Differential Space–Time Modulation With Eigen-Beamforming for Correlated MIMO Fading Channels

Xiaodong Cai, *Senior Member, IEEE,* and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—In this paper, joint differential space–time modulation (DSTM) and eigen-beamforming for correlated multiple-input multiple-output (MIMO) fading channels. While DSTM does not require knowledge of each channel realization, the channel's spatial correlation can be easily estimated without training at the receiver and exploited by the transmitter to enhance the error probability performance. A transmission scheme is developed here that combines beamforming with differential multiantenna modulation based on orthogonal space–time block coding. Error probability is analyzed for both spatially correlated and independent Rayleigh fading channels. Based on the error probability analysis, power loading coefficients are derived to improve performance. The analytical and simulation results presented here corroborate that the proposed scheme can achieve considerable performance gain in correlated channels relative to DSTM without beamforming.**

***Index Terms*—Beamforming, differential optimal transmitter eigen-beamforming and space modulation, error probability analysis, optimal transmitter eigen-beamforming and space block codes.**

I. Introduction

RECENT advances in wireless communications show that multiantenna systems can support high data rates with low error probability [6], [23], [32]. Without any channel knowledge at the transmitter, space–time coding offers an effective fading countermeasure, and thereby reduces error probability [1], [29], [30]. If the receiver can acquire the channel state information (CSI) reliably, coherent detection along with space–time trellis coding [30], or orthogonal space–time block coding (STBC) [1], [29], or recent high-rate high-performance space–time coding [22] can be employed to enable the available spatial diversity. However, channel estimation becomes difficult or requires too many training symbols, when the channel is rapidly changing in a mobile environment, and/or when the number of transmit antennas is large. In such cases, unitary space–time modulation [11] and differential space–time modulation (DSTM) [7], [12], [14], [16], [31] are well motivated because they bypass CSI acquisition at the receiver.

Whenever (even partial) CSI is available at the transmitter, it should be exploited to further improve the performance of multiantenna transmission systems. Since in most cases the transmitter cannot acquire the CSI perfectly, utilization of partial CSI at the transmitter has received considerable attention recently. A general statistical model of partial CSI is presented in [18], [33]. Based on this model, a linear transformation was applied to orthogonal STBC to enhance the symbol error rate (SER) performance in [18]. Since in frequency-division duplex (FDD) systems the transmitter obtains channel knowledge from the feedback channel, two cases of partial CSI, termed *mean* and *covariance feedback*, were studied to maximize channel capacity in [33]; capacity maximization based on covariance feedback was also investigated for multiple-input multiple-output (MIMO) systems in [15]. Optimal beamformers combined with *coherent* STBC were derived in [34] and [35] to minimize symbol error probability, based on channel mean and correlation, respectively. Transmit-beamformer optimization in terms of average signal-to-noise ratio (SNR) and expected mutual information based on partial CSI was studied in [24]. Note that all these works assume that perfect or estimated CSI is available at the receiver.

In this paper, we consider DSTM based on orthogonal STBC, which was also investigated in [7], [8], [16], [31] *without* partial CSI at the transmitter for independent, identically distributed (i.i.d.) fading channels. Since practical multiantenna systems may exhibit strong correlation among fading channels associated with different transmit antennas on the downlink [15], [26], we will focus on spatially *correlated* channels. The channel's spatial correlations depend on the mobile's angular position with respect to the transmitter [4], [26]. This implies that the channel's spatial correlations will typically change slowly, even when the channel coefficients fluctuate relatively fast [10]. While knowledge of each channel realization is not available in a differential space–time transmission system, the channel's spatial correlations can be estimated at the receiver without training and be fed back to the transmitter. We will exploit these channel correlations at the transmitter to combine DSTM with eigen-beamforming. We will also analyze the error performance of DSTM, which is applicable to both i.i.d. and

Manuscript received July 29, 2004; revised May 19, 2005. This paper appeared in part in the *Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 2003*, Hong Kong, April 2003. This work was prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U. S. Government. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yucel Altunbasak.

X. Cai is with the Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33124 USA (e-mail: x.cai@miami.edu).

G. B. Giannakis is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: georgios@ece.umn.edu).

Digital Object Identifier 10.1109/TSP.2006.870637

spatially correlated channels with or without beamforming. Note that error probability performance of DSTM based on orthogonal STBC was not analyzed in [7], [8], [16], and [31]. Our general error probability formula is applicable to DSTM with beamforming in correlated channels investigated in this paper and DSTM without beamforming in i.i.d. channels considered in [7], [8], [16], and [31]. Based on our novel error probability analysis, we will determine the optimal power allocation per beam, and thereby enhance the error performance via loading. An approximate expression for the bit error rate (BER) was derived in [9] for STBC-based DSTM over i.i.d. Rayleigh fading channels. Pairwise error probability and BER were analyzed for the DSTM of [12] and [31], respectively, for i.i.d. Rayleigh fading channels in [5]. Our error performance analysis for STBC-based DSTM is applicable to both i.i.d. and spatially correlated Rayleigh fading channels.

The rest of the paper is organized as follows. Section II presents the signal model and differential reception of multiantenna systems with DSTM and eigen-beamforming. Error probability analysis is carried out and an optimal power loading scheme is developed in Section III. Section IV contains analytical results and simulations to test the merits of the proposed transmission scheme. Finally, conclusions are drawn in Section V.

Notation: Superscripts T and $\mathcal{H}$ stand for transpose, conjugate, Hermitian transpose, respectively; $E[\,]$ denotes expectation over the random variables within the brackets; $\mathrm{Re}(x)$ and $\mathrm{Im}(x)$ represent the real and imaginary parts of x, respectively. Column vectors (matrices) are denoted by boldface lower (upper) case letters. We will use $\mathbf{I}_N$ to denote the $N \times N$ identity matrix, $\mathrm{Tr}(\mathbf{A})$ the trace of $\mathbf{A}$, $\det(\mathbf{A})$ the determinant of $\mathbf{A}$, and $[\mathbf{A}]_{m,n}$ the (m,n)th entry of $\mathbf{A}$.

II. Differential Space–Time Modulation With Eigen-Beamforming

Consider a multiantenna transmission system comprising N_T transmit antennas and N_R receive antennas, signaling over a Rayleigh flat-fading channel. Suppose that the base station (BS) and the mobile user assume the roles of transmitter and receiver, respectively. Let $h_{m,n}$ denote the channel coefficient between the mth transmit and the nth receive antenna, which is modeled as a complex Gaussian random variable with zero mean. The channel matrix is then represented by $[\mathbf{H}]_{m,n} := h_{m,n}$. In a wireless environment where the BS is elevated and unobstructed and the mobile user is surrounded by local scatterers, a ray tracing model is often used to describe the channel [4], [15], [26], [17, p. 60–65]. For this channel model, it has been shown that in practical settings with reasonable antenna spacing, channel gains associated with different transmit antennas exhibit strong correlations, while those associated with different receive antennas are uncorrelated [15]. Let $\mathbf{h}_i$ stand for the ith column of the channel matrix $\mathbf{H}$. Then, $\mathbf{h}_i$ and $\mathbf{h}_j$ are uncorrelated when $i \neq j$, while elements of $\mathbf{h}_1, \forall i$, are correlated. We define the channel correlation matrix as $\mathbf{R}_h := E[\mathbf{h}_i \mathbf{h}_i^{\mathcal{H}}], \forall i$.

While DSTM can be implemented using any unitary matrix constellations [12], [14], we will consider DSTM based on orthogonal STBC as in [7], because the orthogonal STBC matrix is easily constructed and leads to low-complexity symbol-by-symbol decoding. Specifically, the P symbols transmitted in the tth block are first collected in an $N \times N$ orthogonal space–time code matrix as follows [7]:

$$\mathbf{S}_t = \frac{1}{\sqrt{P}} \sum_{p=1}^{P} \left(\mathbf{\Phi}_p s_{t,p}^R + j \mathbf{\Psi}_p s_{t,p}^I \right) \quad t > 0, \tag{1}$$

where $s_{t,p}^R$ and $s_{t,p}^I$ are real and imaginary parts of the complex symbol $s_{t,p}$, respectively, i.e., $s_{t,p} = s_{t,p}^R + j s_{t,p}^I$, and $N \times N$ matrices $\mathbf{\Phi}_p$ and $\mathbf{\Psi}_p$ satisfy the following conditions:

$$\begin{aligned} &\mathbf{\Phi}_p^{\mathcal{H}} \mathbf{\Phi}_p = \mathbf{I}_N, \quad \mathbf{\Psi}_p^{\mathcal{H}} \mathbf{\Psi}_p = \mathbf{I}_N, \quad \forall p \\ &\mathbf{\Phi}_p^{\mathcal{H}} \mathbf{\Phi}_q + \mathbf{\Phi}_q^{\mathcal{H}} \mathbf{\Phi}_p = \mathbf{0}, \quad \mathbf{\Psi}_p^{\mathcal{H}} \mathbf{\Psi}_q + \mathbf{\Psi}_q^{\mathcal{H}} \mathbf{\Psi}_p = \mathbf{0}, \quad p \neq q \\ &\mathbf{\Phi}_p^{\mathcal{H}} \mathbf{\Psi}_q - \mathbf{\Psi}_q^{\mathcal{H}} \mathbf{\Phi}_p = \mathbf{0}, \quad \forall p, q. \end{aligned} \tag{2}$$

Drawing $s_{t,p}$ from M-ary phase-shift keying (M-PSK) constellations, and letting $|s_{t,p}| = 1$, it follows readily from (1) and (2) that the matrix $\mathbf{S}_t$ is unitary, i.e., $\mathbf{S}_t^{\mathcal{H}} \mathbf{S}_t = \mathbf{I}_N$. The $N \times N$ code matrix $\mathbf{C}_t$ for DSTM can then be written recursively as [7], [12], [14]

$$\mathbf{C}_t = \mathbf{S}_t \mathbf{C}_{t-1}, \quad t > 0 \tag{3}$$

with $\mathbf{C}_0 = \mathbf{I}_N$. Since $\mathbf{S}_t$ is unitary, matrix $\mathbf{C}_t$ is unitary too, by design. When the number of transmit antennas $N_T = N$, then N elements of each row of $\mathbf{C}_t$ are transmitted from N_T antennas per time slot. For an arbitrary N_T, however, we may not be able to find a square matrix $\mathbf{S}_t$ with $N = N_T$. In this case, we can construct a square code matrix $\mathbf{S}_t$ with $N > N_T$, find $\mathbf{C}_t$ using (3), and transmit the first N_T columns of $\mathbf{C}_t$ from N_T transmit antennas as in [7]. Mathematically, letting $\mathbf{\Theta} = [\mathbf{I}_{N_T} \mathbf{0}_{N_T \times (N - N_T)}]^T$, the matrix codeword transmitted over N_T antennas in the tth block is $\mathbf{C}_t \mathbf{\Theta}$.

