$ using gradient search algorithms. Unfortunately, several properties of $\sigma^*(S)$ render gradient search algorithms problematic. First, since a rotated constellation has the same correlation properties as the original constellation, $\sigma^*(S)$ does not have a unique global minima. It may have many local minima that are far away from global minima. Second, $\sigma^*(S)$ is not very smooth—in fact, it is not even differentiable everywhere. Both of these facts obviate the direct use of gradient-based search algorithms.

Conway, Harding, and Sloane [1] have considered a similar packing problem in *real* Grassmannian space, and they propose an algorithm that relaxes the optimization problem to overcome the difficulties posed by the undesirable properties of the objective function. Their basic idea is to introduce a family of surrogates f_α for σ^*, parameterized by the positive parameter α. The functionals f_α mimic σ^*: for large values of α, the functionals f_α closely track σ^*, converging to σ^* as α tends to ∞. For small values of α, the functionals f_α overcome the difficulties indicated above while mimicking σ^* less well. The strategy is to minimize f_α for small α and then track this minimum while increasing α. It turns out that a similar technique can also be used in complex Grassmannian space. In the next two subsections, we describe this technique in more detail and discuss various implementation issues.

A. General Description

Ideally, the family of surrogate functionals f_α has the following properties.

1) For all α the functional f_α is smooth.
2) For small values of α the functional f_α has few local minima.
3) The functionals f_α mimic σ^*.

One example of such a family is

$$f_\alpha(S) = \frac{1}{\alpha}\log\left(\sum_{1\leqslant i<j\leqslant |S|} \exp(\alpha\langle \Phi_i^*\Phi_j,\, \Phi_i^*\Phi_j\rangle)\right). \qquad (6)$$

(Note that the results in this correspondence were obtained by directly optimizing the argument of the $\log$ in this expression.)

The purpose of condition 1) is to admit an optimization of f_α by standard gradient descent techniques. The purpose of condition 2) is to enable us to find global minima of f_α relatively easily. In fact, the example given above has only one minimum point for $M = 1$. The purpose of condition 3) is clear, we expect that minimizers of f_α, even for small α, will be "close to" minimizers of σ^*. The intention is to find a minimizer of f_α for small α and to slowly increase the value of α while tracking this minimum point. Any limit of a local minimizer of f_α is guaranteed to be a critical point of σ^*. Any limit point of a global minimizer of f_α is guaranteed to be a global minimizer of σ^*. As we use gradient search algorithms to find the minimum of $f_\alpha(S)$, it is not certain that we will be able to find a global minimum point even for low values of α. Therefore, there is no guarantee that the point we track will lead us to a global minimum of $\sigma^*(S)$. There is still a need to try various random initial conditions and to evaluate, as far as possible, the results from the search procedure for their optimality.

In summary, the algorithm is the following. The search procedure starts with a relatively small value of α, say α_0, and a randomly generated set S. If a good initial set S is known *a priori*, then we can also start from this set. Starting from the initial set S, we use numerical optimization techniques to find a set S_{α_0} such that the value of f_{α_0} is (nearly) locally minimized. We now slightly increase α to α_1 and starting from the set S_{α_0}, find a new set S_{α_1} that (nearly) locally minimizes f_{α_1}. We continue in this manner, each time increasing the value of α slightly and tracking the minimizer of f_α. For very large values of α, f_α would be essentially equivalent to σ^* and minimizing f_α will also essentially minimize σ^*. This optimization technique is often referred to as *the relaxation method* [3].

B. Implementation Details

The optimization problem at hand involves a large number of parameters. For example, in order to transmit 1 bit/s/Hz on a fading channel with coherence interval of length $T = 10$, we need a signal constellation with $L = 2^{10}$ signal points. For $M = 3$ transmit antennas, the corresponding optimization problem involves $(2\cdot 10\cdot 3 - 3^2)\cdot 1024 = 52\,224$ parameters. In order to deal with such a large number of parameters and to execute the optimization algorithm efficiently, we used several suboptimal steps and fine-tuned the optimization algorithm. In this subsection, we highlight some of these implementation details.

As described above, for each α_k, the algorithm starts from a given set $S_{\alpha_{k-1}}$ and attempts to adjust it to find a minimum point of f_{α_k}. A variety of descent algorithms are available for such minimizations. They use local properties of the cost function to iteratively update the set S, thereby reducing the value of the cost function at each iteration. We use two methods for updating S. At each iteration, we either update, in a fixed predetermined order, each matrix in S individually, or we update all matrices in S simultaneously. The first method involves calculating the local properties of the cost function with respect to variations in only one matrix and updating that matrix suitably. While in the second method, we need to compute local properties with respect to the variations of the whole set S. Most of the results in this correspondence were obtained by updating only one matrix at a time, and henceforth we will only describe this case. Note that in updating the matrices' Φ_i's, we have to ensure that the columns of each Φ_i remain orthonormal. (One could consider relaxing the problem further so that this constraint need be satisfied only asymptotically in α, but we have not taken this approach here.)

