Differential Space–Time Modulation

Brian L. Hughes, *Member, IEEE*

***Abstract*—Space–time coding and modulation exploit the presence of multiple transmit antennas to improve performance on multipath radio channels. Thus far, most work on space–time coding has assumed that perfect channel estimates are available at the receiver. In certain situations, however, it may be difficult or costly to estimate the channel accurately, in which case it is natural to consider the design of modulation techniques that do not require channel estimates at the transmitter or receiver.**

We propose a general approach to differential modulation for multiple transmit antennas based on group codes. This approach can be applied to any number of transmit and receive antennas, and any signal constellation. We also derive low-complexity differential receivers, error bounds, and modulator design criteria, which we use to construct optimal differential modulation schemes for two transmit antennas. These schemes can be demodulated with or without channel estimates. This permits the receiver to exploit channel estimates when they are available. Performance degrades by approximately 3 dB when estimates are not available.

***Index Terms*—Differential modulation, group codes, multipath channels, noncoherent communication, space–time coding, transmit diversity.**

I. Introduction

ONE of the goals of third- and fourth-generation cellular systems is to provide broadband data access to highly mobile users. Real-time multimedia services, such as videoconferencing, can require data rates on the order of 2–20 Mb/s. However, the data modes of existing cellular standards, such as IS-136 and GSM, currently support rates two to three orders of magnitude smaller [2]. In order to meet this goal, it is important to develop new wireless communication methods that achieve a higher spectral efficiency (data rate per unit bandwidth) for a given power expenditure.

On multipath radio channels, the tradeoff between spectral efficiency and power consumption can be dramatically improved by deploying multiple antennas at the transmitter and/or receiver [7], [8], [20], [21], [32]. For example, using t antennas at both the transmitter and receiver can increase spectral efficiency by a factor of more than t over comparable single-antenna systems [7]. Space–time coding and modulation strategies, which exploit the presence of multiple transmit antennas, have recently been adopted in third-generation cellular standards (e.g., CDMA 2000 [34] and wideband CDMA [33], [15]), and have also been proposed for wireless local loop (Lucent's BLAST project [38]) and wide-area packet data access (AT&T's Advanced Cellular Internet Service [2]).

Thus far, most research on space–time coding has assumed that perfect estimates of current channel fading conditions are available at the receiver. This is reasonable when the channel changes slowly compared with the symbol rate, since the transmitter can send training symbols (or a pilot tone) which enable the receiver to estimate the channel accurately. Specific codes designed for this situation include the transmit diversity schemes in [9], [10], [23], [36], [37], the layered architecture in [7], [38], the trellis codes in [28], and the block codes in [1], [30].

In some situations, however, we may want to forego channel estimation in order to reduce the cost and complexity of the handset, or perhaps fading conditions change so rapidly that channel estimation is difficult or requires too many training symbols. For example, in frequency-hopping systems, fading conditions may change significantly from one hop to the next; in time-division systems, the channel may change between two successive frames. Channel estimation may also be difficult in high-mobility situations. Consider a vehicle transmitting at a symbol rate of 30 kHz and a frequency of 1.9 GHz. If the vehicle moves at 60 mi/h, the coherence time is on the order of 50–100 symbols [12]. If multiple antennas are used, the path gains between each pair of transmit and receive antennas must be estimated. Thus if five training symbols were used per antenna pair, a system with four transmit and one receive antenna would require 20 training symbols—a significant overhead. Third-generation European cellular standards are required to operate on trains moving up to 500 km/h [5], [12]. At this speed, the coherence time in this example is less than 20 symbols, in which case it is not clear whether accurate channel estimation is possible.

For such situations, it is useful to develop modulation techniques that do not require channel estimates at the transmitter or receiver. For a single transmit antenna, frequency-shift keying (FSK) and differential phase-shift keying (DPSK) can be demodulated without the use of channel estimates or training symbols. It is natural to consider extensions of these schemes to multiple transmit antennas. Motivated by the information-theoretic arguments in [16], Hochwald and Marzetta have proposed the use of unitary space–time block codes, in which the signals transmitted by different antennas are mutually orthogonal. Optimal receivers, error bounds, and design criteria for unitary codes were derived in [11], and some specific code constructions were given in [12]. More recently, Tarokh and Jafarkhani [31] have proposed differential transmit diversity schemes for two antennas. Like FSK and DPSK, all of the schemes in [12] and [31] can be demodulated without channel estimates at the receiver.

Manuscript received August 1, 1999; revised March 1, 2000. This work was supported in part by the National Science Foundation under Grant CCR-9903107, and by the Center for Advanced Computing and Communication. The material in this paper was presented in part at the IEEE Wireless Communications and Networking Conference, New Orleans, LA, September 27–30, 1999 and at the 33rd Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, October 24–27, 1999.

B. L. Hughes is with the Center for Advanced Computing and Communication, Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695-7914 (e-mail: blhughes@eos.ncsu.edu).

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

Publisher Item Identifier S 0018-9448(00)09649-8.

In this paper, we propose a new and general approach to differential modulation for multiple transmit antennas based on group codes. This approach can be applied to any number of transmit and receive antennas, and any signal constellation. We also derive low-complexity differential receivers, error bounds, and modulator design criteria for the case where the number of transmit antennas equals the block length of the group code. We then use the design criteria to construct optimal differential modulation schemes for two transmit antennas. These schemes can be demodulated with or without channel estimates. This permits the receiver to exploit channel estimates when they are available. Performance degrades by approximately 3 dB when estimates are not available. When channel estimates are available, the group codes derived in this paper can also be used as space–time block codes, as in [1], [30].

While this paper was under review, we learned of independent work by Hochwald and Sweldens [13] which proposes a similar approach to differential space–time modulation. Although there are differences in the proposed receivers and the generality of the formulation, the differential encoding method and modulator design criteria in [13] are essentially the same as ours. However, the derivation of optimal modulation schemes for two transmit antennas is unique to this paper.

The rest of the paper is organized as follows. In Section II, we introduce the channel model and provide some necessary background on space–time coding with and without channel estimates at the receiver. In Section III, we introduce our approach to differential modulation for multiple transmit antennas, and derive low-complexity receivers, error bounds, and design criteria. Finally, optimal modulation schemes for two transmit antennas are given in Section IV, and our main conclusions are summarized in Section V.

II. Preliminaries

A. Channel Model

Consider a wireless channel in which data are sent from t transmit antennas to r receive antennas [7], [8], [28], [32]. At the transmitter, data are encoded using t parallel encoders, one for each transmit antenna. The resulting encoded symbols are mapped into a unit-energy constellation $\mathcal{C}$ and modulated onto a pulse waveform of duration T for transmission over the channel. Let $c_{jk} \in \mathcal{C}$ denote the constellation point selected by the encoder of transmit antenna $j = 1, \ldots, t$ at time $k = 1, \ldots, n$.

The signal that arrives at each of the r receive antennas is a superposition of the t fading transmitted signals and noise, as illustrated in Fig. 1. We assume that the delay spread of the multipath is small and that the receiver has obtained symbol, but not phase, synchronization. Moreover, we assume that nT is small compared with the channel coherence time, so that fading conditions can be considered constant over n symbols. At each receive antenna, a demodulator synchronously samples the output of a filter matched to the pulse waveform, thereby producing r decision statistics in each symbol interval. Under these conditions, the relationship between the decision statistics and the transmitted signals is given by

$$y_{ik} = \sum_{j=1}^{t} h_{ij} c_{jk} \sqrt{\rho_t} + n_{ik}, \qquad i = 1, \ldots, r, \;\; k = 1, \ldots, n$$

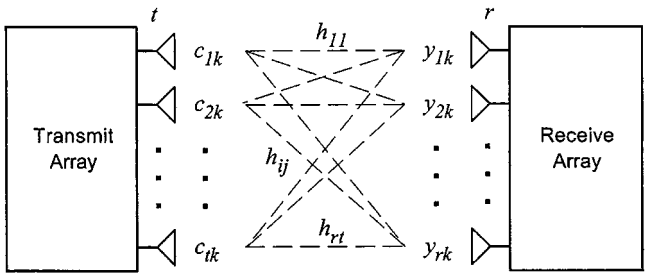

Fig. 1. A flat-fading channel.

where h_{ij} is the complex fading path gain from transmit antenna j to receive antenna i and n_{ik} is a noise variable. Here $\rho_t = \rho/t$ where ρ is the signal-to-noise ratio (SNR) per receive antenna. We assume that the elements in the transmit and receive arrays are spaced so as to produce independent fading between each pair of transmit and receive antennas. The path gains h_{ij} and noise variables n_{ik} are therefore independent and identically distributed, complex Gaussian random variables with probability density function (pdf)

$$p(h) = (1/\pi)\exp(-|h|^2), \qquad h \in \mathbb{C}.$$

Defining the *code matrix* by

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{t1} & c_{t2} & \cdots & c_{tn} \end{bmatrix}$$

we can recast this channel in an equivalent matrix form

$$Y = \sqrt{\rho_t} HC + N \tag{1}$$

where $Y = \{y_{ik}\}$ is the $r \times n$ receive matrix, $H = \{h_{ij}\}$ is the $r \times t$ fading matrix, and $N = \{n_{ik}\}$ is the $r \times n$ noise matrix.

We distinguish between two communication situations for this channel. We say the receiver has perfect *channel state information* (CSI), if the receiver (but not the transmitter) has a perfect estimate of the fading matrix H. If neither the transmitter nor the receiver know the outcome of H, we say there is *no CSI*. In this paper, we are primarily interested in methods for transmitting data without CSI. In order to show why these methods work, however, we make use of results on communication with perfect CSI, which are summarized in the next section.

B. Perfect CSI at the Receiver

Most work on space–time coding has assumed that perfect CSI is available at the receiver. We now summarize results on optimal receivers, error bounds, and design criteria for this situation from [9], [28].