If $h_{m,n}$ are i.i.d., it is natural to transmit the codeword $\mathbf{C}_t \mathbf{\Theta}$ via N_T transmit antennas with equal power. However, when the channel coefficients associated with different transmit antennas are correlated, which is of primary interest in this paper, a modulation scheme exploiting this channel correlation at the transmitter is well motivated. To this end, we will transmit the codeword $\mathbf{C}_t \mathbf{\Theta}$ along the eigenvectors of the channel correlation matrix $\mathbf{R}_h$ with proper power loaded on each eigenvector. This transmission scheme, termed eigen-beamforming, was introduced in [35] for *coherent* STBC over correlated fading channels. The eigen-decomposition of $\mathbf{R}_h$ can be written as $\mathbf{R}_h = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^{\mathcal{H}}$, where the diagonal matrix $\mathbf{\Lambda}$ contains the ordered eigenvalues of $\mathbf{R}_h$ and the unitary matrix $\mathbf{U}$ consists of the corresponding eigenvectors. The transmitted signal during the tth block can then be expressed by the $N \times N_T$ matrix

$$\mathbf{X}_t = \sqrt{P \mathcal{E}_s} \mathbf{C}_t \mathbf{\Theta} \mathbf{D} \mathbf{U}^{\mathcal{H}} \tag{4}$$

where the diagonal matrix $\mathbf{D}$ contains power loading coefficients which will be specified later in Section III-C. In this section, the only constraint we impose on the power loading coefficients is $\sum_{i=1}^{N_t} [\mathbf{D}]_{i,i}^2 = 1$. Using this constraint, we can verify that $\mathrm{Tr}(\mathbf{X}_t \mathbf{X}_t^{\mathcal{H}}) = P \mathcal{E}_s$, where $\mathcal{E}_s$ stands for the energy per

transmitted symbol. We can also write (4) as a recursion initialized by $\mathbf{X}_0 = \sqrt{P\mathcal{E}_s}\boldsymbol{\Theta}\mathbf{D}\mathbf{U}^{\mathcal{H}}$ [cf. (3)]

$$\mathbf{X}_t = \mathbf{S}_t\mathbf{X}_{t-1}, \quad t > 0. \tag{5}$$

Comparing (3) with (5) reveals that the fundamental differential transmission equation is not changed by the loaded transmit eigen-beamforming matrices $\mathbf{D}\mathbf{U}^{\mathcal{H}}$. From (5), it is seen that we need to beamform and power load only in the first transmitted block $\mathbf{X}_0$; and after the first block, signals will be automatically transmitted along eigen-beams without any beamforming operation.

For clarity, we first consider a single receive antenna ($N_R = 1$). The received samples in the tth block can be written in an $N \times 1$ vector $\mathbf{y}_t$ as

$$\mathbf{y}_t = \mathbf{X}_t\mathbf{h} + \mathbf{w}_t \tag{6}$$

where $\mathbf{w}_t$ contains complex additive white Gaussian noise (AWGN) with mean zero and variance $N_0/2$ per dimension. We will detect $\mathbf{S}_t$ based on $\mathbf{y}_{t-1}$ and $\mathbf{y}_t$. As it is common to all differential schemes, we assume that the channel variation is negligible from one space–time block to the next, even though over a large number of blocks the cumulative variation can be indeed fast. For this reason, we omit the time index of $\mathbf{h}$ in (6) for notational brevity.

Defining $\mathbf{y} := [\mathbf{y}_{t-1}^T\ \mathbf{y}_t^T]^T$, we will detect $\mathbf{S}_t$ based on $\mathbf{y}$. Conditioned on $\mathbf{C}_{t-1}$ and $\mathbf{C}_t$, the $2N \times 1$ vector $\mathbf{y}$ is Gaussian with mean zero and covariance matrix

$$\mathbf{R}_y = \begin{bmatrix} \mathbf{X}_{t-1}\mathbf{R}_h\mathbf{X}_{t-1}^{\mathcal{H}} & \mathbf{X}_{t-1}\mathbf{R}_h\mathbf{X}_t^{\mathcal{H}} \\ \mathbf{X}_t\mathbf{R}_h\mathbf{X}_{t-1}^{\mathcal{H}} & \mathbf{X}_t\mathbf{R}_h\mathbf{X}_t^{\mathcal{H}} \end{bmatrix} + N_0\mathbf{I}_{2N}. \tag{7}$$

Hence, the conditional probability density function (pdf) of $\mathbf{y}$ is given by

$$f(\mathbf{y}|\mathbf{C}_{t-1}, \mathbf{C}_t) = \frac{\exp\left(-\mathbf{y}^{\mathcal{H}}\mathbf{R}_y^{-1}\mathbf{y}\right)}{\pi^{2N}\det(\mathbf{R}_y)}. \tag{8}$$

If all symbols are equally likely, the optimum receiver is the maximum-likelihood (ML) detector which finds the $\mathbf{S}_t$ matrix that maximizes the pdf in (8). When the channels associated with different transmit antennas are i.i.d. ($\mathbf{R}_h = \mathbf{I}_N$), the ML detector can be derived from (8) as detailed in [14]. If the channels are correlated and/or $\boldsymbol{\Theta} \neq \mathbf{I}_N$, the likelihood function in (8) is dependent on $\mathbf{C}_{t-1}$ and C_t; and for this reason, the ML detector is not practically feasible.

Instead of the ML criterion,we will rely on the fundamental differential receiver equation

$$\mathbf{y}_t = \mathbf{S}_t\mathbf{y}_{t-1} + \mathbf{w}_t - \mathbf{S}_t\mathbf{w}_{t-1} \tag{9}$$

to derive our receiver, which leads to the decision rule [7], [12]

$$\hat{\mathbf{S}}_t = \arg\max_{\mathbf{S}} \mathrm{Re}\left(2\mathbf{y}_{t-1}^{\mathcal{H}}\mathbf{S}^{\mathcal{H}}\mathbf{y}_t\right). \tag{10}$$

Although (10) coincides with the ML detector in [14] that was derived for i.i.d. channels, it is clearly no longer ML optimum for correlated channels. Due to the code structure in (1), the detector (10) for the codeword $\mathbf{S}_t$ reduces to a symbol-by-symbol detector [7]

$$\hat{s}_{t,p} = \arg\max_s \Big[\mathrm{Re}\left(2\mathbf{y}_{t-1}^{\mathcal{H}}\boldsymbol{\Phi}_p^{\mathcal{H}}\mathbf{y}_t\right)\mathrm{Re}(s) + \mathrm{Re}\left(-j2\mathbf{y}_{t-1}^{\mathcal{H}}\boldsymbol{\Psi}_p^{\mathcal{H}}\mathbf{y}_t\right)\mathrm{Im}(s)\Big]. \tag{11}$$

Defining $z_{p,R} := \mathrm{Re}(2\mathbf{y}_{t-1}^{\mathcal{H}}\boldsymbol{\Phi}_p^{\mathcal{H}}\mathbf{y}_t)$, $z_{p,I} := \mathrm{Re}(-j2\mathbf{y}_{t-1}^{\mathcal{H}}\boldsymbol{\Psi}_p^{\mathcal{H}}\mathbf{y}_t)$ and $z_p := z_{p,R} + jz_{p,I}$, we can also write the decision rule in (11) as

$$\hat{s}_{t,p} = \arg\max_s \mathrm{Re}\left(z_p s^*\right). \tag{12}$$

When $N_R > 1$ receive antennas are employed, the received signal in the tth block can be written in an $N \times N_R$ matrix $\mathbf{Y}_t$ as

$$\mathbf{Y}_t = \mathbf{X}_t\mathbf{H} + \mathbf{W}_t. \tag{13}$$

Using the fact that the noise $\mathbf{W}_t$ is white, we can derive a decision rule similar to (10), which is given by

$$\begin{aligned}\hat{\mathbf{S}}_t &= \arg\max_{\mathbf{S}} \sum_{n=1}^{N_R} \mathrm{Re}\left(2\mathbf{y}_{t-1,n}^{\mathcal{H}}\mathbf{S}^{\mathcal{H}}\mathbf{y}_{t,n}\right) \\ &= \arg\max_{\mathbf{S}} \mathrm{Tr}\,\mathrm{Re}\left(2\mathbf{Y}_{t-1}^{\mathcal{H}}\mathbf{S}^{\mathcal{H}}\mathbf{Y}_t\right)\end{aligned} \tag{14}$$

where $\mathbf{y}_{t,n}$ is the nth column of $\mathbf{Y}_t$ in (13) and the symbol-by-symbol detector is easily derived from (14). Since channel gains associated with different receive antennas are uncorrelated, the error probability results for the single receive antenna generalize easily to multiple receive antennas. For this reason and without loss of generality, we will concentrate on the one receive antenna case in the remaining of this paper.

