Existence and Construction of Noncoherent Unitary Space–Time Codes

Vahid Tarokh, *Member, IEEE*, and Il-Min Kim

Dedicated to D. James E. Mazo on the Occasion of His Retirement

***Abstract*—We consider transmission using N transmit and reception using M receive antennas in a wireless environment assuming that neither the transmitter nor the receiver knows the channel coefficients. For the scenario that the transmission employs noncoherent $T \times N$ unitary space–time codes and for a block-fading channel model where the channel is static during T channel uses and varies from T channel uses to the other, we establish the bound $r \leq \min(T-N, N)$ on the diversity advantage rM provided by the code. In order to show that the requirement $r \leq \min(T-N, N)$ cannot be relaxed, for any given $\mathcal{R}$, N, T, and $r \leq \min(T-N, N)$, we then construct unitary $T \times N$ space–time codes of rate $\mathcal{R}$ that guarantee diversity advantage rM. Two constructions are given that are also amenable to simple encoding and noncoherent maximum-likelihood (ML) decoding algorithms.**

***Index Terms*—Diversity, multiple antennas, noncoherent, space–time codes, wireless communications.**

I. Introduction

Recently, various approaches to transmission in a wireless environment using multiple transmit antennas have been proposed. In this direction, the most practical case is the one that the receiver has the knowledge of the channel [1], [2], [9], [11]–[15], because it 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. Under this assumption, when the transmitter uses a single antenna for transmission, both noncoherent and differential detection schemes exist that neither require the knowledge of the channel nor employ pilot symbol transmission [8]. These well-known techniques motivate the generalization of noncoherent and differential detection for the case of multiple transmit antennas.

We summarize the existing results for the scenario when neither the transmitter nor the receiver has the knowledge of the channel coefficients.

- Transmission schemes that approach the problem using differential detection were first proposed in [10] and independently using another approach in [5]. These were then extended in [6], [7].
- Schemes that are based on noncoherent maximum-likelihood (ML) decoding algorithms are known [3], [4]. These schemes are difficult to implement although some suboptimal approaches to decoding these schemes are known.
- To this date, there are no schemes based on ML noncoherent detection for multiple transmit antennas that are amenable to simple ML decoding algorithms.

Manuscript received September 10, 2001; revised August 16, 2002. This work was supported by the National Science Foundation under the Alan T. Waterman Award, Grant CCR-0139398. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The authors were with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139 USA. They are now with the Division of Engineering and Applied Science, Harvard University, Cambridge, MA 02138 USA (e-mail: vahid@deas.harvard.edu; ilmin@deas.harvard.edu).

Communicated by G. Caire, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2002.805075

This motivates us to consider the design of space–time codes which is decodable by the noncoherent decoder and has simple encoding and decoding algorithms. We will refer to the noncoherently decodable space–time codes as noncoherent space–time codes. The outline of our treatise is given next. In Section II, we will consider transmission in a wireless environment using multiple antennas when neither the receiver nor the transmitter knows the channel coefficients. We first recast the ML detection metric of [3] into an elegant geometric form assuming that the transmission scheme employs noncoherent unitary space–time codes. This geometric interpretation indicates that ML detection is equivalent to the closest subspace detection. We will then interpret the diversity of the unitary space–time code as a function of dimensions of subspaces constructed from the columns of the codewords. A geometric approach was also adopted by Zheng and Tse for obtaining the asymtotic capacity of the multiple-antenna channel at high signal-to-noise ratio (SNR) [16]. They interpreted the capacity as sphere packing in the Grassman manifold. Next, we define the dual code of any noncoherent unitary space–time code and it is proven that the diversity of a unitary space–time code is equal to that of its dual. Using this, we establish a bound on diversity of noncoherent unitary space–time codes. In Section III, we will present two constructions of unitary space–time constellations of arbitrary rates that are amenable to simple encoding and decoding algorithms. In particular, the complexity of the decoding of these constellations is independent of their rates and is at most linear in the product MN. One construction is based on a generalization of phase-shift keying (PSK) constellations and the other is based on orthogonal designs. These constructions show that our bounds cannot be improved.

II. The Channel, Transmission, and Detection Model

We consider a wireless communication system where the base station employs N transmit antennas for transmission and the receiver unit employs M receive antennas for reception. At each time slot t, signals $c_{t,i}$, $i = 1, 2, \ldots, N$ are transmitted simultaneously from the N transmit antennas. The coefficient $\alpha_{i,j}$ is the path gain from transmit antenna i to receive antenna j. We use independent complex Gaussian random variables with variance $1/2N$ per real dimension to model the path gains. The wireless channel is assumed to be quasi-static over a block of length T. This means that the path gains are constant over a frame of length T and vary from one frame to another.

Based on our model, $r_{t,j}$, the signal that is received at time t at receive antenna j, is given by

$$r_{t,j} = \sum_{i=1}^{n} \alpha_{i,j} c_{t,i} + \eta_{t,j} \tag{1}$$

where the noise samples $\eta_{t,j}$ are independent samples of a zero-mean complex Gaussian random variable with variance $1/(2\rho)$ per real dimension. The average energy of the symbols transmitted from each antenna is normalized to be one. Therefore, the average power of the received signal at each receive antenna is one and the SNR is SNR $= \rho$. Let H denote the $N \times M$ matrix whose i, jth element is $\alpha_{i,j}$. Let R_j denote the $T \times 1$ column vector whose tth element is $r_{t,j}$ and R denote the $T \times M$ matrix whose jth column is R_j, then

$$R = CH + N \tag{2}$$

where C is the matrix whose t, ith element is $c_{t,i}$ and N is the $T \times M$ vector whose t, jth element is $\eta_{t,j}$.