A space–time code for the constellation $\mathcal{C}$ consists of a collection of code matrices $C_m, m = 1, \ldots, M$, where $C_m \in \mathcal{C}^{t \times n}$. When H is known at the receiver, the pdf of the received matrix given that C_m is transmitted is

$$p(Y \mid H, C_m) = \frac{\exp(-\mathrm{Tr}\{(Y - \sqrt{\rho_t} HC_m)(Y - \sqrt{\rho_t} HC_m)^\dagger\})}{\pi^{nr}}$$

where "Tr" is the trace and $\dagger$ denotes the conjugate transpose. If the code matrices are equally likely, the optimal receiver is the maximum-likelihood (ML) detector ([18, p. 72]), which reduces to the minimum Euclidean distance detector

$$\begin{aligned}\hat{m} &= \arg\max_m p(Y \mid H, C_m) \\ &= \arg\min_m \operatorname{Tr}\{(Y-\sqrt{\rho_t}HC_m)(Y-\sqrt{\rho_t}HC_m)^\dagger\}. \end{aligned} \quad (2)$$

Here "arg" denotes any argument that achieves the maximum (or minimum). Let $\Pr\{C_0 \to C_1\}$ be the pairwise error probability of this receiver, i.e., the probability of incorrectly decoding C_0 as C_1, in a code consisting of only these two matrices. The Chernoff bound on this error probability takes the form ([28, eq. (9)])

$$\Pr\{C_0 \to C_1\} \le \frac{1}{|I+(\rho_t/4)(C_0-C_1)(C_0-C_1)^\dagger|^r} \quad (3)$$

where I is the identity matrix and $|\cdot|$ denotes the determinant. For large ρ_t, this bound behaves as $(\Lambda_p\rho_t/4)^{-r\nu_p}$ where ν_p and Λ_p depend on the difference matrix C_0-C_1. The parameter $\nu_p=\nu_p(C_0,C_1)$ is equal to the rank of C_0-C_1 and can be interpreted as the *diversity advantage* of the code pair [9]. The maximum diversity is therefore $\nu_p(C_0,C_1)=t$, provided $n\ge t$. The quantity Λ_p can be interpreted as the *coding advantage*, and is given by

$$\Lambda_p(C_0,C_1) = |(C_0-C_1)(C_0-C_1)^\dagger|_+^{1/\nu_p} \quad (4)$$

where $|A|_+$ denotes the product of the nonzero eigenvalues of A, including multiplicities. This is clearly a matrix analog of the *product distance* [4], [35], which arises in single-antenna fading channels. When $\nu_p(C_0,C_1)=t$, note that the product distance (4) reduces to

$$\Lambda_p^*(C_0,C_1) = |(C_0-C_1)(C_0-C_1)^\dagger|^{1/t}. \quad (5)$$

Also, note that $\Lambda_p^*(C_0,C_1)>0$ if and only if $\nu_p(C_0,C_1)=t$.

For large ρ_t, the performance of any space–time code $C_m, m=1,\dots,M$ is determined primarily by the minimum diversity

$$\nu_p \triangleq \min_{m\ne m'} \nu_p(C_m,C_{m'})$$

and to a lesser extent by the minimum coding advantage,

$$\Lambda_p \triangleq \min_{m\ne m':\nu_p(C_m,C_{m'})=\nu_p} \Lambda_p(C_m,C_{m'})$$

If we are interested only in codes with $\nu_p=t$, however, note that we can simply use the single-performance criterion

$$\Lambda_p^* \triangleq \min_{m\ne m'} \Lambda_p^*(C_m,C_{m'})$$

which is positive only if $\nu_p=t$, in which case $\Lambda_p^*=\Lambda_p$.

For example, consider the transmit diversity scheme proposed by Alamouti in [1], in which $t=2$ antennas are used to send two symbols $a,b\in\mathcal{C}$ by transmitting the code matrix

$$C_{a,b} = \begin{bmatrix} a & -b^* \\ b & a^* \end{bmatrix}. \quad (6)$$

For the unit-energy quaternary-phase shift keying (QPSK) constellation $\mathcal{C}=\{1,-1,j,-j\}$, it is easy to verify that the 16 code matrices in this scheme have minimum distance $\Lambda_p^*=2>0$. Therefore, the diversity is $\nu_p=t=2$ and the minimum product distance is $\Lambda_p=\Lambda_p^*=2$.

C. No CSI at the Receiver

In the absence of CSI at the receiver, Hochwald and Marzetta [11] have argued heuristically that the capacity of the multi-antenna channel (1) can be approached for large n or ρ by code matrices with equal-energy, orthogonal rows. Accordingly, they focused attention on codes with the property

$$C_mC_m^\dagger = nI, \qquad \text{for all } m=1,\dots,M \quad (7)$$

which they called *unitary space–time codes*. In this section, we summarize results on optimal receivers, error bounds, and design criteria for unitary codes from [11]. We present this work in a different form than [11], however, in order to more clearly relate it to the results of the previous section. At first glance, it may appear that these results should also follow as a special case ($\mu=0$) of those in [29]; however, the channel is modeled as memoryless in ([29, eq. (2)]), which is inconsistent with our assumption that h_{ij} is fixed for $k=1,\dots,n$.

When C_m is transmitted and H is unknown, the received matrix Y in (1) is Gaussian with conditional pdf

$$p(Y\mid C_m) = \frac{\exp\left(-\operatorname{Tr}\left\{Y\Sigma_m^{-1}Y^\dagger\right\}\right)}{|\pi\Sigma_m|^r}$$

where $\Sigma_m = I+\rho_tC_m^\dagger C_m$. Note that the matrix identity $|I+AB|=|I+BA|$ and the unitary property (7) imply that $|\Sigma_m|$ does not depend on m. Further note that

$$\Sigma_m^{-1} = I - \frac{\rho_t}{n\rho_t+1}C_m^\dagger C_m$$

which follows from the identity

$$(A+BCD)^{-1} = A^{-1}-A^{-1}B(C^{-1}+DA^{-1}B)^{-1}DA^{-1}.$$

Given these results, the ML detector for a unitary code reduces to a quadratic receiver

$$\begin{aligned}\hat{m} &= \arg\max_m p(Y\mid C_m) \\ &= \arg\max_m \operatorname{Tr}\{YC_m^\dagger C_mY^\dagger\}. \end{aligned} \quad (8)$$

A Chernoff bound on the pairwise error probability of this receiver for unitary codes was derived in [11, eq. (18)]. We can rewrite this bound in a compact matrix form as[1]

$$\Pr\{C_0\to C_1\} \le \frac{1}{\left|I+\frac{\rho_t^2n^2}{4(1+\rho_tn)}[I-(1/n^2)C_1C_0^\dagger C_0C_1^\dagger]\right|^r}. \quad (9)$$

As in the previous section, we can extract useful insights on code design by examining the asymptotics of this bound. To the best of our knowledge, the following observations are new, unless otherwise indicated. For large ρ_t, the bound in (9) behaves as $(\Lambda_a\rho_t/4)^{-r\nu_a}$, where ν_a and Λ_a now depend on the cross-product matrix $(1/n)C_0C_1^\dagger$. The diversity advantage

[1]As shown in [11], the bounds in (3) and (9) can both be sharpened by a factor of two, omitted here for simplicity.

$\nu_a = \nu_a(C_0, C_1)$ is equal to the rank $I - (1/n^2)C_1C_0^\dagger C_0C_1^\dagger$. Hochwald and Marzetta [11] have observed that the maximum diversity is $\nu_a(C_0, C_1) = t$, which is achieved when 1 is not a singular value of $(1/n)C_0C_1^\dagger$. Observing that

$$\left|\begin{bmatrix} C_0 \\ C_1 \end{bmatrix} [C_0^\dagger C_1^\dagger]\right| = |n^2 I - C_1C_0^\dagger C_0C_1^\dagger| \tag{10}$$

we see that $\nu_a(C_0, C_1) = t$ if and only if the rows of C_0 and C_1 are linearly independent, which is possible only if $n \geq 2t$.

The coding advantage Λ_a is given by a quantity analogous to the product distance (4)

$$\Lambda_a(C_0, C_1) = |nI - (1/n)C_1C_0^\dagger C_0C_1^\dagger|_+^{1/\nu_a}. \tag{11}$$

When $t = 1$ and the vectors C_0 and C_1 are real, this quantity reduces to $n\sin^2\theta$, where θ is the angle between C_0 and C_1. We therefore propose to call this quantity the *angular distance* between C_0 and C_1. Angular distance provides a design criterion for space–time coding without CSI, which is analogous to the product distance (4) for perfect CSI. From the identity $|I + AB| = |I + BA|$, we see that $\Lambda_a(C_0, C_1)$ is symmetric in C_0 and C_1. For $\nu_a(C_0, C_1) = t$, the angular distance reduces to

$$\begin{aligned}\Lambda_a^*(C_0, C_1) &= |nI - (1/n)C_1C_0^\dagger C_0C_1^\dagger|^{1/t} \\ &= (1/n)\left|\begin{bmatrix} C_0 \\ C_1 \end{bmatrix} [C_0^\dagger C_1^\dagger]\right|^{1/t}.\end{aligned}$$

For large ρ_t, the performance of the code $C_m, m = 1, \ldots, M$ with receiver (8) is determined mainly by

$$\nu_a \triangleq \min_{m \neq m'} \nu_a(C_m, C_{m'})$$

and

$$\Lambda_a \triangleq \min_{m \neq m' : \nu_a(C_m, C_{m'}) = \nu_a} \Lambda_a(C_m, C_{m'}).$$

Once again, if we are interested only in codes with $\nu_a = t$, we can use the single-performance criterion

$$\Lambda_a^* \triangleq \min_{m \neq m'} \Lambda_a^*(C_m, C_{m'})$$

which is positive only if $\nu_a = t$, in which case $\Lambda_a^* = \Lambda_a$.

As an example, consider the performance of the code (6) in the absence of CSI. If $\mathcal{C}$ is any unit-energy phase-shift keying (PSK) constellation, then $C_{a,b}^\dagger C_{a,b} = C_{a,b}C_{a,b}^\dagger = 2I$ for all $a, b \in \mathcal{C}$. Thus the code is unitary and the results above apply. For any code matrices C_0 and C_1, these identities also imply $I - (1/n^2)C_1C_0^\dagger C_0C_1^\dagger = O$, where O is the all-zero matrix. We conclude that $\nu_a = 0$, and hence the code is essentially useless in the absence of CSI.