III. Performance Analysis and Power Loading

DSTM based on orthogonal STBC was studied for i.i.d. channels in [7], [8], [16], and [31], but error probability analysis was not provided. In this section, we will analyze error probability performance of DSTM, which is applicable to DSTM with beamforming over spatially correlated channels considered in this paper and DSTM without beamforming over i.i.d. channels studied in [7], [8], [16], and [31],. For binary phase-shift keying (BPSK) and quadrature phase-shift keying (QPSK), we will provide formulas to calculate the exact BER. However, it is difficult to obtain the exact BER or SER in closed form for any M-PSK other than BPSK and QPSK. For this reason, we will also derive an approximate SER for any M-PSK constellation, which will come handy in deriving the power loading algorithm.

A. Exact BER for BPSK and QPSK

For BPSK constellations, $s = \pm 1$, the decision variable in (12) is simplified as

$$\hat{s}_{t,p} = \mathrm{sign}(z_{p,R}) \tag{15}$$

while for QPSK constellations, $s=(\pm 1 \pm j)/\sqrt{2}$, (12) reduces to

$$\begin{aligned}\hat{s}_{t,p}^{R} &= \operatorname{sign}(z_{p,R}) \\ \hat{s}_{t,p}^{I} &= \operatorname{sign}(z_{p,I}).\end{aligned} \tag{16}$$

If $s_{t,p}^{R}$ takes ± 1 values with equal probability, the BER of $s_{t,p}^{R}$ is given by

$$P_b(e) = P\left(z_{p,R} < 0 | s_{t,p}^{R} = 1\right). \tag{17}$$

It will be shown later that this error probability is the same for $s_{t,p}^{R}$ and $s_{t,p}^{I}$, $\forall p$; thus, it corresponds also to the overall BER. The decision variable $z_{p,R}$ can be expressed as

$$z_{p,R} = \mathbf{y}_{t-1}^{\mathcal{H}} \boldsymbol{\Phi}_p^{\mathcal{H}} \mathbf{y}_t + \mathbf{y}_t^{\mathcal{H}} \boldsymbol{\Phi}_p \mathbf{y}_{t-1} = \mathbf{y}^{\mathcal{H}} \bar{\boldsymbol{\Phi}}_p \mathbf{y} \tag{18}$$

where

$$\bar{\boldsymbol{\Phi}}_p := \begin{bmatrix} \mathbf{0} & \boldsymbol{\Phi}_p^{\mathcal{H}} \\ \boldsymbol{\Phi}_p & \mathbf{0} \end{bmatrix}. \tag{19}$$

It is seen from (18) that $z_{p,R}$ is a quadratic form of the complex Gaussian random vector $\mathbf{y}$. Thus, the Laplace transform of the pdf of $z_{p,R}$ is given by [25, p. 595]

$$\begin{aligned}\phi(\omega) &:= E\left[\exp(-\omega z_{p,R})\right] \\ &= \frac{1}{\det(\mathbf{I} + \omega \mathbf{R}_y \bar{\boldsymbol{\Phi}}_p)} \\ &= \frac{1}{\prod_{i=1}^{2N}(1+\omega\lambda_i)}\end{aligned} \tag{20}$$

where λ_i is the ith eigenvalue of matrix $\mathbf{A} := \mathbf{R}_y \bar{\boldsymbol{\Phi}}_p$. Then, the error probability can be found as [2], [3]

$$P_b(e) = -\sum_{\omega_i > 0} \operatorname{Res}\left[\frac{\phi(\omega)}{\omega}; \omega_i\right] \tag{21}$$

where ω_i is a pole of $\phi(\omega)/\omega$, and $\operatorname{Res}[f(x); x_i]$ denotes the residue of $f(x)$ at x_i. From (20), we have $\omega_i = -1/\lambda_i$, if $\lambda_i \neq 0$. To evaluate the BER in (21), we need to find the eigenvalues of $\mathbf{A}$.

Letting $\mathbf{D}_h := P\mathcal{E}_s \boldsymbol{\Theta} \mathbf{D}^2 \boldsymbol{\Lambda} \boldsymbol{\Theta}^{\mathcal{H}}$, the covariance matrix $\mathbf{R}_y$ in (7) can be written as

$$\mathbf{R}_y = \begin{bmatrix} \mathbf{C}_{t-1}\mathbf{D}_h\mathbf{C}_{t-1}^{\mathcal{H}} & \mathbf{C}_{t-1}\mathbf{D}_h\mathbf{C}_t^{\mathcal{H}} \\ \mathbf{C}_t\mathbf{D}_h\mathbf{C}_{t-1}^{\mathcal{H}} & \mathbf{C}_t\mathbf{D}_h\mathbf{C}_t^{\mathcal{H}} \end{bmatrix} + N_0\mathbf{I}_{2N}. \tag{22}$$

From (22), it appears that $\mathbf{A}$ depends on code matrices $\mathbf{C}_{t-1}$ and $\mathbf{C}_t$, implying that its eigenvalues may depend on all transmitted symbols. However, it will turn out that the eigenvalues of $\mathbf{A}$ are only determined by $s_{t,p}^{R}$, which will enable evaluation of the BER in (21). Define the block diagonal matrix $\mathbf{C} := \operatorname{diag}(\mathbf{C}_{t-1}, \mathbf{C}_t)$, which can be readily verified to be unitary. Then, we have

$$\tilde{\mathbf{A}} := \mathbf{C}^{\mathcal{H}}\mathbf{A}\mathbf{C} = \mathbf{C}^{\mathcal{H}}\mathbf{R}_y\mathbf{C}\mathbf{C}^{\mathcal{H}}\bar{\boldsymbol{\Phi}}_p\mathbf{C} = \bar{\boldsymbol{\Gamma}}\mathbf{B} \tag{23}$$

where

$$\bar{\boldsymbol{\Gamma}} := \begin{bmatrix} \mathbf{D}_h & \mathbf{D}_h + N_0\mathbf{I}_N \\ \mathbf{D}_h + N_0\mathbf{I}_N & \mathbf{D}_h \end{bmatrix}, \tag{24}$$

$\mathbf{B} := \operatorname{diag}(\mathbf{B}_1, \mathbf{B}_1^{\mathcal{H}})$ with $\mathbf{B}_1 := \mathbf{C}_t^{\mathcal{H}}\boldsymbol{\Phi}_p\mathbf{C}_{t-1}$, and since $\mathbf{C}_{t-1}$, $\mathbf{C}_t$, and $\boldsymbol{\Phi}_p$ are unitary, it is easy to verify that $\mathbf{B}_1^{\mathcal{H}}\mathbf{B}_1 = \mathbf{I}_N$. To proceed, we need the following lemma regarding the eigenvalues of $\mathbf{B}_1$ that we prove in Appendix I.

Lemma 1: If $\alpha := 2s_{t,p}^{R}/\sqrt{P}$, then $\mathbf{B}_1$ has two distinct eigenvalues given by $\lambda_{B,1} = (\alpha + j\sqrt{|\alpha^2 - 4|})/2$ and $\lambda_{B,2} = (\alpha - j\sqrt{|\alpha^2 - 4|})/2$.

Note that we have $\lambda_{B,2} = \lambda_{B,1}^{*}$ and $\lambda_{B,2}\lambda_{B,1} = 1$. Due to the special structure of the code matrix $\mathbf{S}_t$, theeigenvalues of $\mathbf{B}_1$ are determined by $s_{t,p}^{R}$, and are not affected by other symbols. The following proposition provides the key result to BER analysis.

Proposition 1: The matrix $\mathbf{A}$ is similar to

$$\mathbf{G} := \begin{bmatrix} \lambda_{B,1}\mathbf{D}_h & \lambda_{B,2}(\mathbf{D}_h + N_0\mathbf{I}_N) \\ \lambda_{B,1}(\mathbf{D}_h + N_0\mathbf{I}_N) & \lambda_{B,2}\mathbf{D}_h \end{bmatrix}. \tag{25}$$

Proof: See Appendix II.

Because $\mathbf{D}_h$ is a constant matrix, $\mathbf{G}$ is determined by $s_{t,p}^{R}$ through $\lambda_{B,1}$ and $\lambda_{B,2}$. The key step in proving this proposition is to apply the permutation matrix $\mathbf{P}$ and obtain (41). It follows from the proof that this proposition is true for any matrix $\mathbf{D}_h$. Hence, if beamforming and power loading are not employed, then $\mathbf{D}_h = P\mathcal{E}_s\boldsymbol{\Theta}\mathbf{R}_h\boldsymbol{\Theta}^{\mathcal{H}}$, and the proposition still holds.

Corollary 1: With $D_{hi} = [\mathbf{D}_h]_{i,i}$, the $2N$ eigenvalues of $\mathbf{A}$ are given by

$$\lambda_i = \frac{1}{2}\left(\alpha D_{hi} \pm \sqrt{\alpha^2 D_{hi}^2 + 4\left(N_0^2 + 2D_{hi}N_0\right)}\right), \quad i = 1, \ldots, N. \tag{26}$$

Proof: See Appendix III.

