Spatial Multiplexing in Correlated Fading via the Virtual Channel Representation

Zhihong Hong, *Member, IEEE*, Ke Liu, *Student Member, IEEE*, Robert W. Heath, Jr., *Member, IEEE*, and Akbar M. Sayeed, *Senior Member, IEEE*

Abstract—**Spatial multiplexing techniques send independent data streams on different transmit antennas to maximally exploit the capacity of multiple-input multiple-output (MIMO) fading channels. Most existing multiplexing techniques are based on an idealized MIMO channel model representing a rich scattering environment. Realistic channels corresponding to scattering clusters exhibit correlated fading and can significantly compromise the performance of such techniques. In this paper, we study the design and performance of spatial multiplexing techniques based on a virtual representation of realistic MIMO fading channels. Since the nonvanishing elements of the virtual channel matrix are uncorrelated, they capture the essential degrees of freedom in the channel and provide a simple characterization of channel statistics. In particular, the pairwise-error probability (PEP) analysis for correlated channels is greatly simplified in the virtual representation. Using the PEP analysis, various precoding schemes are introduced to improve performance in virtual channels. Unitary precoding is proposed to provide robustness to unknown channel statistics. Nonunitary precoding techniques are proposed to exploit channel structure when channel statistics are known at the transmitter. Numerical results are presented to illustrate the attractive performance of the precoding techniques.**

Index Terms—**Beamforming, correlated channels, diversity, precoding, space–time coding, spatial multiplexing, virtual channel representation.**

I. INTRODUCTION

INFORMATION theoretic studies [1], [2] indicate that multiple antennas at the transmitter and receiver, so-called multiple-input multiple-output (MIMO) systems, can dramatically increase the capacity and diversity in wireless communication systems. Over the last few years, several distinct bandwidth-efficient communication techniques including space–time coding [3]–[5] and spatial multiplexing [1], [6] have been developed to exploit the potential of MIMO systems. Spatial multiplexing focuses on the rate advantage whereas space–time coding focuses on the diversity advantage of MIMO systems. Thus far, most of MIMO studies heavily use a statistical channel model that is an idealized abstraction of spatial propagation characteristics and assumes independent and identically distributed (i.i.d.) fading between different transmit–receive antenna pairs. This idealized channel model allows tractable and elegant capacity analysis and space–time code design. In practice, however, the channel coefficients between different transmit–receive antenna pairs exhibit correlation due to clustered scattering in realistic environments and the relatively small antenna spacing. In such realistic conditions, the capacity of MIMO channels can be substantially lower depending on the level of correlation [7], [8]. Therefore, a channel model that accurately captures the characteristics of the propagation environment is needed for realistic capacity assessments, as well as for designing space–time modulation and coding techniques that are matched to channel statistics. Parametric physical models (see, e.g., [9]) based on array processing techniques that explicitly model signal copies arriving from different directions provide one such approach. However, the nonlinear dependence of these models on physical channel parameters, such as angles of departure/arrival, makes them rather difficult to be incorporated in transceiver design, explicit capacity calculations, and space–time code design.

In this paper, we propose a framework for spatial multiplexing in correlated MIMO channels using a *virtual* representation for MIMO channels that has been introduced recently [8]. The virtual representation captures the essence of physical modeling without its complexity, provides a tractable *linear* channel characterization and offers a simple and transparent interpretation of the effects of scattering and array characteristics on channel capacity and diversity. The virtual representation corresponds to a fixed coordinate transformation with respect to spatial basis functions defined by fixed virtual angles of arrival/departure. The resulting virtual channel matrix represents the channel by beamforming in fixed directions. In the context of space–time coding, the most attractive feature of the virtual channel matrix is that different scattering clusters correspond to different nonvanishing submatrices with approximately uncorrelated entries. Analogous to the i.i.d. idealized statistical model, the virtual channel matrix captures the essential degrees of freedom in correlated MIMO channels via the powers in its nonvanishing uncorrelated entries. Thus, the virtual channel representation provides a powerful tool for studying the impact of correlated fading on modulation and coding for MIMO channels. It also provides a natural framework for combining beamforming ideas from array processing with space–time coding techniques.

Z. Hong, K. Liu, and A. M. Sayeed are with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA (e-mail: zhihonghong@ieee.org; kliu@cae.wisc.edu; akbar@dune.ece.wisc.edu).

R. W. Heath, Jr., is with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1804 USA (e-mail: rheath@ece.utexas.edu).

Our approach to spatial multiplexing in correlated MIMO channels is motivated by the analysis of the pairwise-error probability (PEP) in the virtual channel representation. As we will see, the key problem in correlated channels is that some particular codeword error vectors may lie in the channel null space thereby increasing the PEP. The virtual channel matrix, due to its uncorrelated nonvanishing entries, clearly exposes the interaction between the signal space and the channel that causes such degradation in spatial multiplexing performance. Based on our analysis, precoding the transmitted codewords emerges as a simple and effective way for dealing with correlated fading. When the channel statistics are unknown at the transmitter, precoding via *unitary* matrices is proposed for rotating the transmitted codewords to avoid collisions between the codeword error vectors and the channel null space. When channel statistics are known at the transmitter, the structure of nonvanishing entries in the virtual channel correlation matrix is exploited to develop precoding techniques via *nonunitary* matrices that *rotate* and *scale* the codeword vectors to avoid collisions with the channel null space, as well as to match transmitted signal power to the relative channel power in different spatial dimensions.

Precoding techniques have been investigated by several researchers in related contexts. Precoding to rotate the signal constellation and improve robustness of spatial multiplexing is considered in [10] for a polarized channel and in [11] in the presence of spatial correlation for the special case of two transmit and two receive antennas. Both above schemes consider diagonal precoding matrices and both assume knowledge of the (nondiagonal) channel correlation matrix at the transmitter. Linear precoding is considered in [12] for space–time coded system with known fading correlations and in [13] for transmit diversity with random-fading channels. The key advantage of our precoding schemes is due to the uncorrelated nature of the virtual channel matrix: the virtual channel correlation matrix is diagonal regardless of the correlation structure of the original channel matrix. As we will see, this greatly simplifies PEP analysis and also offers direct insights for matching signaling schemes to channel characteristics. In a sense, the proposed precoding techniques effectively combine beamforming ideas with space–time coding via the virtual channel representation.

The paper is organized as follows. In Section II, we provide a brief review of the virtual channel representation. In Section III, we provide a general framework for PEP analysis in correlated MIMO channels based on the virtual channel representation. Section IV focuses on the performance of spatial multiplexing techniques in correlated channels and discusses the mechanisms underlying the degradation of performance in such channels. Motivated by the analysis in Sections III and IV, precoding techniques are proposed in Section V that effectively account for correlated fading depending on whether channel statistics are known at the transmitter or not. Numerical results illustrating the performance of precoding techniques are presented in Section VI. Section VII presents some concluding remarks.

The following notation is used throughout the paper. We use * for conjugate, ^{T} for transpose, ^{H} for conjugate transpose, $\text{tr}(\cdot)$ for the trace operator, $\text{vec}(\cdot)$ for stacking columns of a ma-
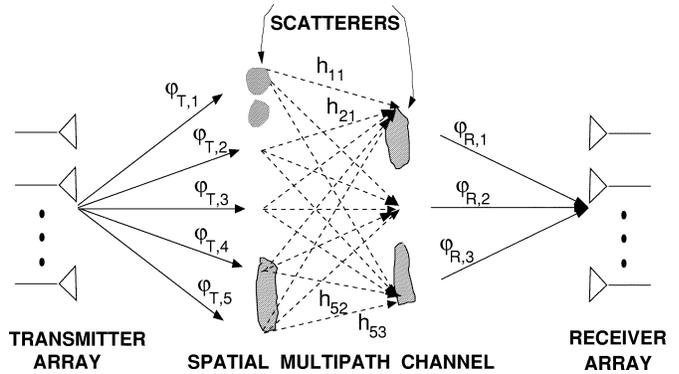
Fig. 1. A schematic illustrating the virtual representation of a physical scattering environment. The virtual representation corresponds to beamforming in fixed directions determined by the resolution of the arrays.

trix, $\otimes$ for Kronecker product [14], $\|\mathbf{A}\|_F^2 = \text{tr}(\mathbf{A}\mathbf{A}^H)$ for the Frobenius norm, and E_s to denote expectation with respect to random variable s when it is not clear from the context. We use boldface lowercase letters to denote vectors and boldface uppercase to denote matrices. $[\mathbf{s}]_k$ is the kth element of the vector $\mathbf{s}$ while $[\mathbf{S}]_{k,l}$ is the element in the kth row and lth column of the matrix $\mathbf{S}$.