III. Differential Space–Time Modulation

For a single transmit antenna, one of the simplest and most effective noncoherent modulation techniques is DPSK. Differentially encoded PSK can be demodulated coherently or noncoherently. Moreover, the noncoherent receiver has a simple form and performs within 3 dB of the coherent receiver on Rayleigh fading channels ([19, p. 774]). It is natural to consider extensions of this technique to multiantenna channels.

Recently, Tarokh and Jafarkhani [31] have proposed a differential modulation scheme for $t = 2$ transmit antennas based on Alamouti's code (6). This scheme shares many of the desirable properties of DPSK: it can be demodulated with or without CSI at the receiver, achieves full diversity in both cases, and there exists a simple noncoherent receiver that performs within 3 dB of the coherent receiver. However, the scheme also has some limitations. First, the encoding procedure significantly expands the signal constellation for nonbinary signaling (e.g., from QPSK to 9QAM). Second, the approach does not seem to extend to complex constellations for $t > 2$, or real constellations for $t > 8$, without a penalty in rate. As noted in [31], the scheme relies on the fact that (6) is a complex orthogonal block design, and such designs do not exist for $t > 2$ [30]. (Real orthogonal designs for $t = 4$ and 8 were derived in [30], and these can be used to construct BPSK modulators for up to eight transmit antennas [31].)

In this section, we present a new approach to differential modulation for multiple-transmit antennas based on group codes. This approach can be applied to any number of antennas and any constellation. The group structure greatly simplifies the analysis of these schemes, and may also lead to simpler and more transparent modulation and demodulation procedures.

A. Unitary Group Codes

Our approach to differential modulation is based on a new class of space–time block codes which possess a group structure. Consider a system with t transmit antennas and constellation $\mathcal{C}$. For any $n \geq t$, let $\mathcal{G}$ be any group of $n \times n$ unitary matrices ($G^\dagger G = GG^\dagger = I$ for all $G \in \mathcal{G}$), and let D be a $t \times n$ matrix such that $DG \in \mathcal{C}^{t \times n}$ for all $G \in \mathcal{G}$. We call the collection of matrices

$$D\mathcal{G} \triangleq \{DG : G \in \mathcal{G}\} \tag{12}$$

a *(multichannel) group code* of length n over the constellation $\mathcal{C}$. The *rate* of this code is given by $R = (1/n)\log_2|\mathcal{G}|$ b/s/Hz, where $|\mathcal{G}|$ denotes the cardinality of $\mathcal{G}$.

Multichannel group codes are a generalization of Slepian's group codes [26] to multiple antennas and complex constellations. For $t = 1$ and groups of real orthogonal matrices, Slepian considered the use of such codes on the Gaussian channel and showed that they possess a high degree of symmetry: each codeword has the same error probability, the ML decoding regions are all congruent, and the set of (Euclidean) distances from a codeword to all of its neighbors is the same for each codeword. In [6], Forney introduced *geometrically uniform codes*, which extend Slepian's idea to groups of arbitrary isometries with respect to Euclidean distance. Since the results in [6], [26] are rooted in Euclidean distance, however, they do not apply directly to fading channels like (1). For the purposes of this paper, however, all that we require is the group structure.

Example 1: For $t = n = 1$, M-ary PSK is a group code with $\mathcal{G} = \{1, \omega_M, \omega_M^2, \ldots, \omega_M^{M-1}\}$ and $D = 1$, where $\omega_M \triangleq \exp(2\pi j/M)$.

Example 2: For $t = 1$ and $n = M$, M-ary *pulse-position modulation* is a group code over $\mathcal{C} = \{\sqrt{M}, 0\}$, with $D = [\sqrt{M}, 0, \ldots, 0]$ and $\mathcal{G} = \{I, \Omega, \Omega^2, \ldots, \Omega^{M-1}\}$, where Ω is the $M \times M$ right-shift matrix

$$\Omega \triangleq \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Example 3: For $t = n = 2$, the pair

$$\mathcal{G} = \left\{ \pm \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \pm \begin{bmatrix} j & 0 \\ 0 & -j \end{bmatrix}, \pm \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \pm \begin{bmatrix} 0 & j \\ j & 0 \end{bmatrix} \right\}$$

$$D = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

is a group code over the QPSK constellation $\mathcal{C} = \{1, j, -1, -j\}$.

The class of group codes is apparently very rich, and includes polyphase codes [39], permutation codes [25], codes from reflection groups [17], all binary linear codes with BPSK modulation [6], [24], and block-circulant unitary codes [12]. From this, it is clear that group codes can be constructed for any number of transmit antennas and any constellation $\mathcal{C}$. We can always choose D to be a $t \times n$ matrix in $\mathcal{C}$ and let $\mathcal{G}$ be any group of $n \times n$ permutation matrices. While permutation groups can always be used, most complex constellations have symmetry properties which permit the use of a wider variety of unitary matrix groups.

The core idea of this paper is that group codes can be differentially encoded in a way similar to PSK. For simplicity, let us consider $\mathcal{G}$ to be the set of possible messages. To initialize transmission, the transmitter sends $X_0 = D$. Thereafter, messages are differentially encoded: to send $G_k \in \mathcal{G}$ in block k the transmitter sends

$$X_k = X_{k-1} G_k, \qquad k = 1, \ldots, K. \tag{13}$$

The group structure ensures that $X_k \in D\mathcal{G}$ whenever $X_{k-1} \in D\mathcal{G}$. Moreover, the rate of the code is essentially

$$R = (1/n) \log_2 |\mathcal{G}|$$

for large K.

In this paper, we consider the structure and performance of differentially encoded group codes, subject to two additional restrictions. First, we assume that $D\mathcal{G}$ is a unitary code, as in (7), so that the results of Section II-C apply. Clearly, $D\mathcal{G}$ is unitary if and only if $DD^\dagger = nI$. Second, we assume for simplicity that $n = t$. All of the results presented here extend in a natural way to $n > t$ and to single-antenna systems; however, this extension requires additional tools and introduces some complications, and so will be treated elsewhere.

Note that unitary group codes can be used in several different ways. First, if we encode messages by

$$X_k = DG_k, \qquad k = 1, \ldots, K \tag{14}$$

rather than (13), then $D\mathcal{G}$ is essentially a space–time block code, as in [1], [12], and [30]. In this case, the results of Section II-B apply when perfect channel estimates are available at the receiver, and the results of Section II-C apply when estimates are not available. Second, $D\mathcal{G}$ can also be differentially encoded, as in (13). When perfect CSI is available, we can still apply the results of Section II-B to decoding the sequence X_k, and then recover G_k from X_k and X_{k-1}. We expect the error probability of this scheme to be approximately twice the error probability of $D\mathcal{G}$ without differential encoding, since an error in X_k tends to result in two errors in the message sequence. (A similar phenomenon occurs with differentially-encoded PSK [19, p. 274].) In this paper, we are mainly interested in the final possibility, in which $D\mathcal{G}$ is differentially encoded and CSI is absent at the receiver. Here, unitary code matrices are used in a different way than in [11], so the results of Section II-C do not apply directly. In the following sections, we derive new receivers, error bounds, and code design criteria for this situation.

B. A Differential Receiver

We now derive a receiver for differentially encoded unitary group codes with $n = t$. In the absence of CSI, the ML detector for the sequence (13) consists of the quadratic receiver (8) applied to the entire received sequence $Y = [Y_0 : \cdots : Y_K]$, where

$$Y_k \triangleq \sqrt{\rho_t} H X_k + N_k, \qquad k = 0, 1, \ldots, K.$$

Even for moderate values of t and K, this receiver is quite complex. We, therefore, seek a simpler suboptimal receiver.

Given the example of DPSK, it is natural to look for a receiver that estimates G_k using only the last two received blocks

$$\bar{Y}_k \triangleq [Y_{k-1} : Y_k].$$

When $G_k = G$, the code matrices that affect $\bar{Y}_k$ are

$$\bar{C}_G \triangleq [X_{k-1} : X_{k-1} G].$$

Note that $n = t$ and $DD^\dagger = nI$ imply $D^\dagger D = tI$. From this, we can easily show that $XX^\dagger = X^\dagger X = tI$ for all $X \in D\mathcal{G}$. It follows that the $t \times 2t$ matrices $\{\bar{C}_G : G \in \mathcal{G}\}$ satisfy $\bar{C}_G \bar{C}_G^\dagger = 2tI$ for all $X_{k-1} \in D\mathcal{G}$, and can therefore be regarded as a unitary block code of length $\bar{n} = 2t$.

If X_{k-1} were known at the receiver, the optimal decoder for this block code would be the quadratic receiver (8), which depends only on the cross-product matrices

$$\bar{C}_G^\dagger \bar{C}_G = \begin{bmatrix} tI & tG \\ tG^\dagger & tI \end{bmatrix}. \tag{15}$$

Since these matrices do not depend on X_{k-1}, however, the receiver does not require knowledge of the past in order to decode the current message. Moreover, this receiver reduces to a simple and elegant form, as shown in Fig. 2

$$\begin{aligned} \hat{G} &= \arg \max_{G \in \mathcal{G}} \operatorname{Tr}\{\bar{Y}_k \bar{C}_G^\dagger \bar{C}_G \bar{Y}_k^\dagger\} \\ &= \arg \max_{G \in \mathcal{G}} \operatorname{Tr}\left\{ [Y_{k-1} : Y_k] \begin{bmatrix} tI & tG \\ tG^\dagger & tI \end{bmatrix} [Y_{k-1} : Y_k]^\dagger \right\} \\ &= \arg \max_{G \in \mathcal{G}} \operatorname{Re} \operatorname{Tr}\{Y_{k-1} G Y_k^\dagger\} \\ &= \arg \max_{G \in \mathcal{G}} \operatorname{Re} \operatorname{Tr}\{G Y_k^\dagger Y_{k-1}\} \end{aligned} \tag{16}$$

where "ReTr" denotes the real part of the trace, and the last step follows from the identity $\operatorname{Tr}(AB) = \operatorname{Tr}(BA)$. In Fig. 2, z^{-1} denotes a one-block delay. Although this receiver is much simpler than ML detection based on $Y_0, \ldots, Y_K$, its complexity grows exponentially with t and R, since $M = 2^{tR}$ comparisons are required.