As in [3], we assume that the transmitted codewords are given by $T \times N$ matrices $\sqrt{T}\Phi_k$, $k = 1, 2, \ldots, L$, where L is the number of codewords and $T \geq N$. It is also assumed that (see [3]) that $\Phi^H\Phi = I_N$, where Φ^H is the Hermitian of Φ and I_N is the $N \times N$ identity matrix i.e., the columns of Φ form a set of orthogonal vectors of unit length. We will refer to the set $\mathcal{C} = \{\sqrt{T}\Phi_1, \sqrt{T}\Phi_2, \ldots, \sqrt{T}\Phi_L\}$ as the unitary space–time code or just the code whenever there is no ambiguity. We will refer to T and N as the *parameters* of the code. The *rate* of the code $\mathcal{C}$ is defined to be $R(\mathcal{C}) = \log_2(L)/T$. Assuming that neither the transmitter nor the receiver knows the channel matrix H, the noncoherent ML detection metric is given by

$$\Phi_{\mathrm{ML}} = \arg \max_{\Phi_\ell \in \{\Phi_1, \ldots, \Phi_L\}} p(R|\Phi_\ell)$$

where $p(R|\Phi_\ell)$ is the conditional probability of R given Φ_ℓ and is given in [3].

By a simple manipulation, the above maximization amounts to that of $\sum_{j=1}^M \|R_j^H \Phi_\ell\|^2$. This is equivalent to a more illuminating ML decoder expression

$$\arg \min_{\Phi_\ell \in \{\Phi_1, \cdots, \Phi_L\}} d^2(R, \mathrm{span}(\Phi_\ell)) \tag{3}$$

where

$$d^2(R, \mathrm{span}(\Phi_\ell)) = \sum_{j=1}^M d^2(R_j, \mathrm{span}(\Phi_\ell))$$

and $d^2(R_j, \mathrm{span}(\Phi_\ell))$ is the Euclidean distance from R_j to the subspace spanned by codeword Φ_ℓ. We will denote $\mathrm{span}(\Phi_\ell)$ by W_{Φ_ℓ}. The expression in (3) shows an important point. The ML decoder metric, previously a subspace projection maximization, can be expressed as a minimum-distance metric. Moreover, to minimize the probability of detection error W_{Φ_ℓ}, $\ell = 1, 2, \ldots, L$ have to be far from each other.

The diversity advantage of the code $\{\Phi_1, \ldots, \Phi_L\}$ has been computed by Hochwald and Marzetta in [3] and has the form of a complicated mathematical expression. At large SNRs, this formula amounts to the minimum number of nonidentity singular values of $\Phi_l^H \Phi_k$ over all $k \neq l$. This, in turn, when viewed in the language of subspace detection, has a simple geometric expression.

Lemma 1: Let U and V denote $T \times N$ matrices such that $U^H U = V^H V = I_N$. Let W_U and W_V, respectively, denote the subspaces of the T-dimensional space generated, respectively, by the columns of U and V. The number of nonidentity singular values of $U^H V$ is equal to $\dim(W_U) - \dim(W_U \cap W_V)$.

Proof: For any arbitrary $T \times N$ matrix U, let $U_1, U_2, \ldots, U_N$ denote the columns of U. Also, for any two vectors v and u, we let $v \cdot u$ denote the inner product of v and u.

Let $k = \dim(W_U \cap W_V)$ and t denote the number of singular values of $U^H V$ that are equal to 1 (counting multiplicities). For any unitary matrices A and B, it can be easily seen that the singular values of $A^H U^H V B$ and $U^H V$ are the same. Clearly, by right multiplication of V by B, we can generate any arbitrary change of coordinates in the column space of B. This means that by choosing A and B carefully, we can assume that the sets $\{(VB)_1, \ldots, (VB)_N\}$ and $\{(UA)_1, \ldots, (UA)_N\}$ have the first k elements in common. By direct computation, it is easy to see that $A^H U^H V B$ (and hence $U^H V$) has k (or more) singular values equal to 1. Thus, the number of unit singular values of $U^H V$ is at least equal to $\dim(W_U \cap W_V)$. Hence $t \geq k$.

Next, let $(c_1, c_2, \ldots, c_N)$ denote an eigenvector of $V^H U U^H V$ corresponding to singular value one of $U^H V$. Then

$$(c_1, c_2, \ldots, c_N) V^H U U^H V = (c_1, c_2, \ldots, c_N)$$

By right multiplying both sides by $(c_1, c_2, \ldots, c_N)^H$ and noticing that

$$\sum_{i=1}^N |c_i|^2 = \left|\sum_{i=1}^N c_i U_i\right|^2$$

we can conclude that

$$\sum_{j=1}^N \left|\left(\sum_{i=1}^N c_i U_i\right) \cdot V_j\right|^2 = \left|\sum_{i=1}^N c_i U_i\right|^2.$$

This means that the length of the projection of the vector $\sum_{i=1}^N c_i U_i$ on the subspace W_V is equal to $\|\sum_{i=1}^N c_i U_i\|$ itself, forcing that $\sum_{i=1}^N c_i U_i \in W_V$. However, $\sum_{i=1}^N c_i U_i \in W_U$, hence $\sum_{i=1}^N c_i U_i \in W_V \cap W_U$. We conclude that the number t of independent eigenvectors of $V^H U U^H V$ corresponding to singular value 1 of $U^H V$ cannot be more than k. Thus, $t \leq k$. Combining this with $t \geq k$, we conclude that $t = k$. □