II. VIRTUAL CHANNEL REPRESENTATION

We consider a narrowband MIMO system with M_t transmit antennas and M_r receive antennas. In the absence of noise, we have $\mathbf{y} = \mathbf{H}\mathbf{s}$, where $\mathbf{s}$ is the M_t-dimensional transmitted vector, $\mathbf{y}$ is the M_r-dimensional received vector, and $\mathbf{H}$ is the $M_r \times M_t$ channel matrix coupling the transmit and receive antennas. The idealized statistical model corresponding to a rich scattering environment assumes that the elements of $\mathbf{H}$ are i.i.d. complex Gaussian random variables. However, the elements of $\mathbf{H}$ are correlated in realistic environments and the statistics of $\mathbf{H}$ are dictated by the scattering and array characteristics, such as angular spreads of scattering clusters and antenna spacing [7], [8]. In this paper, we use the virtual channel representation introduced in [8] to capture the statistical structure of correlated fading channels imposed by clustered scattering environments.

A schematic illustrating the virtual channel representation is shown in Fig. 1. Consider one-dimensional uniform linear arrays (ULAs) of antennas at both the transmitter and receiver for simplicity and assume that far-field assumptions apply. The channel matrix can then be described via the array steering and response vectors given by

$$\mathbf{a}_T(\theta_T) = \frac{1}{\sqrt{M_t}}\left[1, e^{-j2\pi\theta_T}, \ldots, e^{-j2\pi(M_t-1)\theta_T}\right]^T$$

$$\mathbf{a}_R(\theta_R) = \frac{1}{\sqrt{M_r}}\left[1, e^{-j2\pi\theta_R}, \ldots, e^{-j2\pi(M_r-1)\theta_R}\right]^T \quad (1)$$

where θ represents the angle variable and is related to the physical angle φ (measured with respect to the horizontal axis as illustrated in Fig. 1) as $\theta = d\sin(\varphi)/\lambda$, where λ is the wavelength of propagation and d is the antenna spacing. The vector $\mathbf{a}_R(\theta_R)$ represents the signal response at the receiver array due to a point source in the direction φ_R. Similarly, $\mathbf{a}_T(\theta_T)$ represents the array weights needed to transmit a beam focused in the

direction of φ_T. Note that the steering and response vectors in (1) are periodic in θ with period 1.

Parameterized physical models represent $\mathbf{H}$ via signal propagation over multiple paths (see, e.g., [8] and [9])

$$\mathbf{H} = \sum_{l=1}^{L} \beta_l \mathbf{a}_R(\theta_{R,l}) \mathbf{a}_T^H(\theta_{T,l}) \qquad (2)$$

where β_l is the fading gain, $\theta_{R,l}$ represents the angle of arrival (AoA), and $\theta_{T,l}$ represents the angle of departure (AoD) associated with the lth path. The virtual representation, on the other hand, exploits the finite dimensionality of the signal space[1] to use spatial beams in *fixed, virtual* directions (as illustrated in Fig. 1) to capture the effect of the scattering environment [8]. The virtual channel representation can be expressed as

$$\mathbf{H} = \sum_{m=1}^{M_r} \sum_{n=1}^{M_t} [\mathbf{H}_V]_{m,n} \mathbf{a}_R(\theta_{R,m}) \mathbf{a}_T^H(\theta_{T,n}) = \mathbf{A}_R \mathbf{H}_V \mathbf{A}_T^H$$

$$(3)$$

where $\mathbf{A}_R = [\mathbf{a}_R(\theta_{R,1}), \ldots, \mathbf{a}_R(\theta_{R,M_r})]$ ($M_r \times M_r$) and $\mathbf{A}_T = [\mathbf{a}_T(\theta_{T,1}), \ldots, \mathbf{a}_T(\theta_{T,M_t})]$ ($M_t \times M_t$) are full-rank matrices defined by the fixed virtual angles $\{\theta_{T,n}\}$ and $\{\theta_{R,m}\}$. Uniform sampling of θ in the principal period ($\theta \in [-0.5, 0.5)$) is a natural choice for virtual spatial angles

$$\theta_{R,m} = \frac{m-1}{M_r} - 0.5, \ \theta_{T,n} = \frac{n-1}{M_t} - 0.5 \qquad (4)$$

which yields unitary matrices $\mathbf{A}_R$ and $\mathbf{A}_T$. $\mathbf{A}_R$ is an $M_r \times M_r$ discrete Fourier transform (DFT) matrix and $\mathbf{A}_T$ is an $M_t \times M_t$ DFT matrix. Therefore, $\mathbf{H}_V$ is unitarily equivalent to $\mathbf{H}$ and captures all channel information.

Realistic propagation environments can be modeled via a superposition of scattering clusters with limited angular spreads (see, e.g., [8] and [15]). The virtual channel matrix $\mathbf{H}_V$ provides an intuitively appealing "imaging" representation for such environments: different clusters correspond to different nonvanishing submatrices of $\mathbf{H}_V$. Furthermore, it is shown in [8] that the nonvanishing elements of $\mathbf{H}_V$ are approximately uncorrelated under the usual assumption of uncorrelated physical scattering. The virtual channel matrix clearly reveals the capacity and diversity afforded by a given scattering environment. The capacity multiplier provided by a cluster is determined by the size/rank of the corresponding submatrix and depends on the number of virtual angles that lie within the angular spread of the cluster. The number of nonvanishing entries in each submatrix determines the diversity afforded by the cluster and depends on the nature of scattering within the cluster.

A valuable representation of (3) is obtained by stacking the columns of $\mathbf{H}$ as

$$\mathbf{h} = \text{vec}(\mathbf{H}) = [\mathbf{A}_T^* \otimes \mathbf{A}_R] \text{vec}(\mathbf{H}_V)$$

$$= \sum_{m=1}^{M_r} \sum_{n=1}^{M_t} [\mathbf{H}_V]_{m,n} \mathbf{a}_T^*(\theta_{T,n}) \otimes \mathbf{a}_R(\theta_{R,m}) \qquad (5)$$

using the identity $\text{vec}(\mathbf{ABC}) = [\mathbf{C}^T \otimes \mathbf{A}]\text{vec}(\mathbf{B})$. Let $\mathbf{R} = E(\mathbf{h}\mathbf{h}^H)$ denote the correlation matrix of $\mathbf{h}$ and $\mathbf{R}_V$ the correlation matrix of $\mathbf{h}_V = \text{vec}(\mathbf{H}_V)$. $\mathbf{R}$ and $\mathbf{R}_V$ are related by

$$\mathbf{R} = [\mathbf{A}_T^* \otimes \mathbf{A}_R] \mathbf{R}_V [\mathbf{A}_T^T \otimes \mathbf{A}_R^H]$$

$$\approx \sum_{m=1}^{M_r} \sum_{n=1}^{M_t} \sigma_{m,n}^2 [\mathbf{a}_T^*(\theta_{T,n}) \otimes \mathbf{a}_R(\theta_{R,m})]$$

$$\cdot [\mathbf{a}_T^T(\theta_{T,n}) \otimes \mathbf{a}_R^H(\theta_{R,m})] \qquad (6)$$

where $\sigma_{m,n}^2 = E|[\mathbf{H}_V]_{m,n}|^2$ are the diagonal entries of $\mathbf{R}_V$. Note that due to the approximately uncorrelated nature of the elements of $\mathbf{H}_V$, $\mathbf{R}_V$ is approximately diagonal. We assume $\mathbf{R}_V$ to be exactly diagonal in this paper.[2] Furthermore, $\mathbf{R}_V$ may have some zero elements on the diagonal corresponding to the vanishing elements in $\mathbf{H}_V$ due to clustered scattering. We note that $\mathbf{R}$ and $\mathbf{R}_V$ are unitarily equivalent since the Kronecker product of two unitary matrices is also unitary. Thus, (6) is an eigendecomposition of $\mathbf{R}$ and (5) is the corresponding Karhunen–Loeve-like representation for each channel realization. Therefore, the nonvanishing diagonal elements of $\mathbf{R}_V$ that capture the power in the nonvanishing elements of $\mathbf{H}_V$ also determine the eigenvalues of $\mathbf{R}$. As we will see later, this is a very powerful property of the virtual channel matrix from the viewpoint of PEP calculations.