This receiver has an *estimator–correlator* interpretation. If the receiver knew both X_{k-1} and the fading matrix H, then the optimal detector would be the minimum Euclidean distance rule (2). For unitary codes, this reduces to a *correlation receiver*

$$\hat{G} = \arg \max_{G \in \mathcal{G}} \operatorname{Re} \operatorname{Tr}\{H X_{k-1} G Y_k^\dagger\}. \tag{17}$$

We now recognize (16) as a correlation receiver in which HX_{k-1} is estimated by the previous received block

$$Y_{k-1} = \sqrt{\rho_t} H X_{k-1} + N_{k-1}.$$

Thus the differential receiver has the same form as the receiver for perfect CSI, and differs only in the quality of its channel estimate. More generally, this suggests that the same receiver can be used with noisy channel estimates derived from other sources, which lie between these two extremes.

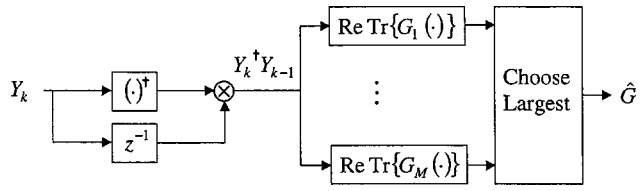

Fig. 2. A differential receiver.

C. Error Bounds and Design Criteria

By *differential space–time modulation* (DSTM), we mean the differential encoder (13) combined with the differential receiver (16). In this section, we derive a bound on the pairwise error probability of DSTM and criteria for optimally designing D and $\mathcal{G}$. As in the previous section, we assume $n = t$ and $DD^\dagger = nI$.

Consider again the detection of G_k in (13) based only on the block $\bar{Y}_k = [Y_{k-1} : Y_k]$. Recall that

$$\bar{C}_G = [X_{k-1} : X_{k-1}G], \qquad G \in \mathcal{G}$$

is a unitary block code of length $\bar{n} = 2t$, and note that

$$C_G \triangleq X_{k-1}G, \qquad G \in \mathcal{G}$$

is a unitary block code of length $n = t$.

When channel estimates are not available at the receiver, the optimal detector for the block code $\bar{C}_G$ is (16). Thus the performance of DSTM is the same as the performance of the block code $\bar{C}_G$ with ML detection (8). Hence the pairwise error probability is bounded by (9). From (15) and the unitary property (7), we have for all G and $G' \in \mathcal{G}$

$$\begin{aligned}\bar{n}I - (1/\bar{n})\bar{C}_G\bar{C}_{G'}^\dagger\bar{C}_{G'}\bar{C}_G^\dagger &= tI - \frac{1}{2}C_GC_{G'}^\dagger - \frac{1}{2}C_{G'}C_G^\dagger \\ &= \frac{1}{2}(C_G - C_{G'})(C_G - C_{G'})^\dagger.\end{aligned} \tag{18}$$

Thus (9) can be written as

$$\Pr\{\bar{C}_G \to \bar{C}_{G'}\} \le \frac{1}{\left|I + \frac{\rho_t^2\bar{n}}{8(1+\rho_t\bar{n})}(C_G - C_{G'})(C_G - C_{G'})^\dagger\right|^r}. \tag{19}$$

For large ρ_t, this bound takes the form $(\Lambda_d\rho_t/4)^{-r\nu_d}$, where ν_d and Λ_d represent the diversity and coding advantage of the pair $\bar{C}_G, \bar{C}_{G'}$. From Section II-C, we know that $\nu_d = \nu_d(C_G, C_{G'})$ is equal to the rank of the left side of (18), which clearly equals the rank of $C_G - C_{G'}$. From Section II-B, we therefore have $\nu_d(C_G, C_{G'}) = \nu_p(C_G, C_{G'})$, which is the diversity advantage of C_G and $C_{G'}$ for perfect CSI. From Section II-C, the coding advantage Λ_d is equal to the angular distance $\Lambda_a(\bar{C}_G, \bar{C}_{G'})$, which from (18) reduces to

$$\begin{aligned}\Lambda_d(C_G, C_{G'}) &\triangleq |\bar{n}I - (1/\bar{n})\bar{C}_G\bar{C}_{G'}^\dagger\bar{C}_{G'}\bar{C}_G^\dagger|_+^{1/\nu_d} \\ &= \left|\frac{1}{2}(C_G - C_{G'})(C_G - C_{G'})^\dagger\right|_+^{1/\nu_p} \\ &= \frac{1}{2}\Lambda_p(C_G, C_{G'})\end{aligned} \tag{20}$$

where Λ_p is the product distance (4). Recall that Λ_p measures coding advantage when perfect CSI is available at the receiver. Using $D^\dagger D = DD^\dagger = tI$, we can express this in terms of the distance between the messages

$$\Lambda_p(C_G, C_{G'}) = t\Lambda_p(G, G'). \tag{21}$$

These results have important implications for the theory and design of differential space–time modulation. First, DSTM based on the group code $D\mathcal{G}$ achieves maximum diversity in the absence of CSI if and only if $D\mathcal{G}$ achieves maximum diversity for perfect CSI. Second, the coding advantage of DSTM without CSI is exactly half the coding advantage of $D\mathcal{G}$ for perfect CSI. Third, the design criteria for DSTM are the same as in Section II-B: choose $D\mathcal{G}$ so that $\nu_p = t$ and such that Λ_p is as large as possible. From (21), we can clearly choose D to be any matrix that satisfies $DD^\dagger = tI$, since the choice does not affect performance.

Comparing with (3) for $\rho_t\bar{n} \gg 1$, we see that (19) is essentially the same as the Chernoff bound *for perfect CSI*, except for a 3-dB loss in ρ_t. This suggests that the pairwise error probability of DSTM suffers a 3-dB loss relative to the performance of the block code $D\mathcal{G}$ with the correlation receiver (17) and perfect CSI. This conclusion can be verified by examining the exact pairwise error probabilities for large ρ_t. This performance loss is due mainly to the suboptimal receiver (16), which uses only the two most recent received blocks to estimate G_k, rather than the entire received sequence.

IV. Optimal Unitary Group Codes

Section III provides a general framework for differential space–time modulation based on unitary group codes. In this section, we characterize all unitary group codes with $|\mathcal{G}| = 2^p$ and $\nu_p = n = t = 2$, and we identify those that are optimal in the sense of achieving the largest minimum product distance Λ_p. All of the codes presented here can be used in two distinct ways: First, when perfect CSI is available at the receiver, we can use them as space–time block codes with encoder (14) and decoder (2), as in Section II-B. Second, when CSI is absent, we can use the differential encoder (13) and detector (16), as in Section III-B. As shown in Section III-C, the design criteria for these two applications are related by $\nu_d = \nu_p$ and $\Lambda_d = \Lambda_p/2$.[2]

For $t = n = 2$, the choice of D affects the constellation but not the distance structure of $D\mathcal{G}$, as shown by (21). We can therefore choose D to be any matrix that satisfies $DD^\dagger = 2I$. In particular

$$D = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \tag{22}$$

is a convenient choice for all of the codes presented below. We say that two codes, $D\mathcal{G}$ and $D'\mathcal{G}'$, are *equivalent* if there is a unitary matrix U such that $\mathcal{G}' = U\mathcal{G}U^\dagger$. From Section II-B and (21), it is easy to see that equivalent codes have the same diversity ν_p and minimum distance Λ_p. Throughout this section and the Appendix, we specify groups by their generators. Let

[2]In principle, the codes in this section could also be used as space–time block codes without CSI at the receiver, as in Section II-C. Since $n = t$ implies $\nu_a = 0$, however, these codes provide no diversity in this context.

$\langle G_1, \ldots, G_m \rangle$ denote the group consisting of all distinct products of powers of $G_1, \ldots, G_m$.

Suppose that $M = 2^p > 1$. In the Appendix, we show that all unitary group codes with $|\mathcal{G}| = M$ and $\nu_p = n = t = 2$ are either *cyclic* or *dicyclic* (cf. Appendix D). For odd $0 < k < M$, the (M, k) *cyclic group code* is defined by (22) and

$$\mathcal{G} = \left\langle \begin{bmatrix} \omega_M & 0 \\ 0 & \omega_M^k \end{bmatrix} \right\rangle \tag{23}$$

where $\omega_M = \exp(2\pi j/M)$. This code takes values in the M-PSK constellation and has $|\mathcal{G}| = M$ code matrices, diversity advantage $\nu_p = 2$, and minimum product distance

$$\begin{aligned} \Lambda_p &= \min_{G \neq G'} \Lambda_p(DG, DG') \\ &= \min_{1 \leq l \leq M-1} 8|\sin(\pi l/M) \cdot \sin(\pi k l/M)| \end{aligned}$$

which is positive for all odd k (cf. Appendix B). For example, the $(2, 1)$ cyclic group code has $\mathcal{G} = \{\pm I\}$ and $\Lambda_p = 8\sin^2(\pi/2) = 8$. For all $M \geq 8$, the *dicyclic group code* is given by (22) and

$$\mathcal{G} = \left\langle \begin{bmatrix} \omega_{M/2} & 0 \\ 0 & \omega_{M/2}^* \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\rangle \tag{24}$$

which takes values in $M/2$-PSK and has M code matrices, diversity advantage $\nu_p = 2$, and minimum product distance $\Lambda_p = 8\sin^2(2\pi/M) > 0$. For example, the code in Example 3 is dicyclic with $M = 8$ and $\Lambda_p = 4$. In the Appendix, we show that every unitary group code with $|\mathcal{G}| = M$ and $\nu_p = n = t = 2$ is equivalent to an (M, k) cyclic code or the dicyclic code. Thus there are at most $M/2 + 1$ nonequivalent unitary group codes with these parameters.

Unitary group codes with maximum Λ_p are given in Table I for all $2 \leq M \leq 32$. All of these codes use the initial matrix (22). Also shown for comparison at the bottom of the table are Alamouti's QPSK code [1], and the differential version of this code proposed by Tarokh and Jafarkhani [31], which takes values in 9QAM.