We have shown that the diversity advantage of the above code is given by rM, where

$$r = \min_{1 \leq k \neq l \leq L} (\dim(W_{\phi_l}) - \dim(W_{\phi_l} \cap W_{\phi_k})). \tag{4}$$

The coding gain of a unitary space–time code at high SNR is computed to be approximately the product $\prod_j (1 - |\lambda_l|^2)$, where λ_j are the nonunit singular values of $\Phi_l^H \Phi_k$. This is hard to compute for design purposes. Code design must assure that $|\lambda_j|$ are small. Otherwise, the subspaces W_{Φ_l} and W_{Φ_k} "almost" collapse on each other in certain directions. This is reminiscent of the expression for probability of pairwise error probability computed by Hochwald and Marzetta [3]. Whenever λ_j is close to unity, the product $(1 - |\lambda_j|^2)\mathrm{SNR}$ is small for all practical SNRs and most of the diversity benefit "corresponding to that singular value" is lost.

To maximize $\prod_j (1 - |\lambda_j|^2)$ is equivalent to maximizing $\sum_j \log(1 - |\lambda_j|^2)$. Using the approximation $\log(1 - x) \simeq -x$, we observe that this is, in turn, equivalent to maximizing $-\sum_j |\lambda_j|^2$. Assuming that there are N nonunit singular values, this is equivalent to maximizing $\sum_{j=1}^N (1 - |\lambda_l|^2)$. The last sum is a measure of distance between the two subspaces W_{Φ_l} and W_{Φ_k} and has an elegant geometric interpretation as described next. If we take an orthonormal basis $w_1, w_2, \ldots, w_N$ for the subspace W_{Φ_l} then

$$\sum_{j=1}^N (1 - |\lambda_j|^2) = \sum_{j=1}^N d^2(w_j, W_{\Phi_k})$$

where $d^2(w_j, W_{\Phi_k})$ is the Euclidean distance of w_j from subspace W_{Φ_k}. We define

$$d^2(W_{\Phi_l}, W_{\Phi_k}) = \sum_{j=1}^N d^2(w_j, W_{\Phi_k}) \tag{5}$$

and refer to it as the *square Euclidean distance* of subspaces W_{Φ_l} and W_{Φ_k}.

Design Criteria for the Unitary Space–Time Codes:

- **Diversity Criteria:** In order to achieve diversity advantage rM, for any two distinct codewords Φ_l and Φ_k

$$\dim(W_{\Phi_l}) - \dim(W_{\Phi_l} \cap W_{\Phi_k}) \geq r.$$

  We will refer to r as the *diversity factor* or just the *diversity* of $\mathcal{C}$ whenever there is no ambiguity.

- **Coding Gain:** If full diversity N is achieved, then the minimum Euclidean distance between subspaces W_{Φ_l} and W_{Φ_k} has to be maximized.

Remark I: In light of the above criteria and the analysis of [3], we can think of N-dimensional subspaces as codewords. In fact, to transmit an N-dimensional subspace W of the T-dimensional complex space $\mathbb{C}^T$, we simply choose an orthonormal basis for W of $T \times 1$ column vectors $\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_N$. We then construct a $T \times N$ matrix Φ from these vectors by letting the kth column of Φ to be $\boldsymbol{c}_k$, $k = 1, 2, \ldots, N$. We then transmit $\sqrt{T}\Phi$ as described before. It is easy to see that $W = W_\Phi$ and that the choice of orthonormal basis for W is immaterial. Thus, we choose to freely mix codewords and subspaces in the sequel, and a code $\{\Phi_1, \Phi_2, \ldots, \Phi_L\}$ will be presented as a set of N-dimensional subspaces $\{W_1, W_2, \ldots, W_L\}$ where $W_k = W_{\Phi_k}$ and *vice versa*.

Definition 2: Let a code consisting of N-dimensional subspaces $\mathcal{C} = \{W_1, W_2, \ldots, W_L\}$ in the T-dimensional complex space be given. Let $\mathcal{C}^\perp = \{W_1^\perp, W_2^\perp, \ldots, W_L^\perp\}$, where $W_k^\perp$ is the orthogonal complement of W_k. We refer to $\mathcal{C}^\perp$ as the dual of $\mathcal{C}$.

We can now prove the following useful lemma.

Lemma 3: Let $\mathcal{C} = \{W_1, W_2, \ldots, W_L\}$ denote a unitary space–time code with parameters T and N. Let $\mathcal{C}^\perp$ denote the dual of $\mathcal{C}$. Then, the diversity factor of $\mathcal{C}$ and $\mathcal{C}^\perp$ are the same.

Proof: Suppose that W_k and W_l denote two distinct codewords. Let $W_k + W_l$ denote the sum of vector spaces W_k and W_l, that is, $W_k + W_l = \{w_k + w_l \mid w_k \in W_k \text{ and } w_l \in W_l\}$.