Given these eigenvalues, we are ready to evaluate the BER using (21). Let us consider the decision variable $z_{p,I}$ in (16). Defining

$$\bar{\boldsymbol{\Psi}}_p := \begin{bmatrix} \mathbf{0} & (j\boldsymbol{\Psi}_p)^{\mathcal{H}} \\ j\boldsymbol{\Psi}_p & \mathbf{0} \end{bmatrix} \tag{27}$$

we can also write $z_{p,I}$ in a quadratic form as $z_{p,I} = \mathbf{y}^{\mathcal{H}}\bar{\boldsymbol{\Psi}}_p\mathbf{y}$. Using (1) and (2), we can show that

$$\mathbf{S}_t^{\mathcal{H}}(j\boldsymbol{\Phi}_p) + (j\boldsymbol{\Phi}_p)^{\mathcal{H}}\mathbf{S}_t = \beta\mathbf{I}_N \tag{28}$$

where $\beta := 2s_{t,p}^{I}/\sqrt{P}$. If wereplace $\bar{\boldsymbol{\Phi}}_p$ in $\mathbf{A}$ and $\mathbf{B}_1$ by $\bar{\boldsymbol{\Psi}}_p$, replace α by β, and use (28), then Lemma 1, Proposition 1 and Corollary 1 also hold true for $z_{p,I}$. Therefore, the error probability given by (21) is the overall BER for BPSK or QPSK.

If channel gains are i.i.d., then $\mathbf{A}$ has only two distinct eigenvalues, each with multiplicity N. When N is large, it may be complicated to find the residue of $\phi(\omega)/\omega$ when calculating BER from (21); in this case, a numerical method can be used to evaluate BER by calculating the inverse Laplace transform of $\phi(\omega)/\omega$ [3]. If beamforming and power loading are not used, Lemma 1 and Proposition 1 still hold, and we can calculate the eigenvalues of $\mathbf{G}$ numerically, which in turn enables us to evaluate the BER. If beamforming and equal power loading are employed, then $\mathbf{G}$ has eigenvalues identical to those without beamforming and power loading. Therefore, these two cases have identical error probability performance, which implies that beamforming alone does not improve error performance. This

motivates appropriate power loading coefficients to reduce the error probability—a subject we will pursue later.

B. Approximate SER for General M-PSK

If $\tilde{\mathbf{h}} := \mathbf{D}\mathbf{U}^{\mathcal{H}}\mathbf{h}$ and $\tilde{\mathbf{C}}_t := \mathbf{C}_t\mathbf{\Theta}$, then $z_{p,R}$ and $z_{p,I}$ in (12) can be expressed as

$$\begin{aligned} z_{p,R} &= 2\sqrt{\mathcal{E}_s}|\tilde{\mathbf{h}}|^2 s_p^R + w_1 + w_3 + w_5 \\ z_{p,I} &= 2\sqrt{\mathcal{E}_s}|\tilde{\mathbf{h}}|^2 s_p^I + w_2 + w_4 + w_6 \end{aligned} \tag{29}$$

where $w_1 := 2\mathrm{Re}[\mathbf{w}_{t-1}^{\mathcal{H}}\mathbf{\Phi}_p^{\mathcal{H}}\mathbf{S}_t\tilde{\mathbf{C}}_{t-1}\tilde{\mathbf{h}}]$, $w_2 := 2\mathrm{Re}[\mathbf{w}_{t-1}^{\mathcal{H}}(j\mathbf{\Psi}_p)^{\mathcal{H}}\mathbf{S}_t\tilde{\mathbf{C}}_{t-1}\tilde{\mathbf{h}}]$, $w_3 := 2\mathrm{Re}[\tilde{\mathbf{h}}^{\mathcal{H}}\tilde{\mathbf{C}}_{t-1}^{\mathcal{H}}\mathbf{\Phi}_p^{\mathcal{H}}\mathbf{w}_t]$, $w_4 := 2\mathrm{Re}[\tilde{\mathbf{h}}^{\mathcal{H}}\tilde{\mathbf{C}}_{t-1}^{\mathcal{H}}(j\mathbf{\Psi}_p)^{\mathcal{H}}\mathbf{w}_t]$, $w_5 := 2\mathrm{Re}[\mathbf{w}_{t-1}^{\mathcal{H}}\mathbf{\Phi}_p^{\mathcal{H}}\mathbf{w}_t]$, and $w_6 := 2\mathrm{Re}[\mathbf{w}_{t-1}^{\mathcal{H}}(j\mathbf{\Psi}_p)^{\mathcal{H}}\mathbf{w}_t]$. When the SNR is high, w_5 and w_6 are negligible; thus, $z_p = z_{p,R} + jz_{p,I}$ can be well approximated by

$$\tilde{z}_p = 2\sqrt{\mathcal{E}_s}|\tilde{\mathbf{h}}|^2 s_p + w \tag{30}$$

where $w = w_1 + w_3 + j(w_2 + w_4)$. Based on (30), it is argued in [7] and [31] that differential modulation incurs 3-dB penalty in SNR relative to orthogonal STBC with coherent detection, since the noise power is doubled. However, SER formulas for coherent modulation cannot be readily applied to calculate an approximate SER for differential modulation based on (30), since the phase of $\tilde{z}_p$ is disturbed by the noise w and we do not know whether the real and the imaginary parts of w are correlated or not. The following fact which we prove in Appendix IV leads to a useful approximate SER expression.

Fact 1: $\mathrm{Re}(w)$ and $\mathrm{Im}(w)$ are uncorrelated and have identical variance $\sigma^2 = 4|\tilde{\mathbf{h}}|^2 N_0$.

Based on this fact, we see that the signal model (30) is the same as that of M-PSK modulation with maximum ratio combining (MRC) [27, p. 266]. This implies that we can approximately compute the SER using [27, p. 271]

$$P_s(e) = \frac{1}{\pi}\int_0^{\frac{(M-1)\pi}{M}} \mathcal{M}\left(-\frac{\mathrm{g_{PSK}}}{\sin^2\theta}\right) d\theta \tag{31}$$

where $\mathrm{g_{PSK}} := \sin^2(\pi/M)$, and $\mathcal{M}(\cdot)$ is the moment-generating function (MGF) of the random variable $\mathcal{E}_s|\tilde{\mathbf{h}}|^2/(2N_0)$. For Rayleigh fading, a closed SER formcan be obtained from (31) by using the partial fraction expansion of the MGF. Since the instantaneous SNR of $\tilde{z}_p$ is $\mathcal{E}_s|\tilde{\mathbf{h}}|^2/(2N_0)$, the approximate SER derived from (30) is 3 dB worse than that of coherent modulation with the same transmitted power. For BSPK and QPSK, we can also consider $z_{p,R}$ only, and derive an approximate BER. This approach will be used in Section IV to calculate the approximate BER for QPSK.

C. Power Loading

Here, we will derive power loading coefficients contained in the diagonal matrix $\mathbf{D}$ in (4). Our goal is to reduce error probability by exploiting the knowledge of $\mathbf{R}_h$. Although exact BER expressions for BPSK and QPSK constellations of Section III-A are useful for benchmarking performance, they are difficult to minimize with respect to $\mathbf{D}$. For this reason, we will derive power loading coefficients based on the approximate SER of Section III-B. While $P_s(e)$ in (31) offers a good approximation of the actual SER at high SNR as shown in Section III-B, we will demonstrate in Section IV that the approximate SER comes also very close to the actual SER at low SNR. Hence, reducing this approximate SER will decrease the actual SER. Similar to [18], [35], we will minimize the Chernoff bound of the approximate SER in (31) which can be found as

$$P_{s,\mathrm{bound}}(e) = \frac{M-1}{M}\frac{1}{\prod_{i=1}^{N_T}\left[1 + \frac{\mathrm{g_{PSK}}\mathcal{E}_s\lambda_{R,i}[\mathbf{D}]_{i,i}^2}{(2N_0)}\right]} \tag{32}$$

where $\lambda_{R,i}$ is the eigenvalue of the channel covariance matrix $\mathbf{R}_h$. To select power loading coefficients, we now formulate the following optimization problem:

$$\begin{aligned} \max_{\mathbf{D}} \quad & \sum_{i=1}^{N_T}\log\left[1 + \frac{\mathrm{g_{PSK}}\mathcal{E}_s\lambda_{R,i}[\mathbf{D}]_{i,i}^2}{(2N_0)}\right] \\ \text{subject to} \quad & \sum_{i=1}^{N_t}[\mathbf{D}]_{i,i}^2 = 1. \end{aligned} \tag{33}$$

Using the Lagrange multiplier method, we can find power loading coefficients as follows:

$$[\mathbf{D}]_{i,i}^2 = \frac{1}{\bar{N}_T} + \frac{2N_0}{\mathrm{g_{PSK}}\mathcal{E}_s}\left(\frac{1}{\bar{N}_T}\sum_{l=1}^{\bar{N}_T}\frac{1}{\lambda_{R,l}} - \frac{1}{\lambda_{R,i}}\right) \tag{34}$$

where $\bar{N}_T$ $(0 < \bar{N}_T \le N_T)$ is the number of beamformers that transmit signals, given the transmitted power budget $\mathcal{E}_s$. Note that the solution in (34) is similar to the optimal power loading for *coherent* STBC of [35] except for a factor 2 in the second term, because the $P_{s,\mathrm{bound}}(e)$ in (32) is 3 dB larger than its counterpart in the coherent STBC of [35]. For the selection of $\bar{N}_T$ and detailed description of the power-loading algorithm, we refer the reader to [35].

Orthogonal STBC is well appreciated for its low decoding complexity, coding, and diversity gains [29]. While the works in [7], [8], [16], and [31] show that orthogonal STBC can be modified to facilitate differential modulation and detection, our work here reveals that the loaded eigen-beamforming derived in [35] for coherent STBC, can also be used in a differential STBC setup.