III. PEP IN SPATIAL MULTIPLEXING SYSTEMS

In this section, we review spatial multiplexing and derive the PEP for spatial multiplexing in correlated channels via the virtual representation.

A. Review of Spatial Multiplexing

Spatial multiplexing is a modulation technique for MIMO communication systems in which independent streams of data are multiplexed in space and subsequently demultiplexed at the receiver [1], [6]. During every discrete-time symbol period, the encoder multiplexes M_t complex symbols $\{[\mathbf{s}]_n\}_{n=0}^{M_t-1}$ from a unit energy constellation to form a complex vector codeword $\mathbf{s}$. The components of this vector codeword are modulated, up-converted and launched into the channel.

Neglecting symbol timing errors and frequency offsets, the $M_r \times 1$ received signal vector after matched filtering and sampling can be written as

$$\mathbf{y} = \sqrt{\frac{\rho}{M_t}} \mathbf{H} \mathbf{s} + \mathbf{v} \qquad (7)$$

where $\mathbf{v}$ is the vector realization of i.i.d. complex circularly symmetric additive white Gaussian noise (AWGN) processes with distribution $\mathcal{N}(\mathbf{0}, (N_o)\mathbf{I}_{M_r})$, where $\mathbf{I}_{M_r}$ denotes the identity matrix of dimension M_r and ρ is the total signal power. The channel $\mathbf{H}$ is assumed to be perfectly known at the receiver (via training symbols, e.g.,) but *unknown to the transmitter*. As will become clearer, limited information about the *statistics* of $\mathbf{H}$ at

---

[1]Due to finite number of antennas and, thus, finite array apertures.

[2]The approximation gets better with larger number of antennas and/or large array apertures [8].

the transmitter can be fruitfully exploited. We assume the use of the maximum-likelihood (ML) decoder at the receiver.

B. Pairwise-Error Probability (PEP)

For random channels the metric of interest is typically the average probability of error. Exact calculation of the symbol or bit-error probability for spatial multiplexing systems, however, is difficult [11], [16]. One solution commonly employed [17] is to upper bound the desired error probability using the union bound and the PEP.

Let $P(\mathbf{s}^{(m)} \to \mathbf{s}^{(k)}|\mathbf{H})$ denote the probability that $\mathbf{s}^{(m)}$ is decoded at the receiver erroneously as $\mathbf{s}^{(k)}$ for a given $\mathbf{H}$. Let the error vector $\mathbf{e}^{(m,k)} = \mathbf{s}^{(m)} - \mathbf{s}^{(k)}$ and define the error correlation matrix of $\mathbf{e}^{(m,k)}$ as $\mathbf{R}^{(m,k)} = \mathbf{e}^{(m,k)}\mathbf{e}^{(m,k)H}$. Using the Chernoff bound $Q(x) \le e^{-x^2/2}$ to upper bound the PEP and the fact that $\mathbf{h}$ has complex normal distribution with zero mean and covariance $\mathbf{R}$, it can be shown that [18]

$$
\begin{aligned}
P\left(\mathbf{s}^{(m)} \to \mathbf{s}^{(k)}\right) &= E_{\mathbf{H}}\left(P\left(\mathbf{s}^{(m)} \to \mathbf{s}^{(k)}|\mathbf{H}\right)\right) \\
&\le \left| \mathbf{I}_{M_r M_t} + \frac{\rho}{4 M_t N_o} \mathbf{R}\left(\mathbf{R}^{(m,k)*} \otimes \mathbf{I}_{M_r}\right) \right|^{-1}.
\end{aligned}
\tag{8}
$$

This is strictly true only for $\mathbf{R}$ nonsingular. As we pointed out, $\mathbf{R}$ is often singular or nearly singular due to clustered scattering. To avoid difficulties with the singular distribution, we can proceed with the above derivation assuming that $\hat{\mathbf{R}} = \mathbf{R} + \epsilon\mathbf{I}$ and then can let ϵ go to zero to arrive at the result in (8). A more formal derivation of (8) is presented in [16] and [19].

Substituting (6) into (8) using the virtual representation, following the identity $|\mathbf{I} + \mathbf{AB}| = |\mathbf{I} + \mathbf{BA}|$ and using the distributive property of the Kronecker product, we have

$$
\begin{aligned}
&P\left(\mathbf{s}^{(m)} \to \mathbf{s}^{(k)}\right) \\
&\le \left| \mathbf{I}_{M_r M_t} + \frac{\rho}{4 M_t N_o} \mathbf{R}_V \left[\mathbf{A}_T^T \mathbf{R}^{(m,k)*} \mathbf{A}_T^* \otimes \mathbf{I}_{M_r} \right] \right|^{-1} \\
&= \prod_{r=0}^{R-1} \left(1 + \frac{\rho}{4 M_t N_o} \lambda_r \left(\mathbf{C}^{(m,k)} \right) \right)^{-1}
\end{aligned}
\tag{9}
$$

where $\mathbf{C}^{(m,k)} = \mathbf{R}_V[\mathbf{A}_T^T \mathbf{R}^{(m,k)*} \mathbf{A}_T^* \otimes \mathbf{I}_{M_r}]$, $R = \text{rank}(\mathbf{C}^{(m,k)})$ and λ_r denotes the rth eigenvalue of $\mathbf{C}^{(m,k)}$.

C. Signaling in the Virtual Channel

Define the DFT of the codeword error correlation matrix as

$$
\begin{aligned}
\mathbf{Q}^{(m,k)} &= \mathbf{A}_T^T \mathbf{R}^{(m,k)*} \mathbf{A}_T^* = \mathbf{A}_T^T \mathbf{e}^{(m,k)*} \mathbf{e}^{(m,k)T} \mathbf{A}_T^* \\
&= \mathbf{q}^{(m,k)} \mathbf{q}^{(m,k)H}
\end{aligned}
\tag{10}
$$

which is rank-1 and $\mathbf{q}^{(m,k)} = \mathbf{A}_T^T \mathbf{e}^{*(m,k)}$ is the DFT of the error vector $\mathbf{e}^{(m,k)}$. From (9), we see that the PEP is governed by the interaction between $\mathbf{R}_V$ and $\mathbf{Q}^{(m,k)} \otimes \mathbf{I}_{M_r}$. The virtual channel representation, thus, suggests the DFT as a natural precoding matrix so that the codewords directly interact with the scattering characteristics captured by $\mathbf{H}_V$. That is, we consider transmitted signals of the form $\mathbf{s} = \mathbf{A}_T \mathbf{s}_V$ so that the channel (7) becomes

$$
\mathbf{y}_V = \sqrt{\frac{\rho}{M_t}} \mathbf{H}_V \mathbf{s}_V + \mathbf{v}_V
\tag{11}
$$
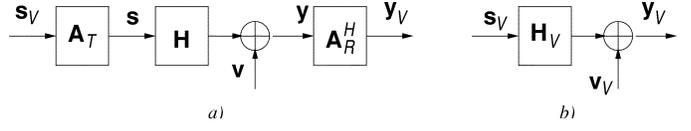
Fig. 2. Signaling in the virtual channel. (a) System depiction with $\mathbf{A}_T$ as a precoder and $\mathbf{A}_R$ as a postcoder. $\mathbf{s}_V$ is the multiplexed symbol vector. (b) Equivalent representation of (a). $\mathbf{s}_V$ directly interacts with $\mathbf{H}_V$.

where $\mathbf{y}_V = \mathbf{A}_R^H \mathbf{y}$ and similarly for $\mathbf{v}_V$. The above equation says that at the transmitter, we first apply a DFT ($\mathbf{A}_T$) to the transmitted codewords $\mathbf{s}_V$ before launching them onto the channel and at the receiver, we first apply a DFT ($\mathbf{A}_R^H$) to the received vector $\mathbf{y}$ before decoding $\mathbf{s}_V$. Thus, (11) corresponds to signaling and reception directly in the virtual (Fourier) domain as illustrated in Fig. 2. Our subsequent development is in the context of (11). In this context, the transmit and receive antennas correspond to the virtual transmit and receive elements (corresponding to beams in virtual directions) and the matrix $\mathbf{Q}^{(m,k)*}$ becomes the codeword error correlation matrix ($\mathbf{R}^{(m,k)}$) corresponding to the actual transmitted vectors [not their DFT as in (10)].