Since our ultimate objective is to minimize σ^* and not to minimize f_α, at least for small α's, it suffices to find a "near" minimizer of f_α. It is sufficient to stay within the domain of attraction of the minimum as α increases and then to converge closer to the minimum when α is large. Therefore, we do not spend a lot of computational resources on pinpointing the local minimum of f_α. We impose the rule that α is increased as soon as the fractional change in the value of the cost function in two consecutive iterations becomes less than a threshold. Depending upon how close to the minimum we want to get, this threshold can be adjusted during the iterations. We stop our search procedure when the decrease in $\sigma^*(S)$ becomes too small to justify further computational resources.

In our implementation we used the steepest descent algorithm to minimize the cost function. A key (and computationally most intensive) step in this algorithm is to compute the steepest descent direction (gradient) of the cost function. In order to simplify computations, we used the families of cost functions f_α that depend on the set S only through the pairwise correlations of the elements of S (cf. (6)). For such cost functions, by applying the chain rule for partial differentiation, we can reduce the problem of finding the steepest descent direction of the cost function to the problem of computing partial derivatives of the pairwise correlation $\langle \Phi_i^* \Phi_j, \Phi_i^* \Phi_j \rangle$ with respect to the parameters of Φ_i and Φ_j.

There are several ways to parameterize the matrices Φ_i's. One possible way is to use the parameterization of $\mathcal{U}(T \times M)$ suggested in the Appendix, and to compute the partial derivatives of the pairwise correlation $\langle \Phi_i^* \Phi_j, \Phi_i^* \Phi_j \rangle$ with respect to these parameters. However, this approach results in cumbersome formulas which are implicit functions of the matrices and are expensive to compute.

A more practical way to parameterize $\mathcal{U}(T \times M)$ is to *overparameterize* the space as follows. Given a matrix $\Phi_i \in \mathcal{U}(T \times M)$, any other matrix $\tilde{\Phi}$ in $\mathcal{U}(T \times M)$ may be written as $U(\Theta)\Phi_i$, where $U(\Theta)$ is a $T \times T$ unitary matrix, parameterized by Θ (see the Appendix). In this way, we can parameterize $\mathcal{U}(T \times M)$ by Θ. Note, however, that this is an overparameterization since, unless $M = T$, $U(\Theta)\Phi_i = \tilde{\Phi}$ does not determine Θ uniquely. Indeed, a subgroup of the group of $T \times T$ unitary matrices leaves the matrix Φ_i invariant. The overparameterization of $\mathcal{U}(T \times M)$ does not present a problem because if we (abstractly) factor out the subgroup of $\mathcal{U}(T \times T)$ which leaves the matrix Φ_i invariant, then a parameterization of the quotient group would uniquely parameterize $\mathcal{U}(T \times M)$. The gradient direction in $\mathcal{U}(T \times M)$ turns out to be the same, however, so there is no need to compute the factorization.

The main advantage of this overparameterization arises from the fact that we can consider perturbing the matrices in S by premultiplying them by matrices $U(\Theta)$ for $\Theta \simeq 0$, since $U(0) = I$. This allows efficient computation of the partial derivatives of the correlation between $U(\Theta)\Phi_i$ and Φ_j by just using entries in matrices Φ_i and Φ_j, and the derivatives of $U^*(\Theta)$ at $\Theta = 0$. Specifically, for $\theta \in \Theta$, we have

$$\left(\frac{\partial}{\partial \theta} \langle \Phi_i^* U^*(\Theta)\Phi_j, \Phi_i^* U^*(\Theta)\Phi_j \rangle \right)\Big|_{\Theta=0}$$

$$= 2\mathrm{Re}\left\{ \left\langle \Phi_i^* \frac{\partial U^*(\Theta)}{\partial \theta}\Phi_j, \Phi_i^* U^*(\Theta)\Phi_j \right\rangle\Big|_{\Theta=0} \right\} \tag{7}$$

$$= 2\mathrm{Re}\left\{ \left\langle \Phi_i^* \frac{\partial U^*(\Theta)}{\partial \theta}\Big|_{\Theta=0} \Phi_j, \Phi_i^* \Phi_j \right\rangle \right\}. \tag{8}$$

Therefore, once we have $\partial U^*(\Theta)/\partial\theta|_{\Theta=0}$ for all $\theta \in \Theta$, partial derivatives of the correlation become an explicit function of Φ_i and Φ_j and can be computed efficiently. In the Appendix, we calculate $\partial U^*(\Theta)/\partial\theta|_{\Theta=0}$ for all $\theta \in \Theta$. By using the chain rule for partial derivatives, we can now compute ∇f_α, the gradient of f_α with respect to Θ and set $\Theta = -c\nabla f_\alpha$, to obtain the updated matrix $U(-c\nabla f_\alpha)\Phi_i$, where $c > 0$ is a step size appropriately chosen. The results in this correspondence were obtained by using this technique to update matrices at each step. It proved to be an efficient and clean way of varying matrices in $\mathcal{U}(T \times M)$.