It is well known that

$$\dim(W_k + W_l) = \dim(W_k) + \dim(W_l) - \dim(W_k \cap W_l)$$

we have

$$\dim(W_k) - \dim(W_k \cap W_l) = \dim(W_k + W_l) - \dim(W_l).$$

It is easy to see that the subspaces $W_k + W_l$ and $W_k^\perp \cap W_l^\perp$ are orthogonal complements of each other. Thus,

$$\dim(W_k + W_l) = T - \dim(W_k^\perp \cap W_l^\perp)$$

and

$$\begin{aligned}\dim(W_k) - \dim(W_k \cap W_l) &= \dim(W_k + W_l) - \dim(W_l)\\ &= T - \dim(W_k^\perp \cap W_l^\perp)\\ &\quad - (T - \dim(W_l^\perp)).\end{aligned}$$

Because $\dim(W_l^\perp) = \dim(W_k^\perp)$, we conclude that

$$\dim(W_k) - \dim(W_k \cap W_l) = \dim(W_k^\perp) - \dim(W_k^\perp \cap W_l^\perp).$$

The result follows by taking minimum over all $1 \le k \ne l \le L$. □

Corollary 4: Let $\mathcal{C}$ denote a unitary space–time code of $T \times N$ matrices (equivalently, of N dimensional subspaces in $\mathbb{C}^T$). The diversity factor r of the code $\mathcal{C}$ is at most $\min(N, T - N)$.

Proof: Clearly, $\dim(W_k) - \dim(W_k \cap W_l) \le N$ for all codewords W_k and W_l. Thus, $r \le N$. Applying the same argument to $\mathcal{C}^\perp$, and by using Lemma 3, we arrive at $r \le T - N$. □

Corollary 5: Let $\mathcal{C}$ denote a unitary space–time code of $T \times N$ matrices (equivalently, of N dimensional subspaces in $\mathbb{C}^T$). Then in order to achieve full diversity N, we must have $T \ge 2N$.

Proof: Suppose that $r = N$. Then $T - N \ge \min(N, T - N) \ge r = N$. Thus, $T \ge 2N$ as desired. □

We will prove that for $T = 2N$, there exist unitary space–time codes that achieve the full spatial diversity $r = N$. This means that the bound of Corollary 5 cannot be improved.

III. Two Constructions of Unitary Space–Time Codes

In this section, for any $T, N, r \le \min(N, T - N)$, and any desired rate $\mathcal{R}$, we will construct unitary space–time codes of rate $\mathcal{R}$ consisting of $T \times N$ codewords having diversity factor r. To this end, we will first present two simple albeit fundamental lemmas.

Lemma 6: Suppose that for any given $\mathcal{R}$ a unitary space–time code with parameters T_0, N, diversity r, and rate $\mathcal{R}$ exists, then for any $T \ge T_0$ and any rate $\mathcal{R}$ a unitary space–time code with parameters T, N, diversity r, and rate $\mathcal{R}$ exists.

Proof: Given $\mathcal{R}$, by assumption a unitary space–time code $\mathcal{C}$ of rate $\mathcal{R}T/T_0$ with parameters T_0, N, and diversity R exists. The elements of $\mathcal{C}$ are N-dimensional subspaces of the T_0-dimensional complex space. Because of the natural embedding of $\mathbb{C}^{T_0}$ in the T-dimensional complex space $\mathbb{C}^T$, whenever $T \ge T_0$, we can view these subspaces as subspaces of $\mathbb{C}^T$. This realization provides us with a code with parameters T, N, diversity r, and rate $\mathcal{R}$. □

Remark II: The reader should notice that our proof is constructive and the constructed code inherits encoding and decoding algorithms of the original code. To encode a sequence of $\mathcal{R}T$ bits, the encoder of $\mathcal{C}$ is applied first to produce a $T_0 \times N$ matrix $\sqrt{T_0}\Phi$. We then append $T - T_0$ rows of all-zero elements to form a matrix Φ^* with the last $T - T_0$ rows having all elements equal to zero. Finally, $\sqrt{T}\Phi^*$ is transmitted.

To decode the $T \times M$-dimensional received matrix R^*, we first construct a $T_0 \times M$-dimensional matrix R by eliminating the last $T - T_0$ rows of R^*. We then decode R using the decoder of $\mathcal{C}$.

Lemma 7: Suppose that given N and T, a unitary space–time code of any desired rate with parameters N, T, and diversity r exists. Then for any rate $\mathcal{R}$ and any $l \ge 0$, a unitary space–time code with parameters $T + l$, $N + l$, diversity r, and rate $\mathcal{R}$ exists.

Proof: By assumption for any given $\mathcal{R}$, a unitary space–time code $\mathcal{C}$ of rate $\mathcal{R}(T + l)/T$ with parameters T, N, and diversity r exists. The elements of $\mathcal{C}$ are subspaces of dimension N in T-dimensional complex space. Let $\mathcal{V}$ denote the orthogonal complement of T-dimensional complex space in $T + l$ dimensional complex space. We consider the code $\mathcal{C}^* = \{\mathcal{V} + W \mid W \in \mathcal{C}\}$. Clearly, the elements of $\mathcal{C}^*$ are $N + l$-dimensional subspaces of $T + l$-dimensional space. Moreover, the diversity factor and the rate of $\mathcal{C}$ are easily seen to be r and $\mathcal{R}$. □

Again the preceding proof is constructive.

Remark III: It is easy to see that $\mathcal{C}^*$ inherits the decoding and encoding algorithms of the original code. To observe this, the encoder for the code $\mathcal{C}^*$ is a modification of the encoder for the code $\mathcal{C}$. Once a subspace W_Φ (a transmission matrix $\sqrt{T}\Phi$) of $\mathcal{C}$ is chosen, first l zeroes are appended to the end of each column of Φ to construct a matrix Φ_*. Let e_k denote the $(T + l) \times 1$ column vector with kth element equal to one and all other elements equal to zero. Then $e_{T+1}, e_{T+2}, \ldots, e_{T+l}$ are appended as new columns to the right of Φ_* to form that $(T + l) \times (N + l)$ matrix Φ^*. Then the matrix $\sqrt{T + l}\Phi^*$ is transmitted.