For R =0.5 b/s/Hz, the only unitary group code with $\nu_p = 2$ is the $(2, 1)$ cyclic group code, which is therefore optimal. For $R = 1$, the $(4, 1)$ and $(4, 3)$ cyclic group codes are both optimal. The $R = 1$ code given in Table I is equivalent to the $(4, 3)$ cyclic group code (see Table III in Appendix B), but has the advantage of taking values in BPSK rather than QPSK. Note that this code has the same code matrices as (6) for binary a and b; thus it is essentially Alamouti's binary code. The differentially encoded version of this code was given in [31]. To the best of our knowledge, the observations that (6) is a group code for binary a and b, and that it is optimal with respect to Λ_p, are new. The group structure may be useful in simplifying the encoding and decoding procedures.

For $R = 1.5$, the dicyclic code given in Example 3 is optimal. Since the underlying group $\mathcal{G}$ is known in algebra as the quaternion group ([14, p. 32]), we call this the *quaternion code*. Note that this code has the same minimum product distance as the optimal $R = 1$ code, but achieves a 50% higher rate. When perfect CSI is available at the receiver, the quaternion code achieves three quarters of the rate of Alamouti's QPSK code, but with a 3-dB higher coding advantage. Moreover, when CSI is absent, the quaternion code can be differentially encoded and detected without expanding the constellation. Fig. 3 gives a plot of the bit-error rate (BER) of this code for one receive antenna $(r = 1)$, with and without CSI. At a BER of 10^{-3}, the performance of the

TABLE I
OPTIMAL UNITARY GROUP CODES $(n = t = 2)$

R	$\mathcal{C}$	$\mathcal{G}$	Λ_p
0.5	BPSK	$\left\langle \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \right\rangle$	8
1.0	BPSK[1,2]	$\left\langle \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\rangle$	4
1.5	QPSK	$\left\langle \begin{bmatrix} j & 0 \\ 0 & -j \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\rangle$	4
2.0	8PSK	$\left\langle \begin{bmatrix} \omega_8 & 0 \\ 0 & \omega_8^* \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\rangle$	1.172
2.5	32PSK	$\left\langle \begin{bmatrix} \omega_{32} & 0 \\ 0 & j\omega_{32} \end{bmatrix} \right\rangle$	0.497
2.0	QPSK[1]/9QAM[2]	-	2

[1] Alamouti [1], [2] Tarokh-Jafarkhani [31]

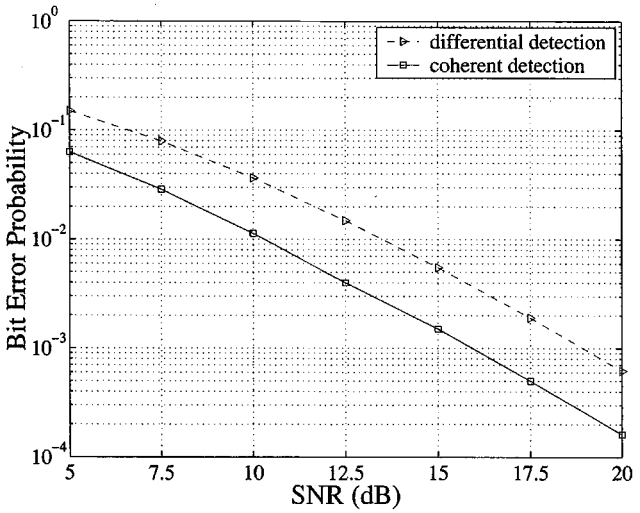

Fig. 3. Bit-error probability of the quaternion code $(r = 1)$.

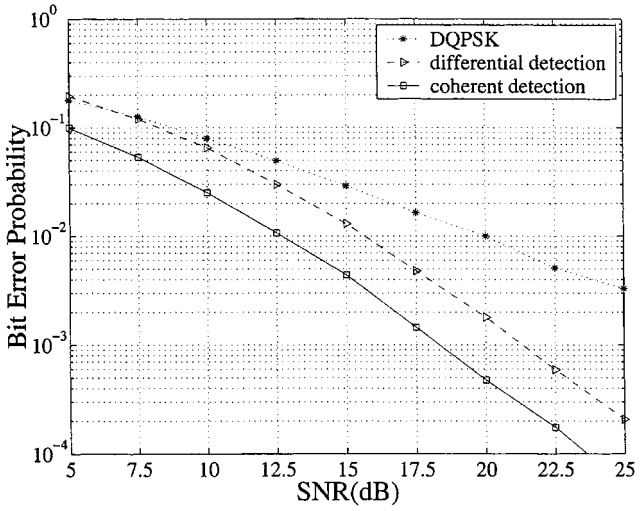

Fig. 4. Bit-error probability of the optimal $R = 2$ code $(r = 1)$.

quaternion code with coherent detection is roughly 1.1 dB better than Alamouti's QPSK code, and with differential detection is about 2.0 dB better than the corresponding code in [31].

For $R = 2$, the best cyclic and dicyclic group codes have the same minimum distance. In Table I, we choose the dicyclic code because the second-nearest neighbors are significantly farther apart. Fig. 4 shows the BER of this code for $r = 1$, along with the performance of single-antenna differential QPSK. Note that this code performs somewhat better than would be expected on the basis of product distance alone.[3] Comparing with [31, Fig. 4] at $\mathrm{BER} = 10^{-3}$, we see that the $R = 2$ code is 1.4 dB worse than Alamouti's QPSK code for coherent demodulation, and only 0.5 dB worse than the corresponding differential code in [31]. In exchange for this loss of performance, the code in Table I can be differentially encoded without changing the signal constellation, preserves the constant modulus property of the constellation (i.e., 8PSK instead of 9QAM), and has a simpler differential encoder and decoder. For coherent transmission, however, the encoder and decoder of the $R = 2$ code are more complex than Alamouti's code, and the constellation is larger (8PSK versus QPSK). Similar observations apply to the optimal $R = 2$ cyclic group code.

Finally, for $R = 2.5$, the $(32, k)$ cyclic group codes are optimal for $k = 7, 9, 23$, and 25. In Table I, we arbitrarily choose the $(32, 9)$ code.

V. Conclusion

We have considered the design of space–time coding and modulation techniques that do not require channel estimates at the transmitter or receiver. We proposed a new and general approach to differential modulation based on unitary group codes, a rich class of space–time block codes which extend Slepian's idea to multiple transmit antennas and complex signal constellations. This approach can be applied to any number of transmit antennas and any target constellation. For the particular case $n = t$, we derived low-complexity differential receivers, error bounds, and design criteria for differential space–time modulation. From these results, it is clear that differentially encoded unitary group codes can be decoded with or without channel estimates at the receiver. This allows the receiver to use channel estimates when they are available; however, performance degrades by 3 dB when estimates are not available. Finally, we used the design criteria to construct optimal unitary group codes for $n = t = 2$. These codes can also be used as space–time block codes when CSI is available at the receiver. As noted earlier, some of these results have been obtained independently by Hochwald and Sweldens [13].

The methods proposed here are not the only way to perform differential modulation with multiple transmit antennas. For example, the differential transmit diversity schemes of Tarokh and Jafarkhani [31], which are based on orthogonal block designs, do not fit within our framework. As pointed out in [31], however, modulators based on orthogonal block designs seem to be limited to $t \leq 8$ for real constellations and to $t = 2$ for complex constellations. The approach presented here is, to the best of our knowledge, the only known approach to designing differential space–time modulation in other situations.

In a sense, this paper raises more questions than it answers. Multichannel group codes are a rich topic for further investigation. The structure and algebraic properties of these codes will be investigated in a companion paper. The proposed approach to differential modulation extends in a natural way to $n > t$ and to single-antenna systems. This extension requires additional tools and introduces some complications, however, and so will be treated elsewhere. Since we have suggested DSTM for applications like frequency hopping and fast-fading channels, it is natural to explore extensions of DSTM that exploit time and frequency diversity as well as space diversity. In particular, DSTM seems to extend in a straightforward way to dispersive channels when combined with orthogonal frequency-division multiplexing. Although the approach presented here permits the design of differential modulation schemes for any number of transmit antennas and any signal constellation, we have only skimmed the surface in terms of the actual design of codes. Our results suggest, however, that the extensive literature on group codes can be leveraged to provide a wealth of space–time block codes and differential space–time modulation schemes.

Appendix

A. Group Design Preliminaries

In this appendix, we characterize all unitary group codes with $\nu_p = t = n = 2$ and $|\mathcal{G}| = 2^p > 1$, and we identify those with maximum product distance Λ_p. We assume familiarity with linear algebra at the level of [27] and group theory at the level of [14, Chs. 1 and 2]. We begin with some useful results on 2×2 unitary matrices.

Consider the matrix

$$G = \begin{bmatrix} a & c \\ b & d \end{bmatrix}.$$

If $GG^\dagger = I$, then the determinant $\Delta = ad - bc$ is a unit-magnitude complex number. Since

$$G^\dagger = G^{-1} = \frac{1}{\Delta}\begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$$

it follows that $d = a^*\Delta$ and $c = -b^*\Delta$. Hence any 2×2 unitary matrix takes the form

$$G = \begin{bmatrix} a & -b^*\Delta \\ b & a^*\Delta \end{bmatrix} \tag{25}$$

where $|a|^2 + |b|^2 = |\Delta| = 1$. We say that G is *diagonal* if $b = 0$ in (25), which we write as $G = \mathrm{diag}\{a, a^*\Delta\}$. We say G is *off-diagonal* if $a = 0$, and we write $G = \overline{\mathrm{diag}}\{b, -b^*\Delta\}$. The nonzero entries in diagonal and off-diagonal unitary matrices always have unit magnitude.

Lemma A1: Let $\Theta = \mathrm{diag}\{x, y\}$ be unitary, and let G be the unitary matrix in (25). *a)* If $x \neq y$, then $G\Theta G^\dagger$ is diagonal if and only if G is diagonal or off-diagonal. *b)* $G\Theta G^\dagger$ is off-diagonal if and only if $|a| = |b| = 1/\sqrt{2}$ and $x = -y$.

[3] This can be explained by observing that each code matrix in the $R = 2$ group code has two nearest neighbors, whereas Alamouti's QPSK code has four.