V. Results and Discussion

As discussed in the previous section, the relaxation technique requires a little fine tuning before it can efficiently output good packings. Among other things, this fine tuning involves adjusting the criterion for increasing α and the rate at which α is increased. We used known optimum packings in $\mathcal{G}(T, 1, \mathbb{C})$ described in [10] as test cases for this fine tuning. Table I shows $\sigma^*(S)$ for the best packings obtained by our computer program along with $\sigma^*(S)$ for the optimal packings. In most cases, the program was able to generate optimal packings within the numerical precision of the machines and the algorithms used. This created some confidence in the effectiveness of the general technique used in this paper. Table I also shows that in two cases the program did not generate optimal packings. This points to some room for further improvement in our methods. This improvement may come from many directions: by a better gradient search algorithm, by using a different family of surrogate functionals, or by updating the set S as a whole as opposed to updating one matrix at a time.

Next, we generated good packings of cardinality 2^T in $\mathcal{G}(T, M, \mathbb{C})$ for $M = 1, 2$, and 3 (see Tables II, III, and IV, respectively). For communications over Rayleigh fading channels, these packings can be used as rate 1 bit/s/Hz unitary space–time constellations. Recall that the values of M correspond to the number of transmit antennas. In all cases, unitary space–time constellations found by our technique improved upon previously best known constellations [5], [10], [4] in terms of having a smaller maximum correlation between signal points. We provide these packings only up to $T = 10$, as our algorithm requires a considerable amount of computing power for large values of T.

TABLE I
Optimal and Achievable $\sigma^*(S)$ in $\mathcal{G}(T, 1, \mathbb{C})$

T	L	$\frac{\log_2(L)}{T}$	$\sigma^*(S)$	Optimal $\sigma^*(S)$
5	11	0.692	0.3465	0.346
5	21	0.878	0.4087	0.400
6	11	0.577	0.2887	0.288
6	31	0.826	0.3783	0.372
7	28	0.687	0.3333	0.3333
7	56	0.830	0.4177	0.378
8	72	0.771	0.3983	0.354
9	19	0.472	0.2490	0.248
9	90	0.721	0.3754	0.3333

TABLE II
Some Good Packings in $\mathcal{G}(T, 1, \mathbb{C})$

T	L	$\frac{\log_2(L)}{T}$	$\sigma^*(S)$	Previously best known $\sigma^*(S)$
5	32	1.000	0.4907	0.515
6	64	1.000	0.5036	0.531
7	128	1.000	0.5140	0.540
8	256	1.000	0.5213	0.545
8	4096	1.500	0.7242	—
9	512	1.000	0.5309	0.580
10	1024	1.000	0.5653	—

TABLE III
Some Good Packings in $\mathcal{G}(T, 2, \mathbb{C})$

T	L	$\frac{\log_2(L)}{T}$	$\frac{\sigma^*(S)}{\sqrt{2}}$	Previously best known $\sigma^*(S)$
5	32	1.000	0.6380	—
6	64	1.000	0.6077	—
7	128	1.000	0.5896	—
8	256	1.000	0.5839	0.66
9	512	1.000	0.5828	—
10	1024	1.000	0.5732	—