To decode a received matrix R^*, we first project the received word in the complex space $\mathbb{C}^T$ (i.e., we remove the last l rows of $\mathcal{R}$) and then decode the projected received word to the elements of $\mathcal{C}$.

We can now prove the following fundamental theorem.

Theorem 8: Suppose that for any arbitrary rate $\mathcal{R}$ and any $r \ge 1$, a unitary space–time code of parameters $N = r, T = 2r$, diversity factor r, and rate $\mathcal{R}$ exists. Then, given T_0, N_0 and $r_0 \le \min(N_0, T_0 - N_0)$, unitary space–time codes of parameters N_0, T_0, diversity factor r_0 of any arbitrary rate exists.

Proof: Let $l_0 = N_0 - r_0$. By assumption, unitary space–time codes of arbitrary rate, parameters $N = r_0, T = 2r_0$, diversity factor r_0 exist. By Lemma 7, unitary space–time codes of arbitrary rate, parameters $N_0 = r_0 + l_0, T_1 = 2r_0 + l_0 = N_0 + r_0$, and diversity factor r_0 exist. Because $r_0 \le \min(N_0, T_0 - N_0)$, we have $r_0 \le T_0 - N_0$.

Thus, $T_1 \leq T_0$. The theorem is proved by $T_0 - T_1$ consecutive applications of Lemma 6. □.

Remark IV: By Remarks II and III, application of both Lemmas 6 and 7 preserve the simplicity of encoding and decoding. Thus, the application of Theorem 8 preserves simplicity of encoding and decoding as well.

In light of Remark IV and Theorem 8, the construction of simply encodable/decodable noncoherent unitary space–time codes reduces to the construction of those of parameters $N = r, T = 2r$, with diversity factor r with simple encoding and decoding algorithms.

A. Construction I: Generalized Noncoherent PSK Constellations

Let $r \geq 2$. We will next construct a noncoherent unitary space–time constellation with parameters $N = r$, $T = 2r$, rate $\mathcal{R}$, and diversity r. We consider sets of vectors $\{e_1, e_2, \ldots, e_N\}$ and $\{e_{N+1}, e_{N+2}, \ldots, e_{2N}\}$, where e_i is the T-dimensional column vector whose ith component is 1 and has all other components equal to 0.

We define the subspace $W_k = \operatorname{span}(v_{1,k}, v_{2,k}, \ldots, v_{N,k})$ where

$$v_{j,k} = \cos(\pi k/2^{T\mathcal{R}})e_j + \sin(\pi k/2^{T\mathcal{R}})e_{j+N} \tag{6}$$

for $1 \leq j \leq N$ and $k = 0, 1, 2, \ldots, 2^{T\mathcal{R}} - 1$.

Encoding of $\mathcal{R}$ bits per transmission time is described next. The $T\mathcal{R}$ input bits choose a value $0 \leq k \leq 2^{T\mathcal{R}} - 1$. Then the vectors $v_{1,k}, v_{2,k}, \ldots, v_{N,k}$ define the columns of the matrix Φ_k for the transmission matrix $\sqrt{T}\Phi_k$. The rate of the code is easily seen to be $\mathcal{R} = T\mathcal{R}/T$ bits per transmission time.

Decoding of the above constellation is described next. As before, we let $R_1, R_2, \ldots, R_M$ denote the received vectors, respectively, at receive antenna $1, 2, \ldots, M$. ML decoding amounts to finding the subspace W_k for which $\sum_{j=1}^{M} \|\mathrm{Proj}_{W_k}(R_j)\|^2$ is maximized, where

$$\|\mathrm{Proj}_{W_k}(R_j)\|^2 = \sum_{l=1}^{N} |R_j \cdot v_{l,k}|^2$$

is the square length of projection of the received word R_j at receive antenna j on subspace W_k. This is equivalent to computing

$$\arg\min_k d^2(R, W_k) = \sum_{j=1}^{M} d^2(R_j, W_k)$$

since

$$d^2(R_j, W_k) = \|R_j\|^2 - \sum_{j=1}^{M} \|\mathrm{Proj}_{W_k}(R_j)\|^2.$$

Clearly,

$$\|\mathrm{Proj}_{W_k}(R_j)\|^2 = \sum_{l=1}^{N} |R_j \cdot v_{l,k}|^2$$

thus, we wish to find k which maximizes

$$\sum_{j=1}^{M} \sum_{l=1}^{N} |R_j \cdot v_{l,k}|^2.$$

Replacing $v_{l,k}$ by $\cos(\pi k/2^{T\mathcal{R}})e_l + \sin(\pi k/2^{T\mathcal{R}})e_{l+N}$, we seek to find k which maximizes

$$\sum_{j=1}^{M} \sum_{l=1}^{N} |A_{j,l}\cos(\pi k/2^{T\mathcal{R}}) + B_{j,l}\sin(\pi k/2^{T\mathcal{R}})|^2 \tag{7}$$

where $A_{j,l} = R_j \cdot e_l$ and $B_{j,l} = R_j \cdot e_{l+N}$. By expanding the above sum and using basic trigonometric equations, we observe that the maximizing k for the sum

$$\frac{\sum_{j=1}^{M}\sum_{l=1}^{N}(|A_{j,l}|^2 + |B_{j,l}|^2)}{2} + \frac{\sum_{j=1}^{M}\sum_{l=1}^{N}(|A_{j,l}|^2 - |B_{j,l}|^2)}{2}\cos(2\pi k/2^{T\mathcal{R}}) + \Re\left(\sum_{j=1}^{M}\sum_{l=1}^{N} A_{j,l}B_{j,l}^H\right)\sin(2\pi k/2^{T\mathcal{R}}).$$

is being sought, where $\Re(\cdot)$ denote the real part function.