IV. Simulations and Numerical Results

We consider a linear array of $N_T = 4$ equispaced (by d) antennas at the transmitter, and $N_R = 1$ antenna at the receiver. We assume that the direction of departure is perpendicular to the transmitter antenna array. Let λ be the wavelength of the transmitted signal, and Δ denote the angle spread. When Δ is small, channel's spatial correlation can be calculated from the "one-ring" channel model as [26]

$$[\mathbf{R}_h]_{m,n} \approx \frac{1}{2\pi}\int_0^{2\pi}\exp\left[\frac{-j2\pi(m-n)d\Delta\sin\theta}{\lambda}\right] d\theta. \tag{35}$$

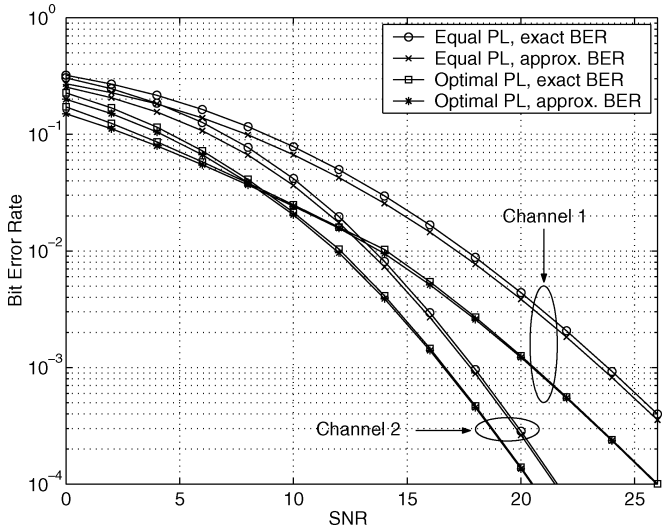

Fig. 1. Exact and approximate BERs of QPSK with known $\mathbf{R}_h$.

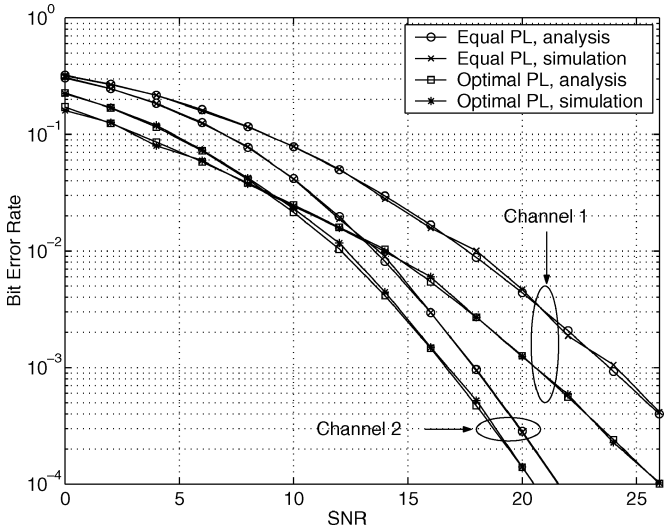

Fig. 2. Simulated and exact BERs of QPSK with known $\mathbf{R}_h$.

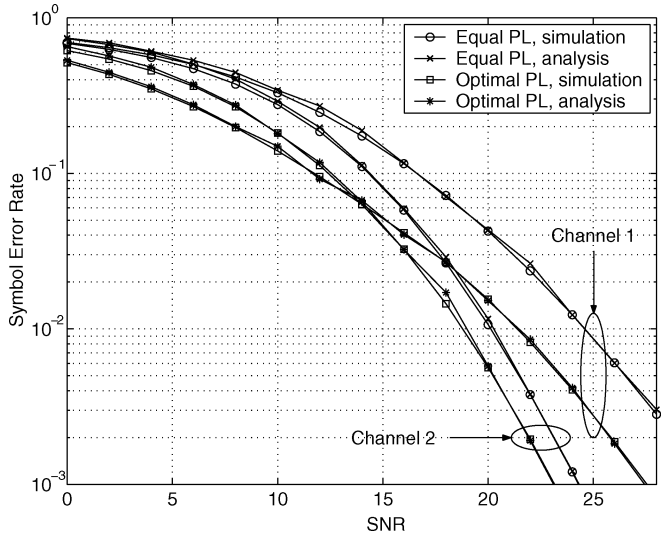

Fig. 3. Simulated and approximate SERs of 8-PSK with known $\mathbf{R}_h$.

Since the channel's spatial correlation is considered as a slowly varying effect similar to shadowing [21], it remains invariant in our simulations. We will consider two channels with different correlations: channel 1 has $d = 0.5\lambda$ and $\Delta = 5°$, while channel 2 has $d = 0.5\lambda$ and $\Delta = 25°$. Channels are normalized so that $\mathrm{Tr}(\mathbf{R}_h) = N_T$. For channel 1, the eigenvalues of $\mathbf{R}_h$ are contained in $\mathbf{\Lambda}_1 = \mathrm{diag}(3.818\,49, 0.180\,79, 0.000\,71, 0.000\,01)$; and for channel 2, we have $\mathbf{\Lambda}_2 = \mathrm{diag}(1.790, 1.741, 0.454, 0.015)$. While channel 1 is highly correlated, channel 2 is less correlated and provides more diversity. QPSK and 8-PSK constellations will be adopted. Since we use a 4×4 orthogonal STBC, the spectral efficiency is 1.5 b/s/Hz for QPSK and 2.25 b/s/Hz for 8-PSK. In all plots, the SNR is defined as SNR $:= \mathcal{E}_s/N_0$.

Fig. 1 displays analytical results of BER performance for QPSK. The correlation matrix of channel 1 has two very small eigenvalues: if we use equal power loading, the power transmitted along two eigenvectors corresponding to these two small eigenvalues is wasted in the SNR region of practical interest. Hence, the optimal power loading outperforms equal power loading by more than 3 dB in the SNR region of interest. Channel 2 is less correlated; thus, the performance gap between optimal and equal power loading is smaller, but still noticeable. Recall that the performance of DSTM with beamforming and equal power loading is the same as that of DSTM without beamforming. Hence, these BER curves demonstrate clearly the advantage of combining DSTM with optimally loaded beamforming. Both exact and approximate BERs are shown in the figure, where it is observed that the approximate BER curve is very close to the exact BER curve even in the low SNR region, which justifies the optimal power loading scheme derived from the approximate BER.

Fig. 2 compares simulated against exact BER for QPSK. In simulations, $\mathbf{R}_h$ is assumed perfectly known at the transmitter. We ran simulations 10^7 times. In each run, we generate two consecutive space–time coding blocks: in the first block $\mathbf{C}_1 = \mathbf{I}$, and in the second block $\mathbf{C}_2 = \mathbf{S}$, where $\mathbf{S}$ is a randomly generated orthogonal STBC matrix. The channel is fixed over a pair of successive blocks in each run but is independently generated from run to run. Hence, our simulations are valid for fast fading channels as long as the channel variation is negligible in two consecutive blocks, even though over a large number of blocks the cumulative variation can be indeed fast, which is a requirement common to all differential schemes. The error probability is obtained by averaging the total number of errors over the number of runs. It is seen that the simulation results match very well with the analytical BER. Fig. 3 depicts SER results obtained from simulations and the approximate SER calculated from the analytical formula in (31) for 8-PSK. Similar to the error probability results in Figs. 1 and 2, we confirm again that optimal power loading offers considerable performance gains for both channels. The approximate SER curve almost coincides with the simulated SER curve, which again corroborates that it is reasonable to derive the power loading scheme from the approximate SER.

While the downlink channel's spatial correlations can be estimated from the uplink channel under certain conditions

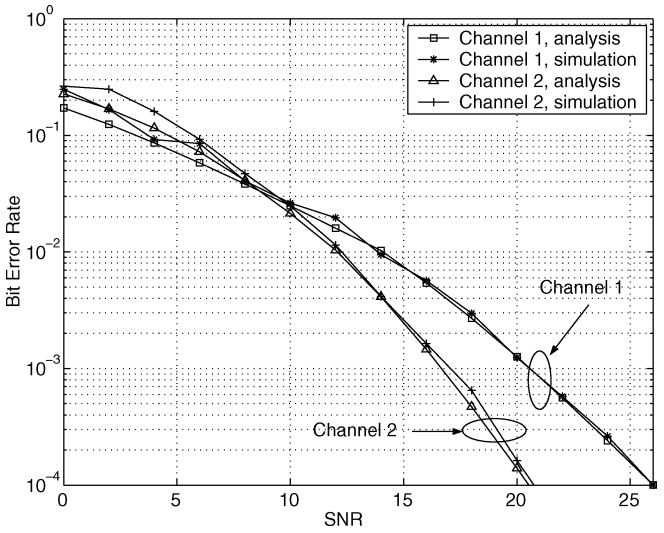

Fig. 4. BER performance of QPSK with estimated $\mathbf{R}_h$.