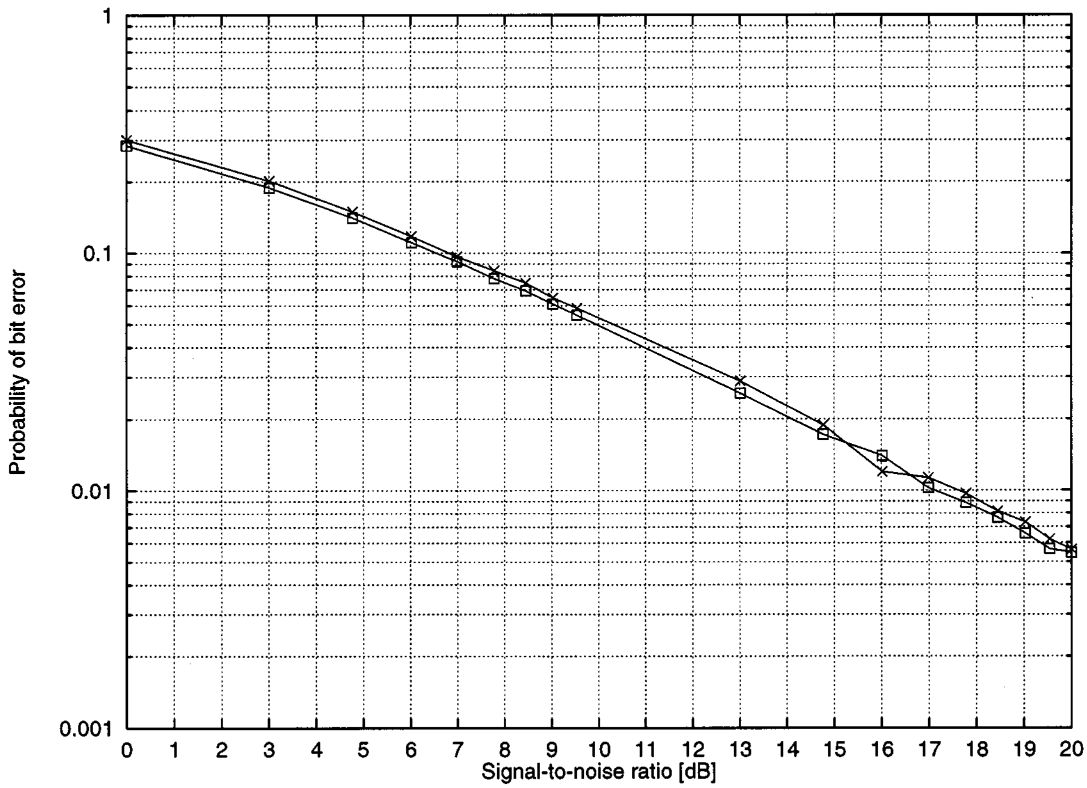

Fig. 2. Bit error rate versus SNR for best new constellation (squares) and best previously known constellation (crosses) for $T = 8$, $M = 1$, and rate 1 bit/s/Hz. The new constellation has 5%–10% better bit error rate.

TABLE IV
SOME GOOD PACKINGS IN $\mathcal{G}(T, 3, \mathbb{C})$

T	L	$\frac{\log_2(L)}{T}$	$\frac{\sigma^*(S)}{\sqrt{3}}$	Previously best known $\sigma^*(S)$
5	32	1.000	0.7791	—
6	64	1.000	0.7091	—
7	128	1.000	0.6853	—
8	256	1.000	0.6590	0.74
9	512	1.000	0.6395	—
10	1024	1.000	0.6266	—

Figs. 2–4 show the average bit error rate for unitary space–time constellations generated here as well as the same for the previously such best known constellations [5], [10], [4] for a coherence interval of length $T = 8$ and rate 1 bit/s/Hz. For one, two, and three transmit antennas, our constellations are respectively, 4%, 12%, and 11% better in terms of the figure of merit. For one transmit antenna, very good packings were already known, and we could improve upon them by only 4%. For two and three transmit antennas, our improvements are more impressive. The smaller maximum correlation between signal points in our constellations translates into a 5%–10%, 15%–30%, and 5%–20% difference in the bit error rate, for one, two, and three transmit antennas, respectively. The exact difference in bit error rate depends on the SNR.

Note that on Rayleigh fading channels (and sufficiently large enough SNR) two different signal constellations have an almost constant ratio of their respective bit error rates. This ratio is in part determined by theminimum distances of signal constellations. This is in stark contrast to the case of Gaussian channels, where the ratio between bit error rates of two constellation changes exponentially with SNR. At first sight, this might seem disappointing. Nevertheless, if such a constellation is used in conjunction with a powerful outer code at bit error rates of several percent (typically 15%–5%) then this percentage decrease in the bit error rate translates into a significant increase in the possible transmission rate. Note also that the improvement for the three-antenna case is not as large as expected from the better minimum distance of our packing. Most probably this is due to the fact that for three transmit antennas the chosen figure of merit is no longer a very good indicator of the performance of a unitary space–time constellation.

Finally, the unitary space–time constellations generated here can be used as a benchmark to assess the optimality of other signal constellations designed for Rayleigh block-faded channels. Our objective in this study was to generate signal constellations with as low bit error rate as possible. Signal constellations which are designed subjected to other constraints such as "easy encoding and decoding" are likely to have a worst bit error rate. Suboptimality of such constellations can be measured by comparing them against the signal constellations generated here [14].

APPENDIX
PARAMETERIZATION OF MATRICES

First, we briefly present a parameterization of T-dimensional unitary matrices which was used extensively to compute the gradient of surrogate functionals. For details regarding this parameterization, we refer the reader to [9]. Later in this appendix, we will apply a similar technique to parameterize $\mathcal{U}(T \times M)$. Throughout this appendix, i denotes $\sqrt{-1}$.