D. Rank and Eigenvalue Characterization

According to the rank and determinant criteria [3], [20], the error rate performance is a function of the rank and the product of the eigenvalues of $\mathbf{C}^{(m,k)}$. The diversity advantage is determined by the rank which we can bound as

$$
\begin{aligned}
R &= \text{rank}\left(\mathbf{C}^{(m,k)}\right) \le \min\left(\text{rank}\left(\mathbf{R}_V\right), \text{rank}\left(\mathbf{Q}^{(m,k)} \otimes \mathbf{I}_{M_r}\right)\right) \\
&= \min\left(\text{rank}\left(\mathbf{R}_V\right), M_r\right)
\end{aligned}
\tag{12}
$$

since $\mathbf{Q}^{(m,k)}$ is a rank-1 matrix. In realistic channels corresponding to clustered scattering, $\mathbf{R}$ is not diagonal but $\mathbf{R}_V$ is still diagonal due to the properties of the virtual representation [8]. Note that the maximum rank of $\mathbf{R}_V$ is $M_t M_r$ and, thus, the absence of coding across time limits spatial multiplexing to a diversity advantage (defined as the minimum rank of all possible $\mathbf{C}^{(m,k)}$) that is at most equal to M_r. Space–time coding can increase the rank of $\mathbf{Q}^{(m,k)}$ to M_t thereby restoring the diversity loss. In addition to the rank of $\mathbf{Q}^{(m,k)}$, the nature of scattering can also reduce the diversity advantage. This is because some diagonal elements of $\mathbf{R}_V$ may be zero or near-zero depending on the scattering geometry and antenna spacing [8]. However, due to the inherent difference in the ranks of $\mathbf{R}_V$ and $\mathbf{Q}^{(m,k)} \otimes \mathbf{I}_{M_r}$, $\mathbf{R}_V$ could have many vanishing (or small) diagonal entries and still yield the maximum diversity advantage (M_r). It all depends on the interaction between $\mathbf{R}_V$ (channel statistics) and $\mathbf{Q}^{(m,k)} \otimes \mathbf{I}_{M_r}$ (code error properties) in (9).

Let $\mathbf{R}_V = \text{diag}\{\mathbf{R}_V(0), \mathbf{R}_V(1), \ldots, \mathbf{R}_V(M_t - 1)\}$ be the diagonal decomposition of $\mathbf{R}_V$ in terms of $M_r \times M_r$ matrices, $\mathbf{R}_V(n) = E[\mathbf{h}_V(n)\mathbf{h}_V^H(n)]$, where $\mathbf{h}_V(n)$ is the nth column of $\mathbf{H}_V$.

Theorem 1: Explicit characterization of eigenvalues of $\mathbf{C}^{(m,k)}$. The M_r eigenvalues of $\mathbf{C}^{(m,k)}$ are given by

$$
\lambda_r\left(\mathbf{C}^{(m,k)}\right) = \sum_{n=0}^{M_t-1} \left| \left[\mathbf{q}^{(m,k)}\right]_n \right|^2 [\mathbf{R}_V(n)]_{r,r},
$$
$$
r = 0, \ldots, M_r - 1. \tag{13}
$$

Proof: We suppress the superscript (m, k) for simplicity of notation. First note that for $\mathbf{Q} = \mathbf{q}\mathbf{q}^H$, $\mathbf{R}_V[\mathbf{Q} \otimes \mathbf{I}_{M_r}] = \mathbf{R}_V[\mathbf{q} \otimes \mathbf{I}_{M_r}][\mathbf{q}^H \otimes \mathbf{I}_{M_r}]$. Furthermore, the nonvanishing eigenvalues of a product of matrices are not changed by changing the order of matrices. Now

$$
\begin{aligned}
[\mathbf{q}^H \otimes \mathbf{I}_{M_r}]&\mathbf{R}_V[\mathbf{q} \otimes \mathbf{I}_{M_r}] \\
&= \left[[\mathbf{q}]_0^* \mathbf{I}_{M_r}, [\mathbf{q}]_1^* \mathbf{I}_{M_r}, \ldots, [\mathbf{q}]_{M_t-1}^* \mathbf{I}_{M_r}\right] \\
&\quad \left[[\mathbf{q}]_0 \mathbf{R}_V(0), [\mathbf{q}]_1 \mathbf{R}_V(1), \ldots, \right. \\
&\qquad \left. [\mathbf{q}]_{M_t-1} \mathbf{R}_V(M_t - 1)\right]^T \\
&= \sum_{n=0}^{M_t-1} |[\mathbf{q}]_n|^2 \mathbf{R}_V(n)
\end{aligned}
\tag{14}
$$

where note that $M_r \times M_r$ matrices $\mathbf{R}_V(n)$ in the last equality are diagonal matrices. Thus, the eigenvalues of $\mathbf{C}^{(m,k)}$ are given by the diagonal entries of the matrix in (14) which are given in (13). $\qquad\square$

The above theorem has a very insightful interpretation: the rth eigenvalue is equal to a weighted sum of the powers of all virtual transmit antennas coupled to the rth virtual receive antenna (via the rth row of $\mathbf{H}_V$). The weighting is given by the magnitude squares of the error vector components, $\{|[\mathbf{q}]_n|^2\}$, corresponding to different virtual transmit antennas. Note that the rth eigenvalue may be zero if either $[\mathbf{R}_V(n)]_{r,r} = 0$ for all n (rth row of $\mathbf{H}_V$ is zero) or if $\mathbf{q}_n = 0$ for those values of n for which $[\mathbf{R}_V(n)]_{r,r} \neq 0$. Rearranging the order of nonzero eigenvalues, the PEP in (9) can be bounded at high signal-to-noise ratios (SNRs) as

$$
\mathrm{PEP}(\mathbf{q}, \mathbf{R}_V) \leq \left(\frac{\rho}{4M_t N_o}\right)^{-R} \left(\left[\prod_{r=0}^{R-1} \lambda_r(\mathbf{q}, \mathbf{R}_V)\right]^{1/R}\right)^{-R}
\tag{15}
$$

and, thus, the diversity gain is R and the coding gain is $[\prod_{r=0}^{R-1} \lambda_r(\mathbf{q}, \mathbf{R}_V)]^{1/R}$. Note that both the diversity and coding gains depend on both the error codeword properties, as well as channel characteristics as evident from Theorem 1. For given $\mathbf{R}_V$, to maximize diversity gain we would like to have as many elements of error vectors $\mathbf{q} \in \mathcal{E}$ be nonvanishing for which the corresponding $[\mathbf{R}_V(n)]_{r,r}$ is nonzero, where $\mathcal{E}$ denotes the set of all possible codeword error vectors. To maximize coding gain, in addition, we would like each $|[\mathbf{q}]_n|$ to be as large as possible over the entire set $\mathcal{E}$.

IV. SPATIAL MULTIPLEXING IN THE VIRTUAL CHANNEL

In this section, we use the PEP analysis in the previous section to get further insight into the performance of spatial multiplexing in correlated channels.

Achieving the upper bound in (12) assumes that column space of $\mathbf{Q}^{(m,k)} \otimes \mathbf{I}_{M_t}$ is not contained in the null-space of $\mathbf{R}_V$. However, due to the unconstrained nature (relative to general space–time coding) of codeword error vectors in spatial multiplexing and due to the vanishing diagonal entries of $\mathbf{R}_V$ in clustered scattering, full intersection between the column space of $\mathbf{Q}^{(m,k)} \otimes \mathbf{I}_{M_t}$ and the range space of $\mathbf{R}_V$ cannot be guaranteed. This results in loss of diversity gain which we quantify in this section.

Exploiting the diagonal nature of $\mathbf{R}_V$ and the structure of the codeword error vectors in spatial multiplexing, we can obtain an exact expression for the diversity advantage that holds in a variety of situations in which the same constellation is used at all virtual transmit antennas.

Theorem 2: Exact diversity advantage. For spatial multiplexing using the same constellation at each virtual transmit antenna, the minimum rank of $\mathbf{C}^{(m,k)}$ over $\mathcal{E}$ satisfies

$$
\begin{aligned}
\min_{\mathbf{q}^{(m,k)} \in \mathcal{E}} &\mathrm{rank}\left(\mathbf{C}^{(m,k)}\right) \\
&= \min_{\mathbf{q}^{(m,k)} \in \mathcal{E}} \mathrm{rank}\left(\mathbf{R}_V\left[\mathbf{Q}^{(m,k)} \otimes \mathbf{I}_{M_r}\right]\right) \\
&= \min_{n=0,1,\ldots,M_t-1} \sum_{l=0}^{M_r-1} I\left([\mathbf{R}_V]_{nM_t+l,nM_t+l} > \epsilon\right)
\end{aligned}
\tag{16}
$$

where $I(\cdot)$ is the indicator function and ϵ is a threshold value for determining the essentially nonvanishing entries of $\mathbf{R}_V$.