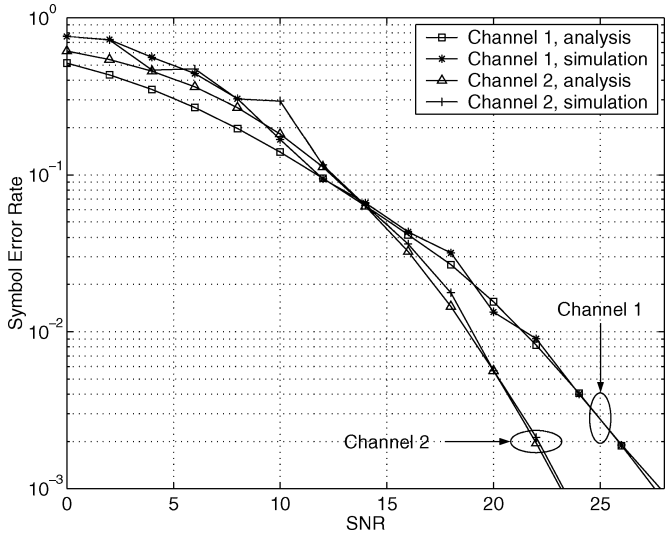

Fig. 5. SER performance of 8-PSK with estimated $\mathbf{R}_h$.

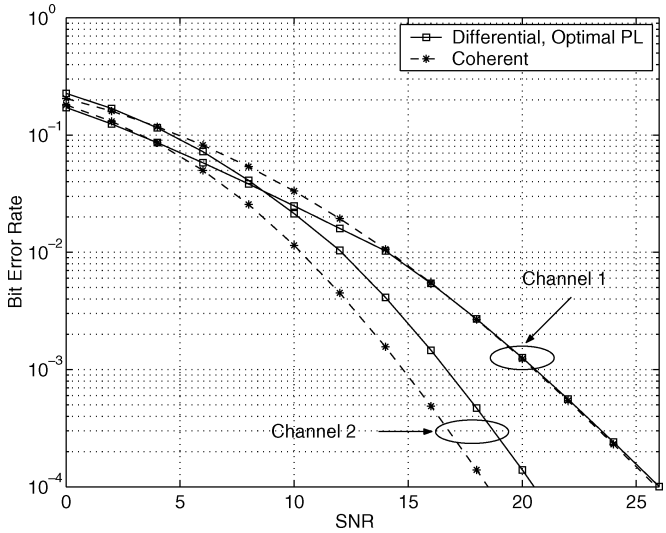

Fig. 6. BER comparison between coherent STBC and DSTM with optimally loaded eigen-beamforming.

[20], we can also estimate channel correlations at the receiver and feed them back to the transmitter. Since spatial correlations change with time very slowly, they only require a low-rate feedback channel, which is often feasible in practice. When channel correlations are not available at the transmitter, no beamforming is employed and the transmitted block is $\mathbf{X}_t = \mathbf{S}_t\mathbf{X}_{t-1}$, $t > 0$ with $\mathbf{X}_0 = \sqrt{P\mathcal{E}_s}\boldsymbol{\Theta}$. At the receiver end, from the detected codeword $\hat{\mathbf{S}}_t$, we can calculate $\hat{\mathbf{X}}_t = \hat{\mathbf{S}}_t\hat{\mathbf{X}}_{t-1}$, and a channel estimate can be obtained as $\hat{\mathbf{h}}_t = \hat{\mathbf{X}}_t^{\mathcal{H}}\mathbf{y}_t$. Note that this channel estimate is not normalized by $\sqrt{P\mathcal{E}_s}$, but this will not affect power loading scheme since the transmitter can normalize eigenvalues of the estimated correlation matrix before calculating power loading coefficients. We form $\hat{\mathbf{h}}_t$ every N_b blocks, where N_b is chosen sufficiently large so that two consecutive samples are uncorrelated. After collecting N_s estimates $\{\hat{\mathbf{h}}_{nN_b}\}_{n=1}^{N_s}$, we average them to estimate the channel correlation as: $\hat{\mathbf{R}}_h = (1/N_s)\sum_{n=1}^{N_s}\hat{\mathbf{h}}_{nN_b}\hat{\mathbf{h}}_{nN_b}^{\mathcal{H}}$. Once the transmitter acquires this channel correlation, signals are transmitted with loaded eigen-beamforming according to (4). This channel correlation estimate is biased by the noise variance. Since the noise variance is easy to measure at the receiver when no data are transmitted, an unbiased channel correlation estimate can be written as $\hat{\mathbf{R}}_h = (1/N_s)\sum_{n=1}^{N_s}\hat{\mathbf{h}}_{nN_b}\hat{\mathbf{h}}_{nN_b}^{\mathcal{H}} - N_0\mathbf{I}_{N_T}$. Note that this channel estimate is only used for estimating the channel correlation, but not for detecting transmitted symbols. While the error in $\hat{\mathbf{S}}_t$ may degrade the channel estimates, it will not cause significant error in $\hat{\mathbf{R}}_h$ because of the averaging involved in estimating $\hat{\mathbf{R}}_h$. This is confirmed by the simulation results in Figs. 4 and 5, which depict the error probability of QPSK and 8-PSK, when $\hat{\mathbf{R}}_h$ estimates are used. In estimating $\mathbf{R}_h$, we use $N_s = 10$, and the noise variance is not subtracted. Comparing Figs. 4 and 5 with Figs. 2 and 3, we see that when the SNR is moderate-high, using $\hat{\mathbf{R}}_h$ instead of $\mathbf{R}_h$ does not degrade error performance.

Since certain practical systems, e.g., 3GPP WCDMA, employ coherent STBC without beamforming [13, p. 97], we compare the BERof DSTM with optimally loaded eigen-beamforming with that of coherent STBC without beamforming in Fig. 6. It is worth to emphasize here that coherent STBC requires CSI at the receiver, and thus, it is not applicable to fast fading channels; whereas DSTM does not require CSI at the receiver, and thus, it performs well in fast fading channels. Both coherent STBC and DSTM use QPSK. When we calculate the BER of coherent STBC, we assume that the receiver knows the channel perfectly. If we take into account the channel estimation error and the transmitted power consumed by training, the BER of coherent STBC will be higher. On the other hand, since DSTM does not require CSI at the receiver, the BER curve in Fig. 6 is its actual BER. We see that for channel 1, the proposed differential modulation scheme actually outperforms the coherent STBC. For channel 2, the error performance gap between these two schemes will be very small after taking into account the channel estimation in coherent STBC. Training in coherent STBC also incurs a data rate loss; hence, DSTM offers higher data rates. Of course, beamforming and optimal power loading can also be employed with coherent STBC to improve error probability performance as in [35]. However, our results here demonstrate that

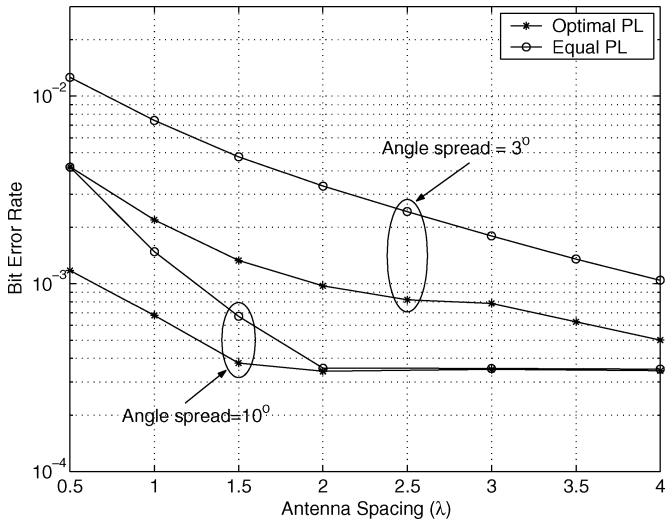

Fig. 7. BER of QPSK versus antenna spacing.

in highly correlated channels, the proposed differential modulation scheme has better or comparable error probability performance, relative to the conventional coherent STBC without beamforming.

Fig. 7 depicts the BER of QPSK versus antenna spacing for typical rural (angle spread $= 3^\circ$) and urban (angle spread $= 10^\circ$) environments [28], where we use the channel correlation computed from (35) and SNR $= 18$ dB. When the angle spread is small, the channels are highly correlated for a wide range of antenna spacings; and thus, the eigen-beamforming and the optimal loading achieve considerable performance gain. On the other hand, when the angle spread is relatively large, the channel correlation is small after the antenna spacing increases beyond 2λ. In this case, the channel correlation matrix is approximately proportional to the identity matrix and beamforming does not improve performance.

V. Conclusion

We have analyzed the error probability performance of differential space–time modulation based on orthogonal space–time block coding. Our performance analysis is useful for both spatially independent and correlated channels. While an exact BER analysis was carried out for BPSK and QPSK, an approximate SER was derived for all M-PSK constellations. Our performance analysis results guided our novel concatenation of differential space–time modulation with eigen-beamforming and derivation of optimal power loading to enhance the error probability performance in correlated channels. Both analysis and simulations illustrated that our scheme can achieve considerable performance gain in correlated channels relative to differential space–time modulation without beamforming.