The sum

$$\frac{\sum_{j=1}^{M}\sum_{l=1}^{N}(|A_{j,l}|^2 + |B_{j,l}|^2)}{2}$$

is independent of k. Thus, we have to compute

$$\arg\max_k \left(\frac{\sum_{j=1}^{M}\sum_{l=1}^{N}(|A_{j,l}|^2 - |B_{j,l}|^2)}{2}\cos\left(\frac{2\pi k}{2^{T\mathcal{R}}}\right) + \Re\left(\sum_{j=1}^{M}\sum_{l=1}^{N} A_j B_j^H\right)\sin\left(\frac{2\pi k}{2^{T\mathcal{R}}}\right)\right). \tag{8}$$

We let

$$A = \sum_{j=1}^{M}\sum_{l=1}^{N}(|A_{j,l}|^2 - |B_{j,l}|^2)$$

and

$$B = \Re(\sum_{j=1}\sum_{l=1}^{N} A_{j,l}B_{j,l}^H)$$

and compute $0 \leq \phi < 2\pi$ such that

$$A = \sqrt{|A|^2 + |B|^2}\cos(\phi) \quad \text{and} \quad B = \sqrt{|A|^2 + |B|^2}\sin(\phi).$$

Using this notation, we observe that k must be chosen to maximize

$$\sqrt{|A|^2 + |B|^2}\cos\left(\phi - \frac{2\pi k}{2^{T\mathcal{R}}}\right). \tag{9}$$

The answer to this maximization problem is clearly $\lfloor \phi 2^{T\mathcal{R}}/2\pi \rfloor$. Noticing that the computation of $R_j \cdot e_1, R_j \cdot e_2, \ldots, R_j \cdot e_T$ is free, we observe that to decode the noncoherent PSK with M receive antennas, at most $3MN + 1$ multiplications, at most $3MN$ additions, one angle computation (application of the $\tan^{-1}(\cdot)$ function), and one division are required. *Thus number of operations is independent of $\mathcal{R}$ and is at most linear in MN.* Clearly, the subspaces W_i and W_j for $i \neq j$ do not have nontrivial intersections, so full diversity is achieved.

We can now summarize our results in the following theorem.

Theorem 9: For any number of transmit antennas N, receive antenna M, coherence time T, rate $\mathcal{R}$, and diversity factor r we have the following.

- A noncoherent unitary space–time code with the above parameters exists only if $r \leq \min(N, T - N)$.
- Conversely, if $r \leq \min(N, T - N)$, then a unitary space–time code $\mathcal{C}$ with the above parameters exists such that encoding of $\mathcal{C}$ is simple (based only on one computation) and decoding complexity of $\mathcal{C}$ is independent of $\mathcal{R}$ and linear in MN.

Proof: The result follows from Theorem 8, Corollary 4, and Construction I. □

B. Construction II: Noncoherent Orthogonal Designs

In this subsection, we will give a second construction of simply encodable, simply ML decodable noncoherent unitary space–time codes. Our construction produces unitary space–time codes of arbitrary rates, parameters $T = 2r$, $N = r$, and diversity r only if $r = 1, 2, 4, 8$.

The construction is based on orthogonal designs. Given $r = 1, 2, 4, 8$, we first consider an $r \times r$ real orthogonal design $\boldsymbol{G}^r$, where we define $\boldsymbol{G}^1 = (x_1)$ to be a 1×1 matrix of indeterminate x_1. The construction of $\boldsymbol{G}^2$, $\boldsymbol{G}^4$ and $\boldsymbol{G}^8$ is due to Radon and Hurwitz and is also presented in [11]. For instance, for $r = 2$, the matrix $\boldsymbol{G}^2 = \left(\begin{smallmatrix} x_1 & x_2 \\ -x_2 & x_1 \end{smallmatrix}\right)$ and for $r = 4$

$$\boldsymbol{G}^4 = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ -x_2 & x_1 & -x_4 & x_3 \\ -x_3 & x_4 & x_1 & -x_2 \\ -x_4 & -x_3 & x_2 & x_1 \end{pmatrix}.$$

For the case that $r = 2$, we also allow the 2×2 complex orthogonal design $\boldsymbol{S}^2 = \left(\begin{smallmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{smallmatrix}\right)$, where s_1^* and s_2^* are, respectively, complex conjugates of s_1 and s_2. The transmission matrix is of the form $\boldsymbol{U}^{2r} = \left(\begin{smallmatrix} I_r \\ \boldsymbol{G}^r \end{smallmatrix}\right)$, for $r = 1, 2, 4, 8$ and $\boldsymbol{V}^4 = \left(\begin{smallmatrix} I_2 \\ \boldsymbol{S}^2 \end{smallmatrix}\right)$. Clearly, the matrices $\boldsymbol{U}^{2r}$, $r = 1, 2, 4, 8$ and $\boldsymbol{V}^4$ are $2r \times r$ matrices. We now describe encoding and decoding using these noncoherent unitary space–time constellations.