Proof: First note that $\mathbf{R}_V[\mathbf{Q} \otimes \mathbf{I}_{M_r}] = \mathbf{R}_V[\mathbf{q}\mathbf{q}^H \otimes \mathbf{I}_{M_r}] = \mathbf{R}_V[\mathbf{q} \otimes \mathbf{I}_{M_r}][\mathbf{q}^H \otimes \mathbf{I}_{M_r}]$ by the distributive property of Kronecker products. Thus, the rank of $\mathbf{C}$ is determined by the rank of $\mathbf{R}_V[\mathbf{q} \otimes \mathbf{I}_{M_r}]$. Recall that $\mathbf{R}_V = \mathrm{diag}\{\mathbf{R}_V(0), \mathbf{R}_V(1), \ldots, \mathbf{R}_V(M_t - 1)\}$. Thus, we have

$$
\begin{aligned}
\mathbf{R}_V[\mathbf{q} \otimes \mathbf{I}_{M_r}] = \\
[[\mathbf{q}]_0 \mathbf{R}_V(0), [\mathbf{q}]_1 \mathbf{R}_V(1), \ldots, [\mathbf{q}]_{M_t-1} \mathbf{R}_V(M_t - 1)]^T.
\end{aligned}
\tag{17}
$$

Clearly, the worst set of error vectors are of the form in which $\mathbf{q}$ is nonzero in only one element (unit error vector). In the case that the same constellation is used on all virtual antennas, all such error vectors are possible. Suppose that $\mathbf{q}$ is only nonzero in the nth element. From (17), the rank of $\mathbf{R}_V[\mathbf{q} \otimes \mathbf{I}_{M_r}]$ for such an error vector is determined by the nonvanishing diagonal elements in $\mathbf{R}_V(n)$ which is exactly the expression corresponding to the index n in (16). Thus, the minimum rank corresponds to $\mathbf{R}_V(n)$ with the smallest number of nonvanishing elements. And this is achievable since all such unit error vectors are possible. This yields the equality in (16). $\qquad\square$

Theorem (16) gives a convenient and explicit expression for the worst-case diversity advantage as a function of the scattering environment. The actual diversity gain can be larger depending on the interaction between the channel and the error codeword vectors. The proof of the theorem yields useful insights in this regard. In particular, it shows that if the nth virtual transmit antenna is not coupled at all to the receivers—$\mathbf{R}_V(n) = \mathbf{0}$ (the nth column of $\mathbf{H}_V$ is identically zero)—no symbols should be transmitted on that virtual transmit antenna; that is $[\mathbf{s}_V]_n$ should be zero. This is because those transmissions are not observable at the receiver at all. This is important in practice since some of the virtual transmit elements might be weakly coupled to the receivers and should not be used for transmission. In such cases, data should only be multiplexed on the virtual transmit antennas corresponding to relatively strong columns of $\mathbf{R}_V$. As our numerical results demonstrate, accounting for such weakly coupled virtual transmit antennas can result in significant improvement in performance. A related important insight is to avoid unit error codeword vectors. Unfortunately, since the symbols at different virtual transmit antennas are independent in spatial mul-

tiplexing, all such error vectors are possible. As we discuss in the next section, applying a precoding transform to the transmitted vectors is an effective way to avoid such error vectors. More importantly, since we cannot introduce dependencies between transmitted symbols in spatial multiplexing, such a precoding approach is the only way of avoiding unit error codeword vectors.

V. UNITARY AND NONUNITARY PRECODING

In this section, we leverage the insights from PEP analysis to propose unitary and nonunitary precoding that improve the robustness of spatial multiplexing in correlated channels.

A. Unknown Channel Statistics—Unitary Precoding

Our PEP analysis shows that the key source of performance degradation is the existence of unit error vectors and the vanishing diagonal entries of $\mathbf{R}_V$ due to clustered scattering. Furthermore, $\mathbf{R}_V$ may not be available at the transmitter. In such cases, it is of interest to develop techniques that minimize the occurrence of unit error vectors and also make spatial multiplexing robust to the vanishing diagonal elements of $\mathbf{R}_V$.

Designing general space–time coding schemes that are robust to correlated channels is an interesting open problem. In spatial multiplexing, since the transmitted vectors are not spatially coded, we are left with the option of applying a precoding transform to the transmitted vectors. In this paper, we focus on linear precoding transforms. Let $\mathbf{W}$ be a $M_t \times M_t$ precoding matrix that is applied to the output of the spatial multiplexer. Instead of transmitting codeword $\mathbf{s}_V$, we transmit $\mathbf{W}\mathbf{s}_V$ in Fig. 2. The receiver observes

$$\mathbf{y}_V = \sqrt{\frac{\rho}{M_t}}\mathbf{H}_V\mathbf{W}\mathbf{s}_V + \mathbf{v}_V. \tag{18}$$

The goal of this section is to find $\mathbf{W}$, without knowledge of $\mathbf{R}_V$ or $\mathbf{H}_V$, to improve the error-rate performance of the communication link.

Due to lack of knowledge of $\mathbf{R}_V$, we choose $\mathbf{W}$ to be unitary since it does not change the spatial distribution of power. A similar approach has been taken in [10], in which a precoding matrix is used to improve the robustness of spatial multiplexing systems in polarization channels. Their precoder assumes knowledge of a nondiagonal $\mathbf{R}_V$, parameterized by the cross-polarization discrimination, but otherwise serves to precondition the transmitted signal vectors. In our case, the precoder is designed to be robust to the nonvanishing diagonal entries of $\mathbf{R}_V$.

From the structure of the virtual representation, a DFT precoder may sound tempting; that is $\mathbf{s} = \mathbf{s}_V$ [we remove the $\mathbf{A}_T$ in Fig. 2(a)] so that $\mathbf{y}_V = \mathbf{H}_V\mathbf{A}_T^H\mathbf{s}_V + \mathbf{v}_V$. In this case, the transmitted codewords undergo a DFT $\mathbf{A}_T^H$ due to the channel propagation before encountering $\mathbf{H}_V$. However, such a DFT precoder is not attractive since it maps a constant error vector of the form $[c, c, \ldots, c]^T$ into the unit vector $[1, 0, \ldots, 0]^T$ which is detrimental to the PEP from Theorem 2. However, with some modification a DFT precoder may be useful, as elaborated below.

From the proof of Theorem 2, diversity advantage is compromised when there are cancellations in the product $\mathbf{R}_V(\mathbf{q}\otimes\mathbf{I}_{M_r})$. Furthermore, such cancellations are particularly acute for unit error vectors. We define a *robust unitary precoder* as one that satisfies the conditions in the following definitions. Let ϵ (not be confused with the ϵ in Theorem 2) be a design parameter.

- Define $J(\mathbf{W}) = \min_{\mathbf{q}\in\mathcal{E},n=0,\ldots,M_t-1}|[\mathbf{W}\mathbf{q}]_n|$. A *robust unitary precoder* satisfies $J(\mathbf{W}) \geq \epsilon$ for $n = 0, 1, \ldots, M_t - 1$ and for all $\mathbf{q} \in \mathcal{E}$, for some $\epsilon > 0$.
- An *optimal robust unitary precoder* is one that maximizes the threshold ϵ over the space of candidate precoders; that is

$$\mathbf{W}_o = \arg\max_{\mathbf{W}}\left[\arg\sup_\epsilon[J(\mathbf{W}) \geq \epsilon]\right]. \tag{19}$$

An optimal precoder may not be unique.

It is clear from Theorem 2 that a robust unitary precoder ensures that all elements of $\mathbf{W}\mathbf{q}$ are sufficiently nonvanishing for all $\mathbf{q} \in \mathcal{E}$. Then, from (13) in Theorem 1, we see that all nonvanishing elements in $\mathbf{R}_V$ contribute to the eigenvalues and, thus, contribute to the diversity gain by maximizing the rank. [Note that $\mathbf{q}$ is replaced with $\mathbf{W}\mathbf{q}$ in (13) with precoding.] An "optimum" precoder makes the modulus of all elements of $\mathbf{q} \in \mathcal{E}$ as large as possible to maximize the magnitude of eigenvalues in (13). This contributes to both the diversity gain and coding gain.