Proof: To prove *a)*, note that

$$G\Theta G^\dagger = \begin{bmatrix} a & -b^*\Delta \\ b & a^*\Delta \end{bmatrix} \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} \begin{bmatrix} a^* & b^* \\ -b\Delta^* & a\Delta^* \end{bmatrix} = \begin{bmatrix} |a|^2x + |b|^2y & ab^*(x-y) \\ a^*b(x-y) & |b|^2x + |a|^2y \end{bmatrix} \tag{26}$$

where $|a|^2 + |b|^2 = |\Delta| = 1$. If $x \neq y$, then $G\Theta G^\dagger$ is diagonal if and only if $ab^* = 0$, which holds if and only if G is diagonal ($b = 0$) or off-diagonal ($a = 0$). To prove *b)*, note that $|x| = |y| = 1$ if Θ is unitary. If $G\Theta G^\dagger$ is off-diagonal then $|a|^2x + |b|^2y = 0$, which is true if and only if $|a| = |b|$ and $x = -y$. □

Lemma A2: Let $\Theta = \mathrm{diag}\{x, y\}$ be unitary. Then Θ has a nondiagonal unitary square root if and only if $x = y$. Moreover, every such root is of the form

$$G = \sqrt{x} \begin{bmatrix} \alpha & d^* \\ d & -\alpha \end{bmatrix} \tag{27}$$

where α is real, d is a nonzero complex number, and $\alpha^2 + |d|^2 = 1$.

Proof: Suppose that $G^2 = \mathrm{diag}\{x, y\}$, where $|x| = |y| = 1$. From (25), we have

$$G^2 = \begin{bmatrix} a^2 - |b|^2\Delta & -ab^*\Delta - a^*b^*\Delta^2 \\ ab + a^*b\Delta & a^{*2}\Delta^2 - |b|^2\Delta \end{bmatrix} = \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix}.$$

For any nonzero complex number u, note that $|u + u'| = |u| + |u'|$ implies $u' = |u'|u/|u|$. Setting $u = -|b|^2\Delta$ and $u' = a^2$, we see that if G is not diagonal ($b \neq 0$) then $a^2 - |b|^2\Delta = x$ implies $\Delta = -x$ and $a^2 = |a|^2x$. Since this further implies $a^{*2}\Delta^2 = a^2$, a nondiagonal root exists only if $x = y$. Substituting $\alpha = a/\sqrt{x}$ and $d = b/\sqrt{x}$ into (25), we obtain (27).

B. Cyclic Groups

Since we are interested only in codes with $\nu_p = t = n = 2$, by the remarks following (5) it suffices to search for unitary group codes $D\mathcal{G}$ that maximize the modified distance

$$\Lambda_p^* \triangleq \min_{G \neq G'} \Lambda_p^*(DG, DG') = 2 \cdot \min_{G \neq G'} \Lambda_p^*(G, G')$$

where the second equality is similar to (21). From Section II-B, $\Lambda_p^* > 0$ if and only if $\nu_p = t$, in which case $\Lambda_p^* = \Lambda_p$. In this section, we characterize all cyclic groups with $|\mathcal{G}| = 2^p$ and $\Lambda_p^* > 0$.

A group $\mathcal{G}$ is *cyclic* if there is a unitary matrix Θ such that

$$\mathcal{G} = \langle\Theta\rangle = \{I, \Theta, \ldots, \Theta^{M-1}\}$$

where M is the *order* of Θ, i.e., the smallest integer such that $\Theta^M = I$. For example, if $M = 2^p$, $0 < k < M$ is odd, and $\omega_M = \exp(2\pi j/M)$, the matrix

$$\Theta = \begin{bmatrix} \omega_M & 0 \\ 0 & \omega_M^k \end{bmatrix}$$

TABLE II
OPTIMAL (M, k) CYCLIC GROUP CODES $(2 \leq M \leq 32)$

M	$\mathcal{C}$	k	Λ_p^*
2	BPSK	1	8
4	QPSK	1,3	4
8	8PSK	3,5	$2\sqrt{2}$
16	16PSK	7,9	1.172
32	32PSK	7,9,23,25	0.497

generates a cyclic group of order M, which we call the (M, k) *cyclic group*. It easily shown that the minimum distance of this group is

$$\Lambda_p^* = 2 \cdot \min_{\substack{0 \leq l, l' < M \\ l \neq l'}} \Lambda_p^*(\Theta^l, \Theta^{l'}) = \min_{1 \leq l \leq M-1} 8|\sin(\pi l/M) \cdot \sin(\pi kl/M)|$$

which is positive for all odd k. For a given M, the smallest distance is $\Lambda_p^* = 8\sin^2(\pi/M)$, which is achieved by $k = 1$ and $M - 1$ ($\omega_M^k = \omega_M$ or ω_M^*). The (M, k) cyclic group codes with maximum Λ_p^* are given in Table II for $M = 2$–32, along with the constellation $\mathcal{C}$, assuming the use of the initial matrix (22). The following lemmas show that every cyclic group code with $\Lambda_p^* > 0$ is equivalent to an (M, k) cyclic group code.

Lemma A3: For $M = 2^p$, every group of diagonal matrices with $|\mathcal{G}| = M$ and $\Lambda_p^* > 0$ is an (M, k) cyclic group, for some odd $0 < k < M$.

Proof: Let $G = \mathrm{diag}\{\lambda_1, \lambda_2\}$ be a generic element of $\mathcal{G}$. If G_0 and $G_1 \in \mathcal{G}$ have the same λ_1 or λ_2, then $\Lambda_p^*(G_0, G_1) = 0$. Thus all of the matrices in $\mathcal{G}$ differ in both λ_1 and λ_2. Since $G^M = I$, λ_1 and λ_2 are both powers of ω_M. Since there are exactly M distinct powers of ω_M, each appears in $\mathcal{G}$ once and only once in each diagonal position. Let $G' = \mathrm{diag}\{\lambda_1', \lambda_2'\} \in \mathcal{G}$ be such that $\lambda_1' = \omega_M$. Note that G' has order M, so $\mathcal{G} = \langle G'\rangle$. Since all of the matrices in $\langle G'\rangle$ differ in λ_2, it follows that λ_2' is an odd power of ω_M. Hence $G' = \mathrm{diag}\{\omega_M, \omega_M^k\}$, for odd $0 < k < M$, thereby proving the lemma. □

Lemma A4: For $M = 2^p$, every cyclic group code $D\mathcal{G}$ with $|\mathcal{G}| = M$ and $\Lambda_p^* > 0$ is equivalent to an (M, k) cyclic group code, for some odd $0 < k < M$.

Proof: Let $D\mathcal{G}$ be a cyclic group code with generator Θ. Recall that a matrix A is said to be *normal* if $AA^\dagger = A^\dagger A$. Normal matrices can be diagonalized by unitary transformations ([27, p. 311]). Since the unitary matrix Θ is clearly normal, there exists a unitary matrix U such that $\Theta' = U\Theta U^\dagger$ is diagonal. Since $\mathcal{G}' = \langle\Theta'\rangle = U\mathcal{G}U^\dagger$, $D\mathcal{G}'$ is equivalent to $D\mathcal{G}$. Note that $\mathcal{G}'$ is a group of diagonal matrices with $\Lambda_p^* > 0$. Hence, by Lemma A3, $\mathcal{G}'$ is an (M, k) cyclic group. □

TABLE III
OPTIMAL CYCLIC GROUP CODES FOR $M = 2$–16 ($k = M/2 + 1$)

R	$\mathcal{C}$	$\mathcal{G}$	Λ_p^*
1	BPSK	$\left\langle \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\rangle$	4
1.5	QPSK	$\left\langle \begin{bmatrix} 0 & j \\ 1 & 0 \end{bmatrix} \right\rangle$	$2\sqrt{2}$
2	8PSK	$\left\langle \begin{bmatrix} 0 & \omega_8 \\ 1 & 0 \end{bmatrix} \right\rangle$	1.172

TABLE IV
DICYCLIC GROUP CODES FOR $M \leq 32$

R	$\mathcal{C}$	$\mathcal{G}$	Λ_p^*
1.5	QPSK	$\left\langle \begin{bmatrix} j & 0 \\ 0 & -j \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\rangle$	4
2	8PSK	$\left\langle \begin{bmatrix} \omega_8 & 0 \\ 0 & \omega_8^* \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\rangle$	1.172
2.5	16PSK	$\left\langle \begin{bmatrix} \omega_{16} & 0 \\ 0 & \omega_{16}^* \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\rangle$	0.304

The groups in Table II can often be represented in smaller constellations by using a nondiagonal generator. For example, the group generated by

$$\Theta = \begin{bmatrix} 0 & \omega_{M/2} \\ 1 & 0 \end{bmatrix}$$

is equivalent to an $(M,\ M/2+1)$ cyclic group, since the eigenvalues of Θ are $\pm\omega_M$. Used with the initial matrix (22), this group code takes values in the $M/2$-PSK constellation. Table III gives the generators and minimum distances of all such groups for $M = 2$–16. From Table II, we see that all of the cyclic groups in Table III are optimal.

C. Dicyclic Groups

Let $\mathcal{G}$ be an arbitrary group and let v denote the maximum order of any element in $\mathcal{G}$. Recall that v always divides $M = |\mathcal{G}|$, and $v = M$ if and only if $\mathcal{G}$ is cyclic. In this section, we consider groups with $v = M/2 = 2^{p-1}$. Such groups can always be generated by two elements. For example, if $v = 2^{p-1} \geq 4$ then

$$\mathcal{G} = \langle \Theta, R \rangle \triangleq \left\langle \begin{bmatrix} \omega_v & 0 \\ 0 & \omega_v^* \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\rangle \tag{28}$$

is a group with $|\mathcal{G}| = 2v$ and $\Lambda_p^* = 8\sin^2(\pi/v) > 0$. Since $R^2 = -I$ and $R\Theta R^{-1} = \Theta^{-1}$, $\mathcal{G}$ is the well-known *dicyclic group* ([3, p. 7]). Table IV gives all of the dicyclic groups for $M \leq 32$, along with their minimum distances and constellations, assuming the initial matrix (22). For $v = 2$, note that $\Theta = R^2 = -I$, so the resulting group is cyclic rather than dicyclic.

In the rest of this section, we show that every unitary group code $D\mathcal{G}$ with $|\mathcal{G}| = 2v$ and $\Lambda_p^* > 0$ is equivalent to the dicyclic group code (28). We first require two technical lemmas.