Appendix I
Proof of Lemma 1

From (1) and (2), it can be shown that

$$\mathbf{S}_t^{\mathcal{H}}\mathbf{\Phi}_p + \mathbf{\Phi}_p^{\mathcal{H}}\mathbf{S}_t = \alpha\mathbf{I}_N \tag{36}$$

and by the definition of $\mathbf{B}_1$, we have

$$\mathbf{B}_1 + \mathbf{B}_1^{\mathcal{H}} = \mathbf{C}_{t-1}^{\mathcal{H}}\left(\mathbf{S}_t^{\mathcal{H}}\mathbf{\Phi}_p + \mathbf{\Phi}_p^{\mathcal{H}}\mathbf{S}_t\right)\mathbf{C}_{t-1}. \tag{37}$$

Combining (36) and (37), we obtain

$$\mathbf{B}_1 + \mathbf{B}_1^{\mathcal{H}} = \alpha\mathbf{I}_N. \tag{38}$$

Since $\mathbf{B}_1^{\mathcal{H}} = \mathbf{B}_1^{-1}$, if λ is an eigenvalue of $\mathbf{B}_1$, then $1/\lambda$ must be an eigenvalue of $\mathbf{B}_1^{\mathcal{H}}$; also, $\mathbf{B}_1$ and $\mathbf{B}_1^{\mathcal{H}}$ have the same corresponding eigenvector $\mathbf{x}$. Multiplying (38) by $\mathbf{x}$ from the right gives $(\lambda + 1/\lambda)\mathbf{x} = \alpha\mathbf{x}$, which implies

$$\lambda + \frac{1}{\lambda} = \alpha. \tag{39}$$

Solving this equation, we obtain the two distinct eigenvalues given in Lemma 1. ■

Appendix II
Proof of Proposition 1

Let columns of matrices $\mathbf{V}_1$ and $\mathbf{V}_2$ consist of eigenvectors of $\mathbf{B}_1$ corresponding to eigenvalues $\lambda_{B,1}$ and $\lambda_{B,2}$, respectively; and define $\mathbf{V} := [\mathbf{V}_1\mathbf{V}_2]$, and $\mathbf{\Lambda}_B := \mathrm{diag}(\lambda_{B,1}\mathbf{I}_{N/2}, \lambda_{B,2}\mathbf{I}_{N/2})$. Then, we have $\mathbf{B}_1\mathbf{V} = \mathbf{V}\mathbf{\Lambda}_B$. Recall that $\lambda_{B,2} = 1/\lambda_{B,1}$ and $\mathbf{B}_1\mathcal{H} = \mathbf{B}_1^{-1}$, we infer that if $\mathbf{v}$ is an eigenvector of $\mathbf{B}_1$ corresponding to eigenvalue $\lambda_{B,1}$, it is also an eigenvector of $\mathbf{B}_1^{\mathcal{H}}$ corresponding to eigenvalue $\lambda_{B,2}$; thus, we have $\mathbf{B}_1^{\mathcal{H}}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}_B^{-1}$. Since $\lambda_{B,2} = \lambda_{B,1}^*$, we can also write this equation as $\mathbf{B}_1^{\mathcal{H}}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}_B^*$. Defining $\bar{\mathbf{V}} := \mathrm{diag}(\mathbf{V}, \mathbf{V})$ and using (23), we obtain

$$\bar{\mathbf{V}}^{\mathcal{H}}\tilde{\mathbf{A}}\bar{\mathbf{V}} = \begin{bmatrix} \mathbf{V}^{\mathcal{H}}\mathbf{D}_h\mathbf{V}\mathbf{\Lambda}_B & \mathbf{V}^{\mathcal{H}}\mathbf{D}_h\mathbf{V}\mathbf{\Lambda}_B^* \\ \mathbf{V}^{\mathcal{H}}\mathbf{D}_h\mathbf{V}\mathbf{\Lambda}_B & \mathbf{V}^{\mathcal{H}}\mathbf{D}_h\mathbf{V}\mathbf{\Lambda}_B^* \end{bmatrix} + N_0\begin{bmatrix} \mathbf{0} & \mathbf{\Lambda}_B^* \\ \mathbf{\Lambda}_B & \mathbf{0} \end{bmatrix}. \tag{40}$$

Define a permutation matrix $\mathbf{P} := \mathrm{diag}(\mathbf{I}_{N/2}, \mathbf{J}_3 \otimes \mathbf{I}_{N/2})$, where $\mathbf{J}_3$ is a 3×3 matrix whose antidiagonal elements are all ones while all other elements are zeros, and $\otimes$ stands for Kronecker product. Note that since $\mathbf{P}$ is a permutation matrix, we have $\mathbf{P}\mathbf{P} = \mathbf{I}_{2N}$. It can be shown that

$$\mathbf{P}\bar{\mathbf{V}}^{\mathcal{H}}\tilde{\mathbf{A}}\bar{\mathbf{V}}\mathbf{P} = \begin{bmatrix} \lambda_{B,1}\mathbf{V}^{\mathcal{H}}\mathbf{D}_h\mathbf{V} & \lambda_{B,2}\mathbf{V}^{\mathcal{H}}\mathbf{D}_h\mathbf{V} \\ \lambda_{B,1}\mathbf{V}^{\mathcal{H}}\mathbf{D}_h\mathbf{V} & \lambda_{B,2}\mathbf{V}^{\mathcal{H}}\mathbf{D}_h\mathbf{V} \end{bmatrix} + N_0\begin{bmatrix} \mathbf{0} & \lambda_{B,2}\mathbf{I}_N \\ \lambda_{B,1}\mathbf{I}_N & \mathbf{0} \end{bmatrix} \tag{41}$$

and thus

$$\begin{aligned} &\bar{\mathbf{V}}\mathbf{P}\bar{\mathbf{V}}^{\mathcal{H}}\tilde{\mathbf{A}}\bar{\mathbf{V}}\mathbf{P}\bar{\mathbf{V}}^{\mathcal{H}} \\ &\quad = \begin{bmatrix} \lambda_{B,1}\mathbf{D}_h & \lambda_{B,2}(\mathbf{D}_h + N_0\mathbf{I}_N) \\ \lambda_{B,1}(\mathbf{D}_h + N_0\mathbf{I}_N) & \lambda_{B,2}\mathbf{D}_h \end{bmatrix} \\ &\quad = \mathbf{G}. \end{aligned} \tag{42}$$

Combining (23) and (42), we obtain

$$\bar{\mathbf{V}}\mathbf{P}\bar{\mathbf{V}}^{\mathcal{H}}\mathbf{C}^{\mathcal{H}}\mathbf{A}\mathbf{C}\bar{\mathbf{V}}\mathbf{P}\bar{\mathbf{V}}^{\mathcal{H}} = \mathbf{G}. \tag{43}$$

Because $\bar{\mathbf{V}}$ and $\mathbf{C}$ are unitary, and $\mathbf{P} = \mathbf{P}^{-1}$, the matrix $\mathbf{A}$ is similar to the matrix $\mathbf{G}$. ■

Appendix III
Proof of Corollary 1

Since $\mathbf{A}$ is similar to $\mathbf{G}$, matrices $\mathbf{A}$ and $\mathbf{G}$ have the same eigenvalues, and it suffices to find the eigenvalues of $\mathbf{G}$. We will

use the following identity regarding the determinant of a block partitioned matrix [19, p. 46]

$$\det\begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{bmatrix} = \det(\mathbf{A}_{11})\det\left(\mathbf{A}_{22}-\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right), \quad \text{if } \det(\mathbf{A}_{11})\neq 0. \quad (44)$$

Suppose that λ is an eigenvalue of $\mathbf{G}$,which is real since $\mathbf{G}$ is Hermitian. Because $\lambda_{B,1}$ is a complex number, we must have $\det(\lambda_{B,1}\mathbf{D}_h-\lambda\mathbf{I}_N)\neq 0$. Using (44), we obtain

$$\begin{aligned}&\det(\mathbf{G}-\lambda\mathbf{I}_{2N})\\ &\quad=\det(\lambda_{B,1}\mathbf{D}_h-\lambda\mathbf{I}_N)\\ &\qquad\times\det\left[(\lambda_{B,2}\mathbf{D}_h-\lambda\mathbf{I}_N)-\lambda_{B,1}(\mathbf{D}_h+N_0\mathbf{I}_N)\right.\\ &\qquad\qquad\left.\times(\lambda_{B,1}\mathbf{D}_h-\lambda\mathbf{I}_N)^{-1}\lambda_{B,2}(\mathbf{D}_h+N_0\mathbf{I}_N)\right]. \quad (45)\end{aligned}$$

Using the identities $\det(\mathbf{A})\det(\mathbf{B})=\det(\mathbf{AB})$, $\lambda_{B,1}\lambda_{B,2}=1$, and noticing that all matrices at the right-hand side of (45) are diagonal, we can write (45) as

$$\det(\mathbf{G}-\lambda\mathbf{I}_{2N})=\det\left[(\lambda_{B,1}\mathbf{D}_h-\lambda\mathbf{I}_N)(\lambda_{B,2}\mathbf{D}_h-\lambda\mathbf{I}_N)-(\mathbf{D}_h+N_0\mathbf{I}_N)^2\right]. \quad (46)$$

Since $\lambda_{B,1}+\lambda_{B,2}=\alpha$, (46) can also be expressed as

$$\det(\mathbf{G}-\lambda\mathbf{I}_{2N})=\prod_{i=1}^{N}\left[\lambda^2-\alpha D_{hi}\lambda-\left(N_0^2+2N_0D_{hi}\right)\right]. \quad (47)$$

Letting $\det(\mathbf{G}-\lambda\mathbf{I}_{2N})=0$, we have

$$\lambda^2-\alpha D_{hi}\lambda-\left(N_0^2+2N_0D_{hi}\right)=0,\quad i=1,\ldots,N. \quad (48)$$

Solving these N equations, we obtain the eigenvalues of $\mathbf{G}$ as given in (26). ■