Finding an optimum robust unitary precoder is difficult since $\mathbf{W}_o$ depends on the number of transmit antennas, as well as the specific type of constellation. Optimization is difficult since robust unitary precoders satisfy a max–min criterion, where both the maximum and minimum are effectively taken over discrete spaces. However, a unitary $\mathbf{W}$ can be represented using a reduced set of coefficients. We use this fact to propose two different unitary precoders.

Diagonal Precoder: The simplest class of unitary matrices take the form

$$\mathbf{W} = \text{diag}\left\{1, e^{j\theta_1}, \ldots, e^{j\theta_{M_t-1}}\right\}. \tag{20}$$

Clearly, $\mathbf{W}^H\mathbf{W} = \mathbf{I}_{M_t}$ by design and there are only $M_t - 1$ free parameters since we can assume the first coefficient is one without loss of generality. However, this precoder does not solve the problem of avoiding unit error codeword vectors. In conjunction with a DFT precoder, though, this simple precoder may be effective. Suppose that we transmit signals of the form $\mathbf{s} = \mathbf{W}\mathbf{A}_T\mathbf{s}_V$. From (7) and (18), the channel equation in this case becomes $\mathbf{y}_V = \mathbf{H}_V\mathbf{A}_T^H\mathbf{W}\mathbf{A}_T\mathbf{s}_V + \mathbf{v}_V$. The intuition behind this precoder is that the matrix $\mathbf{A}_T^H\mathbf{W}\mathbf{A}_T$ (which is analogous to an all pass filter) will smear the error vectors over all frequencies (virtual angles) so that all elements of $\mathbf{A}_T^H\mathbf{W}\mathbf{A}_T\mathbf{q}$ [which replaces $\mathbf{q}$ in (13) in this case] will be nonvanishing for both unit error vectors, as well as constant error vectors (which cause problems in a DFT precoder).

For reasonable M_t, we propose the following exhaustive search technique.

Search 1: Uniformly quantize the phase values to $\Phi = \{2\pi n/N\}_{n=0}^{N-1}$ and let $\mathcal{W}$ be the set of N^{M_t-1} candidate precoding matrices. Choose the $\mathbf{W} \in \mathcal{W}$ such that (19) is satisfied.

We consider an example of a spatial multiplexing system with QPSK symbols and $M_t = M_r = 2$. The optimal precoding matrix is found and the simulation results are shown in Fig. 4 in

Section VI, where we show significant improvements with precoding. Given the performance improvements, a natural question is whether a nondiagonal unitary $\mathbf{W}$ can even further improve performance.

General Unitary Precoder: Consider a general unitary precoding matrix $\mathbf{W}$. Since the criterion in (19) does not readily admit a closed form solution, a numerical search is needed. One approach is to characterize a unitary matrix using a canonical representation, say using the Givens rotations, as in [11]. We then quantize the angles of the rotations to enumerate a family of candidate precoding matrices and optimize. For large numbers of antennas this becomes computationally intensive due to the number of parameters that need to be quantized. In this case, we propose the following design based on a random search.

Search 2: Generate a set of random complex $M_t \times M_t$ matrices from the complex Gaussian distribution. Let $\mathcal{W}$ be the set of orthogonal matrices constructed from the QR decomposition of the random matrices [21]. Choose the $\mathbf{W} \in \mathcal{W}$ such that (19) is satisfied.

Clearly, Search 2 is not guaranteed to be optimal. We find, however, that it is relatively easy to find a good $\mathbf{W}$ as we will show in Section VI

B. Known Channel Statistics—Nonunitary Precoding

Robust precoders do not exploit knowledge of the channel statistics and structure at the transmitter—the precoding matrix is designed to make the system robust to channel structure in clustered scattering environments. We now present a general nonunitary precoding design which can incorporate knowledge of channel structure at the transmitter. The second-order statistics of the channel, captured by $\mathbf{R}_V$, typically change more slowly than the channel realizations and, thus, it may be possible to convey $\mathbf{R}_V$ back to the transmitter.

Consider a nonunitary precoding matrix

$$\hat{\mathbf{W}} = \mathbf{D}\mathbf{W} = \mathrm{diag}\left\{\sqrt{p_0}, \ldots, \sqrt{p_{M_t-1}}\right\} \mathbf{W} \qquad (21)$$

where $\mathbf{W}$ is a unitary precoding matrix (as in the previous section) and p_n is chosen to satisfy the power constraint $\sum_{n=0}^{M_t-1} p_n = M_t$. Generally, if $p_n \neq 1$ for $n = 0, \ldots, M_t - 1$, $\hat{\mathbf{W}}$ is a nonunitary matrix. When $\mathbf{D} = \mathbf{I}_{M_t}$, $\hat{\mathbf{W}}$ becomes unitary. $\mathbf{D}$ could be interpreted as a scaling or power allocation matrix. Since we have the knowledge of $\mathbf{R}_V$ at the transmitter, instead of the robustness criterion, we seek a $\hat{\mathbf{W}}$ that minimizes the maximum PEP (9) for all possible error vectors. Let $\mathcal{D}$ and $\mathcal{W}$ denote sets of candidate $\mathbf{D}$ and $\mathbf{W}$ matrices, respectively. We define an optimal nonunitary precoding matrix as one that minimizes the maximum PEP of codeword error $\mathbf{D}\mathbf{W}\mathbf{q}$

$$(\mathbf{D}_o, \mathbf{W}_o)$$

$$= \arg \min_{\mathbf{D} \in \mathcal{D}, \mathbf{W} \in \mathcal{W}} \max_{\mathbf{q} \in \mathcal{E}} \left| \mathbf{I}_{M_r M_t} + \frac{\rho}{4 M_t N_o} \mathbf{R}_V \left[\mathbf{Q} \otimes \mathbf{I}_{M_r} \right] \right|^{-1}$$

$$(22)$$

where now $\mathbf{Q} = \mathbf{D}\mathbf{W}\mathbf{q}\mathbf{q}^H \mathbf{W}^H \mathbf{D} = \mathbf{D}\check{\mathbf{q}}\check{\mathbf{q}}^H \mathbf{D}$. The eigenvalues in (13) now become

$$\lambda_r(\mathbf{C}) = \sum_{n=0}^{M_t-1} p_n |[\check{\mathbf{q}}]_n|^2 [\mathbf{R}_V(n)]_{r,r}. \qquad (23)$$
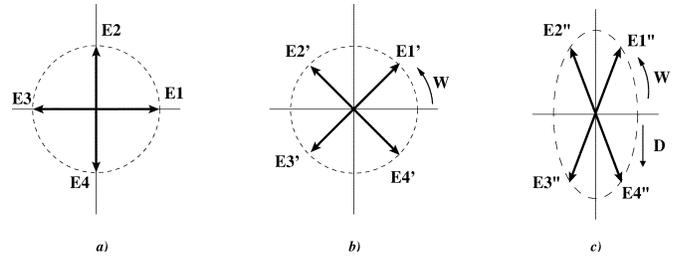
Fig. 3. Geometric representation of effects of precoding: a) original error constellation; b) unitary precoding; c) nonunitary precoding.

The use of unitary or nonunitary precoding can be motivated by considering the geometry of signal constellations and channel statistics. An example of signal error vector constellation is illustrated in Fig. 3(a). Since $\mathbf{H}_V$ may be sparse, the channel matrix has a certain "null space", captured by the vanishing diagonal elements of $\mathbf{R}_V$ in the context of PEP. If an error vector lies in this channel null space, it cannot be detected at the receiver and significantly compromises the PEP. In order to improve performance, unitary precoding rotates the error vector constellation to some angle to avoid error vectors (particularly unit error vectors) lying in the channel null space. This corresponds to making as many elements of $\check{\mathbf{q}}$ in (23) nonvanishing as possible over all $\mathbf{q} \in \mathcal{E}$. This contributes primarily to diversity gain and some to coding gain. Without knowledge of channel statistics at the transmitter, unitary precoding can provide robust performance since it tries to minimize the projection of error vectors onto the channel null space. When the transmitter has knowledge of $\mathbf{R}_V$, nonunitary precoding not only prevents an error vector from lying in the channel null space, but it also increases the projection of all the error vectors onto the channel range space via judicious choice of $\{p_n\}$. This corresponds to making $p_n |[\check{\mathbf{q}}]_n|^2$ as large as possible in (23) over all $\mathbf{q} \in \mathcal{E}$. This increases the coding gain compared with unitary precoding. A schematic illustrating the effects of unitary and nonunitary precoding is shown in Fig. 3(b) and (c). A similar channel geometry explanation was also given in [11], where only a particular 2×2 system with diagonal precoding was considered.