Lemma A5: Let $\mathcal{G}$ be such that $v = |\mathcal{G}|/2 = 2^{p-1}$ and $\Lambda_p^* > 0$. If Θ is a diagonal element of order v and $\mathcal{G} = \langle \Theta, R \rangle$, then *a)* R is not diagonal, *b)* R^2 is a nonprimitive element of $\langle \Theta \rangle$, and *c)* $R\Theta R^\dagger$ is a primitive element of $\langle \Theta \rangle$.

Proof: If R is diagonal, then $\mathcal{G}$ is a group of diagonal matrices with $\Lambda_p^* > 0$. By Lemma A3, $\mathcal{G}$ must be cyclic and therefore contains an element of order $2v$, which contradicts the assumption that v is the maximum order. Hence R cannot be diagonal, which proves *a)*. Since $|\mathcal{G}| = 2v$ all elements in $\mathcal{G}$ must be either in $\langle \Theta \rangle$ or the coset $R\langle \Theta \rangle$. Since $R \notin \langle \Theta \rangle$, it follows that $R^2 \notin R\langle \Theta \rangle$ and hence $R^2 \in \langle \Theta \rangle$. If R^2 is primitive in $\langle \Theta \rangle$, then R has order $2v$, a contradiction. Thus R^2 is a nonprimitive element in $\langle \Theta \rangle$, proving *b)*. Finally, note that $\langle \Theta \rangle$ is a subgroup of $\mathcal{G}$ with index $|\mathcal{G}|/v = 2$. Since all subgroups of index 2 are normal ([14, p. 45]), $\langle \Theta \rangle$ is a normal subgroup of $\mathcal{G}$. Hence $G\langle \Theta \rangle G^{-1} = \langle \Theta \rangle$ for all $G \in \mathcal{G}$ ([14, p. 41]), where $G^{-1} = G^\dagger$. In particular, $R\langle \Theta \rangle R^\dagger = \langle R\Theta R^\dagger \rangle = \langle \Theta \rangle$, which implies $R\Theta R^\dagger$ is primitive in $\langle \Theta \rangle$, proving *c)*. □

Lemma A6: Let k be an odd integer such that $1 < k < v$, where $v = 2^{p-1} \geq 4$. Then for every $0 \leq q < v$ such that $(k-1)q = 0 \bmod v$, there exists an integer $0 \leq l < v$ such that

$$q + l(k+1) = 0 \bmod v \tag{29}$$

except for $k = v - 1$ and $q = v/2$.

Proof: Let $[n, m]$ denote the greatest common divisor of n and m. From elementary number theory ([22, p. 102]), the congruence $(k-1)q = 0 \bmod v$ has exactly $[k-1, v]$ solutions in $0 \leq q < v$

$$q = \frac{iv}{[k-1, v]}, \qquad 0 \leq i < [k-1, v]. \tag{30}$$

For $q = 0$, note that $l = 0$ always satisfies (29). We therefore restrict attention to $q > 0$.

Suppose first that $1 < k < v-1$. From [22, p. 102], the congruence (29) has a solution l if and only if $[k+1, v]$ divides q. It follows that there is a solution *for every* q *in* (30) if and only if $[k+1, v]$ divides $v/[k-1, v]$, or equivalently, if and only if $[k+1, v] \cdot [k-1, v]$ divides v. Since $v = 2^{p-1}$ and $k < v-1$, $[k+1, v]$ and $[k-1, v]$ are both powers of 2, no greater than $v/2$. Moreover, since $k-1$ and $k+1$ are consecutive even integers, only one is divisible by 4; hence $[k+1, v] = 2$ or $[k-1, v] = 2$. It follows that $[k+1, v] \cdot [k-1, v]$ is a power of 2 not greater than v, which always divides v. Thus to every q in (30) there is an l that satisfies (29), which proves the lemma for all $k \neq v-1$. If $k = v-1$, we have $[k-1, v] = 2$ and hence the only nonzero number in (30) is $q = v/2$. Since $k = v-1$ and $q = v/2$ clearly admits no solution l in (29), the proof is complete. □

Lemma A7: Every group code $D\mathcal{G}$ with $v = |\mathcal{G}|/2 = 2^{p-1}$ and $\Lambda_p^* > 0$ is equivalent to the dicyclic group code (28). Moreover, there is no group code with $v = |\mathcal{G}|/2 = 2$ and $\Lambda_p^* > 0$.

Proof: Let $D\mathcal{G}$ be an arbitrary group code with $v = |\mathcal{G}|/2 = 2^{p-1}$ and $\Lambda_p^* > 0$, and let $\Theta \in \mathcal{G}$ be any element of order v. Since the subgroup $\langle \Theta \rangle \subset \mathcal{G}$ is cyclic and has positive minimum distance, by Lemma A4 it is equivalent to a (v, k) cyclic group. We can therefore assume $\Theta = \mathrm{diag}\{\omega_v, \omega_v^k\}$ for some odd $0 < k < v$. Let R be any

matrix in $\mathcal{G} - \langle\Theta\rangle$. Since $|\mathcal{G}| = 2v$, every element in $\mathcal{G}$ must be either in $\langle\Theta\rangle$ or in $R\langle\Theta\rangle$; hence $\mathcal{G} = \langle\Theta, R\rangle$.

Suppose first that $k = 1$, in which case $\Theta = \omega_v I$. By Lemma A5, R^2 is a nonprimitive element of $\langle\omega_v I\rangle$, and hence $R^2 = \omega_v^{2u} I$ for some integer $0 \le u < v/2$. By Lemmas A2 and A5, R is a nondiagonal matrix of the form (27), where $x = \omega_v^{2u}$. Using (5) and (27), we can explicitly calculate the distance between the group matrices I and $R\Theta^l$ for all $1 \le l \le v$

$$\Lambda_p^*(I, R\Theta^l) = \left|1 - x\omega_v^{2l}\right| = \left|1 - \omega_v^{2(u+l)}\right|.$$

Since this vanishes for $l = v - u$, it follows that $\Lambda_p^* = 0$, which is a contradiction. We, therefore, conclude that $k \neq 1$. In particular, for $v = |\mathcal{G}|/2 = 2, k = 1$ is the only possibility. Hence, there exists no group with $v = |\mathcal{G}|/2 = 2$ and $\Lambda_p^* > 0$.

Now suppose $k \neq 1$ and $v \ge 4$. By Lemma A5, R is nondiagonal and $R\Theta R^\dagger$ lies in $\langle\Theta\rangle$ and is therefore diagonal. Since $\omega_v \neq \omega_v^k$, it follows from Lemma A1*a)* that R is off-diagonal. Lemma A5 further asserts that R^2 is a nonprimitive element of $\langle\Theta\rangle$. By Lemma A2, it follows that R is of the form

$$R = \sqrt{x}\begin{bmatrix} 0 & d^* \\ d & 0 \end{bmatrix} \tag{31}$$

where $|d| = 1$ and $R^2 = xI \in \langle\Theta\rangle$. For $\Theta = \mathrm{diag}\{\omega_v, \omega_v^k\}$, the only elements in $\langle\Theta\rangle$ of the form xI are the matrices $\Theta^q = \omega_v^q I$, where $0 \le q < v$ is such that $\omega_v^q = \omega_v^{kq}$. Hence $x = \omega_v^q$ for some q that satisfies $(k-1)q = 0 \bmod v$. Given k and q, we can use (5) and (31) to calculate the distance

$$\Lambda_p^*(I, R\Theta^l) = \left|1 - x\omega_v^{l(k+1)}\right| = \left|1 - \omega_v^{q+l(k+1)}\right|.$$

Note that if there exists an l such that $q + l(k+1) = 0 \bmod v$, then $\Lambda_p^*(I, R\Theta^l)$ vanishes and hence $\Lambda_p^* = 0$. Since $1 < k < v$ is odd, Lemma A6 shows that $\Lambda_p^* = 0$ for all k and q, with the possible exception of $k = v - 1$ and $q = v/2$.

We have now proved that, if $\mathcal{G} = \langle\Theta, R\rangle$ has $\Lambda_p^* > 0$, then the only possible values of k and q are $k = v - 1$ and $q = v/2$. It follows that $\Theta = \mathrm{diag}\{\omega_v, \omega_v^*\}$ and $x \triangleq \omega_v^q = -1$. Substituting x into (31), we obtain $R = \overline{\mathrm{diag}}\{jd, jd^*\}$. Defining the unitary matrix $U \triangleq \mathrm{diag}\{jd, 1\}$, we observe that $D\mathcal{G}$ is equivalent to $D\mathcal{G}'$, where

$$\mathcal{G}' = \langle\Theta', R'\rangle, \Theta' \triangleq U\Theta U^\dagger = \mathrm{diag}\{\omega_v, \omega_v^*\}$$

and

$$R' \triangleq URU^\dagger = \overline{\mathrm{diag}}\{1, -1\}.$$

Since $\mathcal{G}'$ is the dicyclic group (28), which has positive minimum distance, we conclude that every group with $v = |\mathcal{G}|/2 = 2^{p-1}$ and $\Lambda_p^* > 0$ is equivalent to the dicyclic group code (28), thereby completing the proof. □

D. Other Extensions and Optimal Groups

Let $M = 2^p > 1$ and let $D\mathcal{G}$ be any unitary group code with $|\mathcal{G}| = M$, $\Lambda_p^* > 0$, and maximum element order v. In Sections B and C of this appendix, we showed that $D\mathcal{G}$ is equivalent to an (M, k) cyclic group code if $v = M$, and $D\mathcal{G}$ is equivalent to the dicyclic group code (28) if $v = M/2$. In this section, we show that there are no other possibilities. In particular, we prove $M \le 2v$. Since v divides M, this implies $v = M$ or $M/2$; hence $\mathcal{G}$ is either cyclic or dicyclic.

Lemma A8: If $D\mathcal{G}$ is such that $|\mathcal{G}| = 2^p > 1$ and $\Lambda_p^* > 0$, then $\mathcal{G}$ contains $-I$ and this is the only element of order 2.