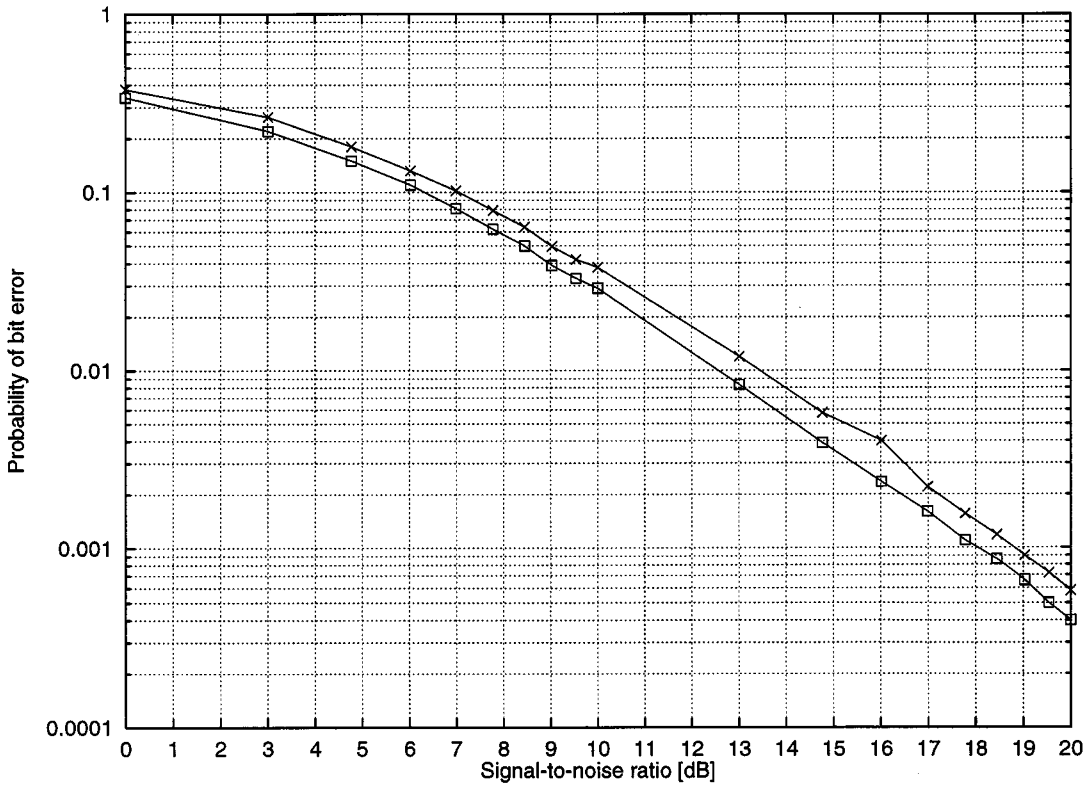

Fig. 3. Bit error rate versus SNR for best new constellation (squares) and best previously known constellation (crosses) for $T = 8$, $M = 2$, and rate 1 bit/s/Hz. The new constellation has 15%–30% better bit error rate.