To further convey the intuition of unitary and nonunitary precoding, consider a concrete example of a 4×4 $\mathbf{H}_V = [\mathbf{h}_V(1) \quad \mathbf{h}_V(2) \quad \mathbf{h}_V(3) \quad \mathbf{h}_V(4)]$ of the form

$$\mathbf{H}_V = \begin{bmatrix} \cdot & \times & \times & \times \\ \cdot & \times & 0 & \times \\ \cdot & \times & 0 & \times \\ \cdot & 0 & 0 & \times \end{bmatrix} \qquad (24)$$

where $\cdot$ stands for nonvanishing elements with variance 0.02 and $\times$ stands for nonvanishing elements with unit variance. As we mentioned earlier, nonvanishing elements of $\mathbf{H}_V$ correspond to the nature of scattering. In this example, $[\mathbf{s}_V]_1$ is weakly supported by $\mathbf{h}_V(1)$ due to low power in $\mathbf{h}_V(1)$. Similarly, $[\mathbf{s}_V]_3$, that corresponds to the third-virtual transmit angle and $\mathbf{h}_V(3)$, only couples with one virtual receive angle and, thus, there is no diversity gain for this symbol. Therefore, we would expect the reliability of transmission of $[\mathbf{s}_V]_1$ and $[\mathbf{s}_V]_3$ to be poor. In

order to improve performance, we could spread the information transmitted by $[\mathbf{s}_V]_1$ and $[\mathbf{s}_V]_3$ over other transmit angles (corresponding to $\mathbf{h}_V(2)$ and $\mathbf{h}_V(4)$ with more power). Unitary precoding is designed to achieve this goal. On the other hand, $[\mathbf{s}_V]_1$ couples with receivers so weakly that it may not be used for transmission at all. Therefore, instead of equally distributing transmit power among all the virtual transmit angles, it is natural to allocate more transmit power to the more reliable transmit angles if such *a priori* information is available at the transmitter. The diagonal weighting matrix $\mathbf{D}$ in nonunitary precoding is designed to achieve this goal. It is also clear that in the extreme case of a fully populated $\mathbf{H}_V$, which is analogous to an i.i.d. channel [8], neither unitary precoding nor nonunitary precoding would be needed.

A numerical search is generally needed to obtain $\hat{\mathbf{W}}$. When the dimension is large, the search for $\hat{\mathbf{W}}$ that jointly optimizes $\mathbf{W}$ and $\mathbf{D}$ may be very computationally intensive. An alternative suboptimal solution is the following.

Search 3: Let $\mathbf{W}$ be an optimal robust unitary precoder obtained, for example, using Search 1 or Search 2 in Section V-A. Let $\mathcal{D}$ denote the space of $\mathbf{D}$ matrices obtained, for example, by quantization and normalization. Choose the $\mathbf{D}$ that satisfies (22) for the given $\mathbf{W}$ and known $\mathbf{R}_V$.

Using a $\mathbf{W}$ that is an optimal robust precoder has a number of important advantages. First, if $\mathbf{R}_V$ is not available, then the transmission is still robust. Second, there are fewer parameters to optimize in $\mathbf{D}$ than in $\mathbf{W}$. Since the channel statistics change over time on a relative slower time scale than changes in channel realizations, the optimization of $\mathbf{D}$ can be done in real-time. We now give another optimization criterion for $\mathbf{D}$ based on mutual information.[3] Assume that the codewords $\mathbf{s}_V$ are complex Gaussian distributed, $E[\mathbf{s}_V \mathbf{s}_V^H] = \mathbf{I}$, (the transmitted signal is now $\mathbf{DWs}_V$, and $E[\mathbf{DWs}_V \mathbf{s}_V^H \mathbf{W}^H \mathbf{D}^H] = \mathbf{D}^2$). Therefore, it can be shown using Jensen's and Hadamard inequalities, the average mutual information between $\mathbf{y}$ and $\mathbf{s}$ is [2]

$$\mathcal{I}(\mathbf{y}, \mathbf{s}) = E_{\mathbf{H}_V} \log \det \left(\mathbf{I}_{M_t} + \frac{\rho}{M_t N_o} \mathbf{H}_V^H \mathbf{H}_V \mathbf{D}^2 \right)$$
$$\leq \sum_{n=0}^{M_t-1} \log \left(1 + \frac{p_n}{\sigma_n} \right) \qquad (25)$$

where $\sigma_n = M_t N_o/(\nu_n \rho)$, $\nu_n = E[\mathbf{h}_V^H(n)\mathbf{h}_V(n)]$. The optimal $\mathbf{D}$ then maximizes (25) under the power constraint. This is a standard waterfilling problem [23] and the solution is

$$p_n = (\mu - \sigma_n)^+ \qquad (26)$$

here $(x)^+$ denotes the positive part of x and μ is chosen so that $\sum_{n=0}^{M_t-1}(\mu - \sigma_n)^+ = M_t$.

C. Virtual Transmit Antenna Selection

The mutual information based power allocation procedure in (26) explicitly illustrates an important point that is implicit in (22). In some cases, a particular virtual transmit angle will contribute a negligible amount of power to the receiver [e.g., the first virtual antenna in (24)] making this virtual angle *infeasible*—the power budget may not be sufficient to

[3]In [22], it is argued that maximizing mutual information also minimizes PEP.
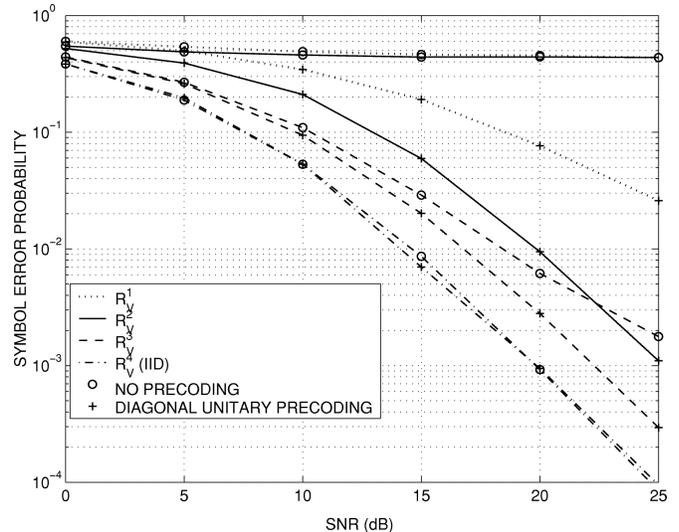
Fig. 4. No precoding versus diagonal precoding as a function of the number of nonvanishing elements of $\mathbf{R}_V$.

allocate any power to this virtual antenna. This problem can be alleviated by applying a technique known as stochastic antenna subset selection [24]. The idea is to allow only a subset of K_t of the M_t virtual angles to be used. The size of K_t can be determined *a priori* from $\mathbf{R}_V$ and then Search 3 can be performed over candidate subset. Alternatively, the number and locations of nonzero p_n in (26) can be used to determine the active virtual transmit antennas and then Search 3 can be used to find $\mathbf{D}$ for these antennas.

VI. NUMERICAL RESULTS

A. Unitary Precoding Without Channel Statistics

In this experiment, we consider a spatial multiplexing system with quaternary phase-shift keying (QPSK) symbols and $M_t = M_r = 2$. We consider the performance with and without precoding as a function of the number of nonzero diagonal elements of $\mathbf{R}_V$. We denote $\mathbf{R}_V$ with k nonzero elements as $\mathbf{R}_V^k$, $k = 1, 2, 3, 4$, $\mathbf{R}_V^1 = \text{diag}\{1, 0, 0, 0\}$, $\mathbf{R}_V^2 = \text{diag}\{1, 1, 0, 0\}$, $\mathbf{R}_V^3 = \text{diag}\{1, 1, 1, 0\}$, $\mathbf{R}_V^4 = \text{diag}\{1, 1, 1, 1\}$. We used 100 000 Monte Carlo simulations to estimate the average symbol-error rate for each $\mathbf{R}_V$.