Encoding: Encoding requires binary phase-shift keying (BPSK) constellation for the codes $\boldsymbol{U}^{2r}$, $r = 1, 2, 4, 8$ and PSK constellation for the case of $\boldsymbol{V}^4$. Suppose that a constellation $\mathcal{A}$ of size 2^b is given. At the transmitter, rb bits choose r constellation symbols $c_1, c_2, \ldots, c_r$. The variables $x_1, \ldots, x_r$ are then replaced by the constellation symbols $c_1, c_2, \ldots, c_r$ in the transmission matrix $\boldsymbol{U}^{2r}$ (or s_1, s_2 are replaced by c_1, c_2 in the transmission matrix $\boldsymbol{V}^4$). All the elements of the transmission matrix are then scaled by a constant k to form a unitary space–time matrix. Then the elements of the ith row of $\sqrt{T}\boldsymbol{U}^{2r}$ (resp., $\sqrt{4}\boldsymbol{V}^4$) are sent from antennas $1, 2, \ldots, N = r$ simultaneously at times $i = 1, 2, \ldots, T = 2r$, respectively.

Decoding: The decoder has to compute the values $x_1, x_2, \ldots, x_r$ in the signal constellation $\mathcal{A}$ that maximizes $\sum_{j=1}^{M} \sum_{l=1}^{N} \|R_j \cdot \boldsymbol{U}_l^{2r}\|^2$, where $\boldsymbol{U}_l^{2r}$ is the lth column of $2r \times r$ matrix $\boldsymbol{U}^{2r}$. We denote the k, jth entry of R and $\boldsymbol{G}^r$, respectively, by $R_{k,j}$ and $\boldsymbol{G}_{k,j}^r$. We note that $\boldsymbol{G}_{k,j}^r = (\boldsymbol{G}_{k,j}^r)^*$ as $\boldsymbol{G}^r$ is assumed to be a real orthogonal design. The above ML decoding problem amounts to maximization of

$$f(x_1, x_2, \ldots, x_r) = \sum_{j=1}^{M} \sum_{l=1}^{N} \left| R_{l,j} + \sum_{k=1}^{N} R_{k+N,j} \boldsymbol{G}_{k,l}^r \right|^2$$

over $x_1, x_2, \ldots, x_r \in \mathcal{A}$.

Because $\boldsymbol{G}^r$ is an orthogonal design, the above maximization is equivalent to maximization of $\sum_{i=1}^{r} d^2(x_i, p_i)$, where p_i are complex numbers generated only by linear processing of elements of R at most N^2M multiplications and N^2M additions [11]. In this light, to perform ML detection, the farthest points of $\mathcal{A}$ from p_i, $i = 1, 2, \ldots, r$ have to be computed. This requires an additional $N = r$ comparisons. We hence observe that ML decoding in this case is also independent of the transmission rate and requires at most N^2M multiplications, N^2M additions, and N comparisons.

Example: To illustrate the decoding described above, we consider the code $\boldsymbol{V}^4$ which has the format

$$\boldsymbol{V}^4 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ s_1 & s_2 \\ -s_2^* & s_1^* \end{pmatrix}.$$

We also consider $M = 1$ receive antennas and the matrix

$$R = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{pmatrix}.$$

ML detection is equivalent to maximization of

$$f(s_1, s_2) = |r_1 + r_3 s_1^* - r_4 s_2|^2 + |r_2 + r_3 s_2^* + r_4 s_1|^2$$

over $s_1, s_2 \in \mathcal{A}$. Expanding the above, we observe that preceding maximization is equivalent to maximization of

$$|r_1^* r_3 + r_2 r_4^* + s_1|^2 + |r_2^* r_3 - r_1 r_4^* + s_2|^2$$

over all $s_1, s_2 \in \mathcal{A}$. Letting

$$p_1 = -(r_1^* r_3 + r_2 r_4^*) \quad \text{and} \quad p_2 = -(r_2^* r_3 - r_1 r_4^*)$$

the above maximization is equivalent to maximizing $\sum_{i=1}^{2} d^2(s_i, p_i)$ and, in turn, equivalent to finding the farthest points of $\mathcal{A}$ from, respectively, p_1 and p_2.

Acknowledgment

The authors sincerely thank Dr. B. Hochwald and Dr. T. L. Marzetta for many valuable discussions.

References

[1] S. M. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[2] J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," in *Proc. IEEE Vehicular Technology Conf (VTC'96)*, pp. 136–140.

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

[4] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space–time constellation," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2335–2367, Sept. 2000.

[5] B. M. Hochwald and W. Sweldens, "Differential unitary space–time modulation," *IEEE Trans. Commun.*, vol. 48, pp. 2041–2052, Dec. 2000.

[6] B. L. Hughes, "Differential space–time modulation," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2567–2578, Nov. 2000.

[7] H. Jafarkhani and V. Tarokh, "Multiple antenna differential detection from generalized orthogonal designs," *IEEE Trans. Inform. Theory*, to be published.

[8] J. G. Proakis, *Digital Communications*, 2nd ed. New York: McGraw-Hill, 1989.

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

[10] V. Tarok and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 1169–1174, July 2000.

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

[12] ——, "Space–time block codes for high data rate wireless communication: Performance results," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 451–460, Mar. 1999.