APPENDIX IV
PROOF OF FACT 1

Since w_1 is uncorrelated with w_4 and w_3 is uncorrelated with w_2, we need to prove that w_1 and w_2 are uncorrelated, and w_3 and w_4 are uncorrelated. Letting $\mathbf{a}:=2\mathbf{\Phi}_p^{\mathcal{H}}\mathbf{S}_t\tilde{\mathbf{C}}_{t-1}\tilde{\mathbf{h}}$ and $\mathbf{b}:=2(j\mathbf{\Psi}_p)^{\mathcal{H}}\mathbf{S}_t\tilde{\mathbf{C}}_{t-1}\tilde{\mathbf{h}}$, we have $w_1=\mathrm{Re}(\mathbf{w}_{t-1}^{\mathcal{H}}\mathbf{a})$ and $w_2=\mathrm{Re}(\mathbf{w}_{t-1}^{\mathcal{H}}\mathbf{b})$. Since $\mathrm{Re}(\mathbf{w}_{t-1})$ and $\mathrm{Im}(\mathbf{w}_{t-1})$ are uncorrelated and they have the same variance, it follows that $E[\mathbf{w}_{t-1}\mathbf{w}_{t-1}^{\mathcal{H}}]=E[\mathbf{w}_{t-1}^{*}\mathbf{w}_{t-1}^{T}]=N_0\mathbf{I}_N$, and $E[\mathbf{w}_{t-1}\mathbf{w}_{t-1}^{T}]=E[\mathbf{w}_{t-1}^{*}\mathbf{w}_{t-1}^{\mathcal{H}}]=\mathbf{0}$. The correlation between w_1 and w_2 is given by

$$\begin{aligned}E[w_1w_2]&=\frac{1}{4}E\left[\left(\mathbf{a}^{\mathcal{H}}\mathbf{w}_{t-1}+\mathbf{a}^{T}\mathbf{w}_{t-1}^{*}\right)\left(\mathbf{w}_{t-1}^{\mathcal{H}}\mathbf{b}+\mathbf{w}_{t-1}^{T}\mathbf{b}^{*}\right)\right]\\ &=\frac{1}{4}N_0(\mathbf{a}^{\mathcal{H}}\mathbf{b}+\mathbf{b}^{\mathcal{H}}\mathbf{a})\\ &=jN_0\tilde{\mathbf{h}}^{\mathcal{H}}\tilde{\mathbf{C}}_{t-1}^{\mathcal{H}}\mathbf{S}_t^{\mathcal{H}}\left(\mathbf{\Psi}_p\mathbf{\Phi}_p^{\mathcal{H}}-\mathbf{\Phi}_p\mathbf{\Psi}_p^{\mathcal{H}}\right)\mathbf{S}_t\tilde{\mathbf{C}}_{t-1}\tilde{\mathbf{h}}. \quad (49)\end{aligned}$$

From (2), it can be shown that $\mathbf{\Psi}_p\mathbf{\Phi}_p^{\mathcal{H}}-\mathbf{\Phi}_p\mathbf{\Psi}_p^{\mathcal{H}}=\mathbf{0}$. Hence, $E[w_1w_2]=0$, which implies that w_1 and $w_2$are uncorrelated. Similarly, we can show that w_3 and w_4 are uncorrelated. It is straightforward to verify that the variance of w_1 is $\sigma_1^2=|\mathbf{a}|^2N_0/2=2|\tilde{\mathbf{h}}|^2N_0$; and similarly, the variance of w_2, w_3, and w_4 equals $2|\tilde{\mathbf{h}}|^2N_0$. ■

REFERENCES

[1] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[2] M. Brehler and M. K. Varanasi, "Asymptotic error probability analysis of quadratic receivers in Rayleigh-fading channels with applications to a unified analysis of coherent and noncoherent space–time receivers," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2383–2399, Sep. 2001.

[3] J. K. Cavers and P. Ho, "Analysis of the error performance of trellis-coded modulation in Rayleigh-fading channels," *IEEE Trans. Commun.*, vol. 40, no. 1, pp. 74–83, Jan. 1992.

[4] T.-A. Chen, M. P. Fitz, W.-Y. Kuo, M. D. Zoltowski, and J. H. Grimm, "A space–time model for frequency nonselective Rayleigh fading channels with applications to space–time modems," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 7, pp. 1175–1190, Jul. 2000.

[5] E. Chiavaccini and G. M. Vietta, "Further results on differential space–time modulations," *IEEE Trans. Commun.*, vol. 51, no. 7, pp. 1093–1101, Jul. 2003.

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

[7] G.Ganesan and P. Stoica, "Differential detection based on space–time block codes," *Wireless Personal Commun.*, vol. 21, pp. 163–180, 2002.

[8] ——, "Differential modulation using space–time block codes," *IEEE Signal Process. Lett.*, vol. 9, no. 2, pp. 57–60, Feb. 2002.

[9] C. Gao and A. M. Haimovich, "BER analysis of MPSK space–time block codes with differential detection," *IEEE Commun. Lett.*, vol. 7, no. 7, pp. 314–316, Jul. 2003.

[10] J. S. Hammerschmidt, C. Brunner, and C. Drewes, "Eigenbeamforming—A novel concept in array signal processing," presented at the VDE/ITG Eur. Wireless Conf., Dresden, Germany, Sep. 2000.

[11] B. M. Hochwald and T. L. Marzetta, "Unitary space–time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.

[12] B. M. Hochwald and W. Sweldens, "Differential unitary space–time modulation," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2041–2052, Dec. 2000.

[13] H. Holma and A. Toskala, *WCDMA for UMTS—Radio Accss for Third Generation Mobile Communications*. New York: Wiley, 2000.

[14] B. L. Hughes, "Differential space–time modulation," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2567–2578, Nov. 2000.

[15] S. A. Jafar, S. Vishwanath, and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," in *Proc. Int. Conf. Communications (ICC)*, Helsinki, Finland, May 2001, pp. 2266–2270.

[16] H. Jafarkhani and V. Tarokh, "Multiple transmit antenna differential detection from generalized orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2626–2631, Sep. 2001.

[17] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.

[18] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space–time block coding," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 611–627, Mar. 2002.

[19] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*. New York: Academic, 1985.

[20] Y.-C. Liang and F. P. S. Chin, "Downlink channel covariance matrix (DCCM) estimation and its applications in wireless DS-CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 2, pp. 222–232, Feb. 2001.

[21] Lucent, Nokia, Siemens, and Ericsson. A standardized set of MIMO radio propagation channels, 3GPP document TSGR1 #23 R1-01-1179. [Online]. Available: http://www.3gorg/ftp/tsg_ran/WG1_RL1/TSGR1_22/Docs/Zips/R1-01-11%79.zip

[22] X. Ma and G. B. Giannakis, "Full-diversity full-rate complex-field space-time coding," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2917–2930, Nov. 2003.

[23] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[24] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct. 1998.
[25] M. Schwartz, W. R. Benett, and S. Stein, *Communication Systems and Techniques*. Piscataway, NJ: IEEE Press, 1996.
[26] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multi-element antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.
[27] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*. New York: Wiley, 2000.
[28] R. A. Soni, R. M. Buehrer, and R. D. Benning, "Intelligent antenna system for CDMA2000," *IEEE Signal Process. Mag.*, vol. 19, no. 4, pp. 54–67, Jul. 2002.
[29] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space–time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1456–1467, Jul. 1999.
[30] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
[31] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 7, pp. 1169–1174, Jul. 2000.
[32] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.
[33] E. Visotsky and U. Madhow, "Space–time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.
[34] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space–time block coding based on channel mean feedback," *IEEE Trans. Signal Proc.*, vol. 50, no. 10, pp. 2599–2613, Oct. 2002.
[35] ——, "Optimal transmitter eigen-beamforming and space-time block coding based on channel correlations," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1673–1690, Jul. 2003.

Xiaodong Cai (S'00–M'01–SM'05) received the B.S. degree from Zhejiang University, China, the M.Eng. degree from the National University of Singapore, Singapore, and the Ph.D. degree from the New Jersey Institute of Technology, Newark, all in electrical engineering.

From February 2001 to June 2001, he was a Member of Technical Staff at Lucent Technologies, Whippany, NJ. From July 2001 to October 2001, he was a Senior System Engineer at Sony technology center, San Diego, CA. From November 2001 to July 2004, he was a Postdoctoral Research Associate in the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. Since August 2004, he has been an Assistant Professor in the Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL. His research interests lie in the areas of communication theory, signal processing, and networking. His current research focuses on cooperative communications and collaborative signal processing for wireless ad hoc and sensor networks.

Georgios B. Giannakis (F'97) received the Diploma degree in electrical engineering from the National Technical University of Athens, Greece, in 1981 and the M.Sc. degree in electrical engineering, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering from the University of Southern California (USC), Los Angeles, in 1983, 1986, and 1986, respectively.

After lecturing for one year at USC, he joined the University of Virginia, Charlottesville, in 1987, where he became a Professor of electrical engineering in 1997. Since 1999, he has been a Professor with the Department of Electrical and Computer Engineering at the University of Minnesota, Minneapolis, where he now holds an ADC Chair in Wireless Telecommunications. His general interests span the areas of communications and signal processing, estimation and detection theory, time-series analysis, and system identification—subjects on which he has published more than 220 journal papers, 380 conference papers, and two edited books. His current research focuses on transmitter and receiver diversity techniques for single- and multiuser fading communication channels, complex-field and space–time coding, multicarrier, ultra-wide-band wireless communication systems, cross-layer designs, and sensor networks.

Dr. Giannakis is the (co-)recipient of six paper awards from the IEEE Signal Processing (SP) Society and the IEEE Communications Society (1992, 1998, 2000, 2001, 2003, and 2004). He also received the SP Society's Technical Achievement Award in 2000. He served as Editor-in-Chief for the IEEE SIGNAL PROCESSING LETTERS, as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS, as Secretary of the SP Conference Board, as Member of the SP Publications Board, as Member and Vice-Chair of the Statistical Signal and Array Processing Technical Committee, as Chair of SP for the Communications Technical Committee, and as a Member of the IEEE Fellows Election Committee. He has also served as a Member of the IEEE SP Society's Board of Governors, the Editorial Board for the PROCEEDINGS OF THE IEEE, and the steering committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.