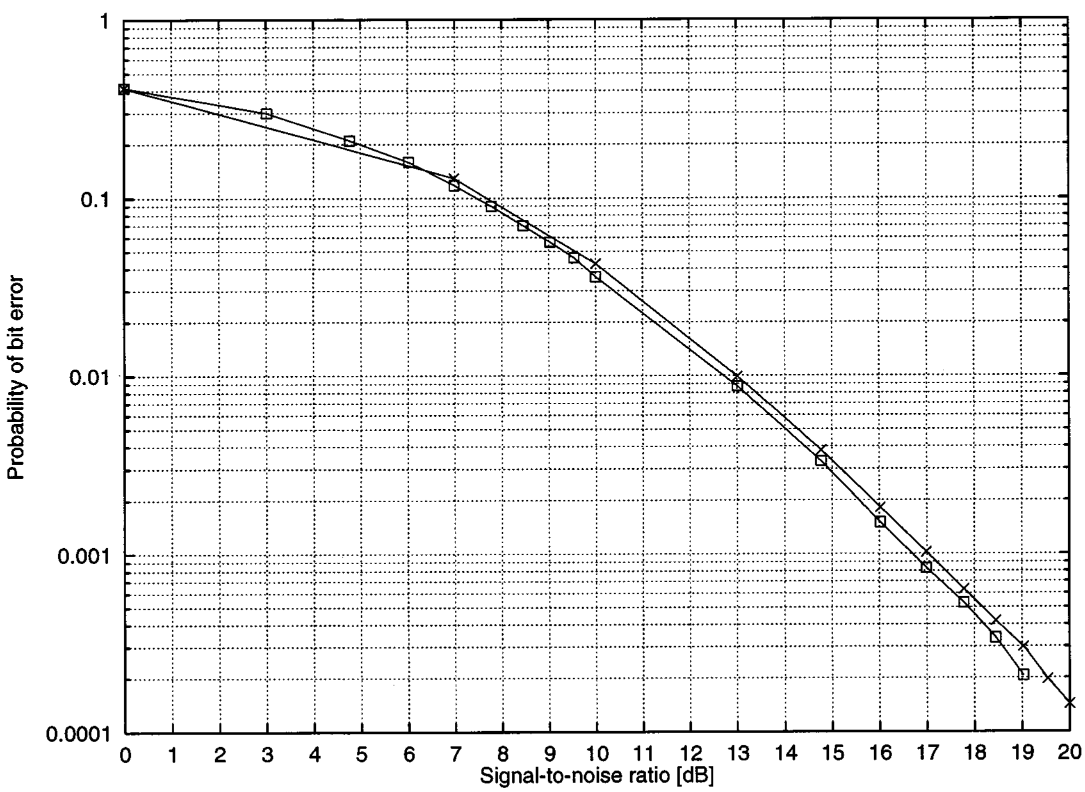

Fig. 4. Bit error rate versus SNR for best new constellation (squares) and best previously known constellation (crosses) for $T = 8$, $M = 3$, and rate 1 bit/s/Hz. The new constellation has 5%–20% better bit error rate.

We start by introducing some basic unitary matrices U^{pq}. Let $U^{pq}(\phi_{pq}, \sigma_{pq})$ for $p < q$ and $\phi_{pq}, \sigma_{pq} \in [-\pi, \pi)$ denote the T-dimensional unitary matrix given by

$$U^{pq}_{jk}(\phi_{pq}, \sigma_{pq}) = \begin{cases} 1, & \text{if } j = k \text{ and } j \neq= p, q \\ \cos(\phi_{pq}), & \text{if } j = k \text{ and } j = p, q \\ -\sin(\phi_{pq})\exp(-i\sigma_{pq}), & \text{if } j = p \text{ and } k = q \\ \sin(\phi_{pq})\exp(i\sigma_{pq}), & \text{if } j = q \text{ and } k = p \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

To illustrate one such U^{pq}, we have

$$\begin{pmatrix} \cos(\phi) & 0 & -\sin(\phi)\exp(-i\sigma) & 0 \\ 0 & 1 & 0 & 0 \\ \sin(\phi)\exp(i\sigma) & 0 & \cos(\phi) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

which is the four-dimensional basic unitary matrix $U^{13}(\phi, \sigma)$. Note that $U^{pq*}(\phi, \sigma) = U^{pq}(-\phi, -\sigma)$. Post-multiplication (pre-multiplication, respectively) by a basic unitary matrix U^{pq} can be used to change the pth and qth entries of a row vector (a column vector, respectively), without changing other entries, in such a way that the squared sum of magnitudes of pth and qth entries remains the same.

We will show that any T-dimensional unitary matrix U can be written as a unique product of these basic unitary matrices with a diagonal matrix. Given an T-dimensional unitary matrix U, consider the unitary matrix $U' = UU^{1T*}(\phi_{1T}, \sigma_{1T})$ obtained by post-multiplying U with the complex conjugate of $U^{1T}(\phi_{1T}, \sigma_{1T})$. We choose the parameters ϕ_{1T} and σ_{1T} in such a way that U' satisfies the following two constraints:

- $U'_{1T} = 0$.
- Either $U'_{11} = 0$ or $-\frac{\pi}{2} \leqslant \arg(U'_{11}) \leqslant \frac{\pi}{2}$.

It can be verified that these conditions uniquely specify $\phi_{1T} \in [-\pi, \pi)$ and $\sigma_{1T} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ (with the convention that we set σ_{1T} to zero in case it is indeterminate).

Similarly, we can post-multiply U' by

$$U^{1(T-1)^*}(\phi_{1(T-1)}, \sigma_{1(T-1)})$$

to obtain U'' such that

- $U''_{1(T-1)} = 0$.
- Either $U''_{11} = 0$ or $-\frac{\pi}{2} \leqslant \arg(U''_{11}) \leqslant \frac{\pi}{2}$.

Again, using the same convention, these conditions uniquely specify $\phi_{1(T-1)} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ and $\sigma_{1(T-1)} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Note that the range of $\phi_{1(T-1)}$ is different from the range of ϕ_{1T} as a result of the condition $-\frac{\pi}{2} \leqslant \arg(U'_{11}) \leqslant \frac{\pi}{2}$.

We can continue this process until all elements of the first row are zero with the exception of the first element. Hence, we obtain

$$U^{T-1} = U\left(U^{12}(\phi_{12}, \sigma_{12})U^{13}(\phi_{13}, \sigma_{13}) \cdots U^{1T}\ (\phi_{1T}, \sigma_{1T})\right)^* \quad (10)$$

$$U = U^{T-1}U^{12}(\phi_{12}, \sigma_{12})U^{13}(\phi_{13}, \sigma_{13}) \cdots U^{1T}(\phi_{1T}, \sigma_{1T}) \quad (11)$$

where U^{T-1} is a unitary matrix with only one nonzero entry in its first row, namely, at position 11. As the norm of a row of a unitary matrix is 1, $U^{T-1}_{11} = \exp(i\delta_1)$ for some $\delta_1 \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. From the orthogonality of the rows of a unitary matrix, it follows that $U^{T-1}_{j1} = 0$ for $j \neq 1$. Therefore, U^{T-1} can be represented as follows:

$$U^{T-1} = \begin{pmatrix} \exp(i\delta_1) & 0 \\ 0 & V \end{pmatrix} \quad (12)$$

where V is an $(T-1)$-dimensional unitary matrix. Note that in order to pass from a T-dimensional unitary matrix to a $(T-1)$-dimensional matrix, we need $2T-1$ parameters: $2 \times (T-1)$ parameters represented by ϕ's and σ's and one parameter represented by δ.