[13] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space–time codes for high data rates wireless communications: Performance criteria in the presence of channel estimation errors, mobility and multiple paths," *IEEE Trans. Commun.*, vol. 47, pp. 199–207, Feb. 1999.

[14] A. Wittneben, "Base station modulation diversity for digital SIMULCAST," in *Proc. IEEE Vehicular Technology Conf. (VTC'91)*, May 1991, pp. 848–853.

[15] ——, "A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation," in *Proc. IEEE Int. Communications Conf. (ICC'93)*, 1993, pp. 1630–1634.

[16] L. Zheng and D. N. C. Tse, "Communication on the Granssmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359–383, Feb. 2002.

Capacity of Multiple-Transmit Multiple-Receive Antenna Architectures

Angel Lozano, *Senior Member, IEEE*, and
Antonia Maria Tulino, *Member, IEEE*

***Abstract*—The capacity of wireless communication architectures equipped with multiple transmit and receive antennas and impaired by both noise and cochannel interference is studied. We find a closed-form solution for the capacity in the limit of a large number of antennas. This asymptotic solution, which is a sole function of the relative number of transmit and receive antennas and the signal-to-noise and signal-to-interference ratios (SNR and SIR), is then particularized to a number of cases of interest. By verifying that antenna diversity can substitute for time and/or frequency diversity at providing ergodicity, we show that these asymptotic solutions approximate the ergodic capacity very closely even when the number of antennas is very small.**

***Index Terms*—Adaptive antennas, antenna arrays, asymptotic analysis, channel capacity, diversity, fading channels, multiantenna communication, multiuser detection.**

I. Introduction

With the explosive growth of both the wireless industry and the Internet, the demand for mobile data access is expected to increase dramatically in the near future. As a result, the ability to support higher capacities will be paramount. Capacity can be pushed by exploiting the space dimension inherent to any wireless communication system. Nonetheless, due to economical and environmental aspects, it is highly desirable not to increase the density of base stations. Under such constraint, antenna arrays are the tools that enable spatial processing on a per-base-station basis. Recognizing this potential, the use of arrays at base-station sites is becoming universal. Array-equipped terminals, on the other hand, had not been contemplated in the past because of size and cost considerations. However, recent results in information theory have shown that, with the simultaneous use of multiple transmit and receive antennas, very large capacity increases can be unleashed [1]–[4]. At the same time, it is reasonable to expect that terminals supportive of progressively higher data rates will tend to be naturally larger in size and, consequently, they will be able to accommodate multiple closely spaced antennas. Hence, the deployment of arrays at both base stations and terminals appears as an attractive scenario for the evolution of mobile data access.

Great progress has been made toward understanding the information-theoretical capacity and the performance of multiple-antenna architectures with thermal noise as the only impairment (see [5]–[16]). Within the context of a wireless system, however, the dominant impairment is typically not thermal noise, but rather cochannel interference. Thus, the objective of the present work is to extend this understanding to the realm of spatially colored interference. We invoke, as central tool, recent results on the asymptotic distribution of the singular values of random matrices and their application to randomly spread code-division multiple access (CDMA) [17]–[19]. Although these distributions pertain asymptotically in the number of antennas, the results we derive therefrom become virtually universal under ergodic conditions.

Since the focus is on mobile systems, we consider only "open-loop" architectures wherein the transmitter does not have access to the instantaneous state of the channel. Only large-scale information—defined as information that varies slowly with respect to the fading rate—is available to the transmitter.

This correspondence is organized as follows. In Section II, the metrics and models are introduced. In Section III, the noise-limited capacity is reviewed using the tools of asymptotic analysis. Such analysis is generalized, in Section IV, to environments containing spatially colored interference. The main result therein is an expression of the asymptotic capacity in the presence of both noise and interference. Finally, Section V concludes the correspondence.

II. Definitions and Models

A. Propagation Model

With M transmit and N receive antennas, the channel responses from every transmit antenna to every receive antenna can be assembled into an $N \times M$ random matrix $\mathbf{G}$ whose underlying random process is presumed zero-mean and ergodic. The propagation scenario and the spatial arrangement of the antennas determine the correlation among the entries of $\mathbf{G}$. The scenario we consider, typical of a mobile system, is based on the existence of an area of local scattering around each terminal. Accordingly, the power angular spread is expected to be very large—possibly as large as 360°—at the terminals rendering the antennas therein basically uncorrelated. At the base station, the angular spread tends to be small [20], [21] but the antennas can be also decorrelated by spacing them sufficiently apart [22]–[25]. Consequently, we focus on channel matrices containing only independent entries.

Since the elements of $\mathbf{G}$ are identically distributed, it is possible to define a normalized channel matrix $\mathbf{H}$ with unit-variance entries such that $\mathbf{G} = \sqrt{g}\,\mathbf{H}$.

B. "Open-Loop" Capacity

Perfect channel estimation at the receiver [26]–[28] is presumed.[1] The impairment comprises additive white Gaussian noise (AWGN) as

Manuscript received May 3, 2001; revised April 8, 2002.

A. Lozano is with the Wireless Communications Research Department, Bell Laboratories, Lucent Technologies, Holmdel, NJ 07733 USA (e-mail: aloz@lucent.com).

A. M. Tulino is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08540 USA (e-mail: atulino@princeton.edu).

Communicated by D. N. C. Tse, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2002.805084

[1]The penalty associated with channel estimation is small as long as the coherence time of the channel—measured in symbols—is large enough with respect to the number of transmit antennas [27].