Proof: To show that $-I$ is the only possible element of order 2, let $F \in \mathcal{G}$ be such that $F \neq I$ and $F^2 = I$. It follows that the eigenvalues of F satisfy $\lambda_1, \lambda_2 \in \{-1, 1\}$. Since $\lambda_i = 1$ implies $\Lambda_p^*(I, F) = 0$, both eigenvalues are -1. From the proof of Lemma A4, F is a normal matrix, which can be diagonalized by a unitary matrix U. Hence $F = U(-I)U^\dagger = -I$. To show that $\mathcal{G}$ contains $-I$, let $G \in \mathcal{G}$ be any element with order $r > 1$. Since r divides $|\mathcal{G}|$, r is even. Setting $F = G^{r/2}$, we have $F \neq I$ and $F^2 = I$, and hence $F = -I \in \mathcal{G}$. □

Lemma A9: If $D\mathcal{G}$ is such that $|\mathcal{G}| = 2^p > 1$ and $\Lambda_p^* > 0$, then $|\mathcal{G}| \le 2v$.

Proof: Since the result follows from Lemma A8 for $v = 2$, we can assume $v \ge 4$. Suppose that $|\mathcal{G}| > 2v$ and $\Lambda_p^* > 0$. Since $|\mathcal{G}|$ is a prime power, the first Sylow Theorem ([14, p. 94]) asserts that every proper subgroup $\mathcal{G}_0 \subset \mathcal{G}$ is normal in some larger subgroup $\mathcal{G}_1 \subset \mathcal{G}$ of order $|\mathcal{G}_1| = 2|\mathcal{G}_0|$. Let $\mathcal{G}_0 = \langle\Theta\rangle$ where Θ is any element of order v. Then $\mathcal{G}_1$ is a group with $\Lambda_p^* > 0$ and order $2v$. From Lemma A7, it follows that $\mathcal{G}_1$ is equivalent to the dicyclic group (28). We can, therefore, assume $\mathcal{G}_1 = \langle\Theta, R_1\rangle$, where $\Theta = \mathrm{diag}\{\omega_v, \omega_v^*\}$ and $R_1 = \overline{\mathrm{diag}}\{1, -1\}$.

Let $\mathcal{G}_d$ be the subgroup of diagonal matrices in $\mathcal{G}$. Since $\mathcal{G}_d$ contains $\langle\Theta\rangle$, we have $|\mathcal{G}_d| \ge v$. Conversely, by Lemma A3, $\mathcal{G}_d$ is a cyclic group and, therefore, its order is bounded by the maximum element order: $|\mathcal{G}_d| \le v$. It follows that $\mathcal{G}_d = \langle\Theta\rangle$, which shows that every diagonal matrix in $\mathcal{G}$ is in $\langle\Theta\rangle$. For any off-diagonal matrix $G \in \mathcal{G}$, note that $R_1^{-1}G$ is diagonal and therefore contained in $\langle\Theta\rangle$. Thus every off-diagonal matrix in $\mathcal{G}$ is contained in $R_1\langle\Theta\rangle \subset \mathcal{G}_1$. We conclude that $\mathcal{G}_1$ consists of all of the diagonal and off-diagonal matrices in $\mathcal{G}$.

Since $|\mathcal{G}| > |\mathcal{G}_1|$, it follows from the Sylow Theorem that $\mathcal{G}_1$ is normal in some larger subgroup $\mathcal{G}_2 \subset \mathcal{G}$. Let R_2 be any element in $\mathcal{G}_2 - \mathcal{G}_1$. Since $\mathcal{G}_1$ is normal in $\mathcal{G}_2$, $R_2\Theta R_2^\dagger$ must be in $\mathcal{G}_1$ and is therefore either diagonal or off-diagonal. If $R_2\Theta R_2^\dagger$ is diagonal then, by Lemma A1*a)* with $x = \omega_v$ and $y = \omega_v^*$, R_2 is either diagonal or off-diagonal. However, this implies $R_2 \in \mathcal{G}_1$, which contradicts our choice of R_2. Hence, $R_2\Theta R_2^\dagger$ must be off-diagonal. By Lemma A1*b)*, this is possible only if $x = -y$, which requires $v = 4$. For $v = 4$, Lemma A8 implies that all elements of $\mathcal{G}$ have order 4, except $\pm I$. Moreover, every matrix of order 4 is a square root of $-I$. By Lemma A2, each nondiagonal root of $-I$ is of the form

$$j\begin{bmatrix} \alpha & d^* \\ d & -\alpha \end{bmatrix} \tag{32}$$

where α is real, d is complex, and $\alpha^2 + |d|^2 = 1$. Since $\Theta = \mathrm{diag}\{j, -j\}$ if $v = 4$, we see that the diagonal entries of R_2 and $R_2\Theta$ both have zero real part only if R_2 is off-diagonal ($\alpha = 0$). This implies $R_2 \in \mathcal{G}_1$, another contradiction. Since

both possibilities lead to contradictions, we conclude that $|\mathcal{G}| \leq 2v$. □

Conclusions: We now combine the results of the Appendix–D to prove the optimality of the unitary group codes given in Section IV. Recall from Appendix-B that $\nu_p = 2$ if and only if $\Lambda_p^* > 0$, in which case the minimum product distance is given by $\Lambda_p = \Lambda_p^*$.

For $M = 2$, there is only one group with $\Lambda_p^* > 0$: the $R = 0.5$ cyclic group in Table II. For $M = 4$, there exists no dicyclic group (cf. Lemma A7), so the two cyclic groups in Table II are both optimal. We prefer the $R = 1$ group in Table III, however, which is equivalent to the $(4, 3)$ cyclic group and takes values in BPSK. For $M = 8$, the $R = 1.5$ dicyclic group in Table IV is optimal. For $M = 16$, the best cyclic and dicyclic groups have the same minimum distance $\Lambda_p^* = 1.172$. Since the second-nearest neighbors are at a distance of 4 in the dicyclic group and 1.531 in the $(16, 7)$ and $(16, 9)$ cyclic groups, we choose the dicyclic group. For $M = 32$, the four cyclic groups in Table II are optimal, of which the $(32, 9)$ group has a particularly simple generator: $\Theta = \mathrm{diag}\{\omega_{32}, j\omega_{32}\}$.

ACKNOWLEDGMENT

The author is grateful to an anonymous referee for suggesting (10), which simplified the subsequent discussion of angular distance, and to Carmela Cozzo for producing Figs. 3 and 4 and for helpful discussions.

REFERENCES

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

[2] L. J. Cimini Jr., J. C.-I. Chuang, and N. R. Sollenberger, "Advanced cellular internet service (ACIS)," *IEEE Commun. Mag.*, vol. 36, pp. 150–159, Oct. 1998.

[3] H. S. M. Coxeter and W. O. J. Moser, *Generators and Relations for Discrete Groups*. New York: Springer-Verlag, 1980.

[4] D. Divsalar and M. K. Simon, "The design of trellis-coded MPSK for fading channels: Performance criteria," *IEEE Trans. Commun.*, vol. 36, pp. 1004–1012, Sept. 1988.

[5] "FRAMES multiple access proposal for the UMTS radio interface—SMG2," presented at the Workshop on UMTS Radio Technologies, Dec. 1996.

[6] G. D. Forney Jr., "Geometrically uniform codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1241–1260, Sept. 1991.

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

[8] G. J. Foschini and M. J. Gans, "On limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–355, Mar. 1998.

[9] J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," in *Proc. IEEE VTC'96*, 1996, pp. 136–140.

[10] ——, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," *IEEE Trans. Commun*, vol. 47, pp. 527–537, Apr. 1999.

[11] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, submitted for publication.

[12] B. M. Hochwald, T. L. Marzetta, T. L. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inform. Theory*, submitted for publication.

[13] B. M. Hochwald and W. Sweldens, Differential unitary space-time modulation, in *IEEE Trans. Commun.*, July 1999, submitted for publication.

[14] T. W. Hungerford, *Algebra*. New York: Springer-Verlag, 1974.

[15] L. Jalloul, K. Rohani, K. Kuchi, and J. Chen, "Performance analysis of CDMA transmit diversity methods," in *Proc. 1999 Veh. Tech. Conference (VTC'99)*, Fall 1999, pp. 1326–1330.

[16] 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.

[17] T. Mittelholzer and J. Lahtonen, "Group codes generated by finite reflection groups," *IEEE Trans. Inform. Theory*, vol. 42, pp. 519–527, Jan. 1996.

[18] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.

[19] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[20] G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communications," in *Proc. IEEE GLOBECOM'96*, 1996, pp. 1809–1814.

[21] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.

[22] M. R. Schroeder, *Number Theory in Science and Communication*, 3rd ed. New York: Springer, 1997.

[23] N. Seshadri and J. H. Winters, "Two signaling schemes for improving the error performance of frequency-division-duplex (FDD) transmission systems using transmitter antenna diversity," *Int. J. Wireless Inform. Networks*, vol. 1, no. 1, 1994.

[24] D. Slepian, "A class of binary signaling alphabets," *Bell Syst. Tech. J.*, vol. 35, pp. 203–234, 1956.

[25] ——, "Permutation modulation," *Proc. IEEE*, vol. 53, pp. 228–236, Mar. 1965.

[26] ——, "Group codes for the Gaussian channel," *Bell Syst. Tech. J.*, vol. 47, pp. 575–602, Apr. 1968.

[27] G. Strang, *Linear Algebra and Its Applications*, 3rd ed. New York: Harcourt-Brace-Jovanovich, 1988.

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

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

[30] 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.

[31] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE J. Select. Areas Commun.*, to be published.

[32] I. E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," AT&T Bell Labs, Internal Tech. Memo, June 1995.

[33] "Space-Time Block Coded Transmit Antenna Diversity for WCDMA," Texas Instuments Inc., Helsinki, Finland, UMTS SMG2-LI, Tech. doc. 662/98, Dec. 14–18, 1998.

[34] The CDMA 2000 Candidate Submission, TIA 45.5 Subcommittee, June 2, 1998. Draft.

[35] S. G. Wilson and Y. S. Leung, "Trellis-coded phase modulation on Rayleigh fading channels," in *Proc. IEEE ICC'87*, June 1987.

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

[37] ——, "Base station modulation diversity for digital SIMULCAST," in *Proc. IEEE VTC*, May 1993, pp. 505–511.

[38] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. ISSSE-98*, Sept. 29, 1998.

[39] L. H. Zetterberg, "A class of codes for polyphase signals on a bandlimited Gaussian channel," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 385–395, July 1965.