We can operate on V in a similar manner by using matrices $U^{2j}(\phi_{2j}, \sigma_{2j})$ for $j = T, T-1, \ldots, 3$ and pass to a $(T-2)$-dimensional unitary matrix. Continuing this process, we pass to a one-dimensional unitary matrix which consists of just one entry $\exp(i\delta_T)$. Therefore, by this process, we can parameterize the bounded and closed space of T-dimensional unitary matrices by

$$1 + 3 + 5 + \cdots + (2T-1) = T^2$$

parameters. Except ϕ_{jT}, for $j = 1, 2, \ldots, T-1$ and δ_T, which take values in the interval $[-\pi, \pi)$, all other parameters belong to the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Thus, out of T^2 parameters, T belong to the interval $[-\pi, \pi)$, other remaining parameters belong to the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Let Θ be the collection of these T^2 parameters.

Partial derivatives of $U(\Theta)$ at $\Theta = 0$ are easy to evaluate and are given by

$$\left.\frac{\partial U(\Theta)}{\partial \phi_{pq}}\right|_{\Theta=0} = U^{pq}(\frac{\pi}{2}, 0) \quad (13)$$

$$\left.\frac{\partial U(\Theta)}{\partial \sigma_{pq}}\right|_{\Theta=0} = \mathbf{0}_{T\times T} \quad (14)$$

$$\left.\frac{\partial U(\Theta)}{\partial \delta_k}\right|_{\Theta=0} = ie_k.e_k^* \quad (15)$$

where e_k is a unit norm column vector of length T with kth element equal to 1. As discussed in Section IV-B, these partial derivatives were used to calculate the gradient of the cost function.

In a similar manner, we can also parameterize the set $\mathcal{U}(T \times M)$. We start by premultiplying $U \in \mathcal{U}(T \times M)$ with the basic rotation matrix $U^{1T}(\phi_{1T}, \sigma_{1T})$. We choose ϕ_{1T} and σ_{1T} in such a way that the last element of the first column becomes zero and the argument of the first element of the first column lies in the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$. We continue in this manner until only the first element of the first column is nonzero. This process uses $2T-1$ parameters. Again using orthogonality of columns, we can show that this procedure reduces the problem of parameterizing a matrix in $\mathcal{U}(T \times M)$ to the problem of parameterizing a matrix in $\mathcal{U}((T-1) \times (M-1))$. Therefore, to parameterize a matrix in $\mathcal{U}(T \times M)$ we need

$$(2T-1) + (2T-3) + \cdots + (2T-(2M-1)) = 2TM - M^2$$

parameters. Out of these $2TM - M^2$ parameters, M parameters belong to the interval $[-\pi, \pi)$ and other remaining parameters belong to the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

Acknowledgment

The authors wish to thank two reviewers for their helpful suggestions that led to many improvements in this manuscript.

References

[1] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Exper. Math.*, vol. 5, pp. 139–159, 1996.

[2] G. J. Foschini, "Layered space-time architecture for wireless communication in a flat fading environment when using mulit-element antennas," *Bell Labs. Tech. J.*, vol. 1, pp. 41–50, 1996.

[3] H. Hardin and N. J. A. Sloane. GOSSET: A General-Purpose Program for Designing Experiments. [Online]. Available: www.research.att.com/njas/gosset/

[4] B. Hochwald, T. Marzetta, T. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1962–1973, Sept. 2000.

[5] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," Lucent Technol., Bell Labs, Tech. Rep.BL0112170-980512-07TM, 1998.
[6] K. Hoffman and R. Kunze, *Linear Algebra*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1971.
[7] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need 'perfect side-information' be?," *IEEE Trans. Inform. Theory*, submitted for publication.
[8] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, Jan. 1999.
[9] F. D. Murnaghan, *The Unitary and Rotation Groups, vol. III of Lectures on Applied Mathematics*. Washington, DC: Spartan, 1962.
[10] T. J. Richardson and R. Urbanke, "Some good multiple-antenna codes," Tech. Memo.
[11] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, 2nd ed. Berkeley, CA: Publish or Perish, 1979, vol. IV.
[12] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criteria and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.
[13] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs., Tech. Rep. BL0112170-950615-07TM, 1995.
[14] D. Warrier and U. Madhow, "Noncoherent communication in space and time," *IEEE Trans. Inform. Theory*, submitted for publication.

Multiple Transmit Antenna Differential Detection From Generalized Orthogonal Designs

Hamid Jafarkhani, *Senior Member, IEEE,* and Vahid Tarokh, *Member, IEEE*

***Abstract*—We explicitly construct multiple transmit antenna differential encoding/decoding schemes based on generalized orthogonal designs. These constructions generalize the two transmit antenna differential detection scheme that we proposed before.**

***Index Terms*—Antenna arrays, differential detection, orthogonal designs, space–time block codes, transmitter diversity.**

I. INTRODUCTION

Recently, coding for multiple-antenna transmission in a wireless environment has become an extremely active subject of research. In this direction, it is most practical to assume that the receiver has the knowledge of the channel [1], [3], [9], [11]–[15]. This is a reasonable assumption because the receiver has to estimate the channel for synchronization and carrier recovery purposes.

In rare occasions, it may be assumed that neither the receiver nor the transmitter has the knowledge of the channel. In the case of one transmit antenna, a differential detection scheme exists that neither requires the knowledge of the channel nor employs pilot symbol transmission [8]. This well-known technique motivates the generalization of differential detection for the case of multiple transmit antennas. Hochwald and Marzetta considered this problem in [4] where they proposed unitary space–time codes. These codes are very interesting but have exponential encoding and decoding complexity. In some sense, they are somewhat like Shannon's random codes which indicate the possibility of the existence of structured codes. In a subsequent work, Hochwald *et al.* [5] came up with a second construction that has polynomial encoding but exponential decoding complexity which makes their use formidable in practical situations. At about the same time, we proposed a truly differential coding scheme based on orthogonal designs [10]. This scheme is the first scheme that provides simple encoding/decoding algorithms for differential detection. Followed by our work, Hughes [7] introduced another construction based on group codes. Independently, Hochwald and Sweldens [6] presented a similar construction. Later, Clarkson *et al.* [2] provided a suboptimal decoding algorithm for the schemes in [5] and [6] which does not suffer from an exponential complexity. The complexity of the suboptimal decoding algorithm is polynomial in the number of antennas and the rate and still much more than the decoding complexity of the scheme of [10] which is linear in the number of antennas and the rate.

It was not clear to some members of the research community that the original scheme of [10] can be generalized to more than two transmit antennas. The purpose of this work is to present such generalization and clarify this misconception. In doing so, we employ the theory of space–time block coding introduced in [11].

The outline of the manuscript follows next. In Section II, the system model for transmission using N transmit antennas is considered and space–time block codes are reviewed assuming coherent detection. In Section III, the new differential encoding algorithm is presented. The corresponding decoding algorithm is presented in Section IV. We discuss different versions of the system in Section V. Finally, simulation results and some conclusions are provided in Section VI.

II. SPACE–TIME BLOCK CODING ASSUMING COHERENT DETECTION

In this section, we model a multiple-antenna wireless communication system under the assumption that fading is quasi-static and flat. We consider a wireless communication system where the base station contains N transmit antennas and the decoder contains M receive antennas. At each time slot t, signals $\mathcal{C}_{t,n}$, $n = 1, 2, \ldots, N$ are transmitted simultaneously from the N transmit antennas. The coefficient $\alpha_{n,m}$ is the path gain from transmit antenna n to receive antenna m. We use independent complex Gaussian random variables with variance 0.5 per real dimension to model the path gains. The wireless channel is assumed to be quasi-static so that the path gains are constant over a frame and vary from one frame to another.

Based on our model, the signal $r_{t,m}$ which is received at time t at antenna m is given by

$$r_{t,m} = \sum_{n=1}^{N} \alpha_{n,m} \mathcal{C}_{t,n} + \eta_{t,m} \tag{1}$$

where the noise samples $\eta_{t,m}$ are independent samples of a zero-mean complex Gaussian random variable with variance $1/(2\text{SNR})$ per complex dimension. The average energy of the symbols transmitted from each antenna is normalized to be $1/N$. Therefore, the average power of the received signal at each receive antenna is 1 and the signal-to-noise ratio is SNR.

Manuscript received May 27, 2000; revised February 21, 2001.
H. Jafarkhani was with AT&T Labs-Research, Red Bank, NJ 07701 USA. He is now with Broadcom Corporation, Irvine, CA 92619-7013 USA (e-mail: hamidj@broadcom.com).
V. Tarokh was with AT&T Labs-Research, Florham Park, NJ 07932 USA. He is now with the Department of Electrical and Computer Engineering, the Massachusetts Institute of Technology (MIT), Cambridge, MA 02139 USA (e-mail: vahid@mit.edu).
Communicated by M. L. Honig, Associate Editor for Communications.
Publisher Item Identifier S 0018-9448(01)05459-1.