First consider the case of no precoding as illustrated in Fig. 4. The symbol-error rate is about 0.5 for $\mathbf{R}_V^1$ and $\mathbf{R}_V^2$ since the symbol stream transmitted from the second virtual antenna is lost in the null-space of the channel. For the case of $\mathbf{R}_V^3$, there is a first-order diversity, since the minimum rank is one as follows from Theorem 2. Finally, $\mathbf{R}_V = \mathbf{I}_4$ corresponds to a rich channel and yields a second-order diversity gain and maximum coding gain.

Now we show how even diagonal precoding improves the diversity advantage in sparse-scattering environments. By quantizing the angles with $N = 100$, we used Search 1 to find the following optimal diagonal precoding matrix with $\epsilon = 0.6299$

$$\mathbf{W} = \text{diag}\{1, e^{j0.93}\}. \qquad (27)$$

The performance improvement obtained using this precoder is illustrated in Fig. 4 and compared with the case of no pre-
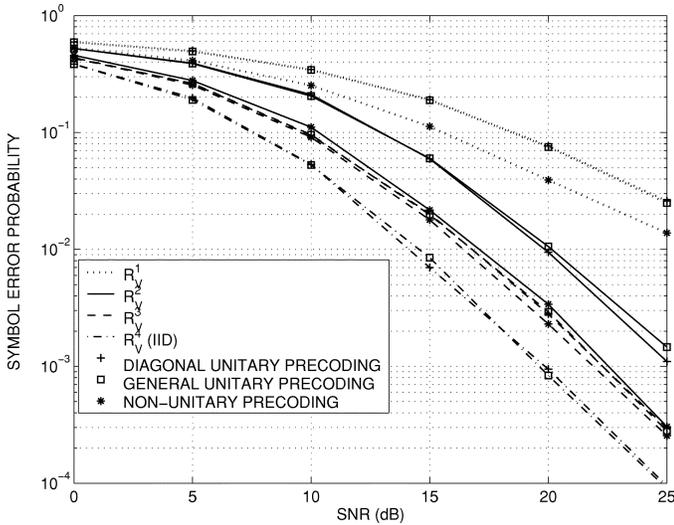
Fig. 5.   Diagonal, general unitary, and nonunitary precoding as a function of the number of nonvanishing elements of $\mathbf{R}_V$.

Fig. 6.   Effects of unitary and nonunitary precoding on 2×2 virtual channels with two nonzero elements.

coding. Even with a single nonzero element, as predicted by Theorem 2, there is a first-order diversity gain since both data streams get coupled to the nonvanishing $[\mathbf{H}_V]_{1,1}$ due to rotation. As $\mathbf{R}_V$ fills out with 2, 3, and 4 nonzero diagonal entries there is a second-order diversity advantage as predicted by Theorem 2. Since each additional nonzero element increases the received power, there is an increase in coding gain as predicted by Theorem 1.

Now consider a general unitary precoder. Using Search 2, we searched over $10\,000$ random 2×2 matrices and found the following optimal precoding matrix with $\epsilon = 0.6488$:

$$\mathbf{W} = \begin{bmatrix} -0.2436 - 0.7934i & 0.4194 + 0.3679i \\ -0.5194 - 0.2037i & -0.8167 + 0.1478i \end{bmatrix}. \quad (28)$$

From Theorem 1, we expect that there is a marginal coding gain compared with precoder (27) due to the marginal difference in ϵ. In Fig. 5, we compare the precoder (28) to (27) and find that they have comparable performance. This is possible due to the suboptimality of the search procedure and also due to the fact that our design criterion is not exactly motivated by the PEP. The comparison shows that the diagonal precoder jointly used with DFT has the advantage of less intensive search and comparable performance to the general unitary precoder.

B. Nonunitary Precoding With Channel Statistics

To illustrate how nonunitary precoding can further improve the interaction between the codeword and the channel, we compare a unitary precoder with a nonunitary precoder that exploits knowledge of $\mathbf{R}_V$. First, we search for unitary precoder that yields the $\mathbf{W}$ in (28). Then we use (22) to search for $\mathbf{D}$ for each $\mathbf{R}_V^k$, $k = 1, 2, 3$. For $\mathbf{R}_V^1$ and $\mathbf{R}_V^2$, the optimal $\mathbf{D}$ from Search 3 is trivial. The optimal $\mathbf{D}$ for $\mathbf{R}_V^3$ is $\mathrm{diag}\{1.1576, 0.8124\}$. The comparison of unitary precoding and nonunitary precoding is shown in Fig. 5. Nonunitary precoding outperforms unitary precoding by up to 2 dB when $\mathbf{R}_V$ is sparse ($\mathbf{R}_V^1$, $\mathbf{R}_V^2$). For denser scattering ($\mathbf{R}_V^3$), the improvement of nonunitary precoding over unitary precoding is apparent though reduced. For the fully populated channel ($\mathbf{R}_V = \mathbf{I}$), the nonunitary precoding matrix and
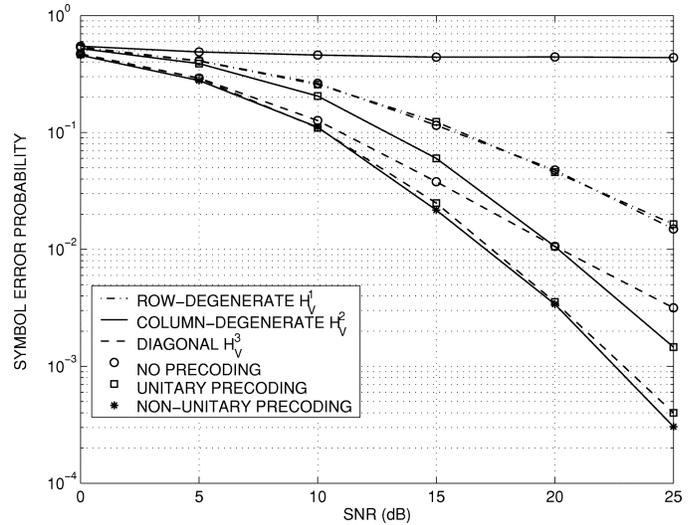
unitary precoding are identical since the optimal $\mathbf{D}$ in this case is $\mathbf{I}$.

C. Impact of Precoding on Degenerate Virtual Channels

We have shown that precoding can improve the performance of spatial multiplexing in correlated channels. Now we use an example to explicitly illustrate the diversity and coding advantage due to precoding using Theorem 1 and 2. We use three different possible 2×2 representations for $\mathbf{H}_V$: row degenerate, column degenerate, and diagonal

$$\mathbf{H}_V^1 = \begin{bmatrix} \times & \times \\ 0 & 0 \end{bmatrix}, \mathbf{H}_V^2 = \begin{bmatrix} \times & 0 \\ \times & 0 \end{bmatrix}, \mathbf{H}_V^3 = \begin{bmatrix} \times & 0 \\ 0 & \times \end{bmatrix}. \quad (29)$$

For $\mathbf{H}_V^1$, Theorem 2 shows that the minimum diversity advantage of this system is one and the only eigenvalue is given by Theorem 1 as $\lambda = \|\mathbf{q}\|^2$, which is determined by the chosen constellation. Clearly, the precoder cannot improve performance since both the diversity advantage and the minimum eigenvalue are fixed. Without precoding, $\mathbf{H}_V^2$ provides no diversity gain by Theorem 2. However, with precoding, Theorem 1 reveals that $\mathbf{C}$ has two nonzero eigenvalues $\lambda_1 = \lambda_2 = |[\mathbf{Wq}]_1|^2$, so the diversity gain is increased to two and the degenerate column suggests a trivial power allocation scheme. The coding gain is further improved by using nonunitary precoding. Similarly, for $\mathbf{H}_V^3$, without precoding, diversity advantage is one according to Theorem 2 and this is improved to two by precoding according to Theorem 1. The coding gain is given by $(\lambda_1 \lambda_2)^{1/2} = |[\mathbf{Wq}]_1 [\mathbf{Wq}]_2|$. Theorem 1 shows that we also improve the coding gain by maximizing the eigenvalues by our search criterion. The comparison of three virtual representation with and without precoding is shown in Fig. 6 and confirms our intuition and analytical results.

D. Virtual Transmit Antenna Selection Based on Second-Order Channel Statistics

In this experiment, we conclude by comparing all the proposed precoding approaches in a spatial multiplexing system
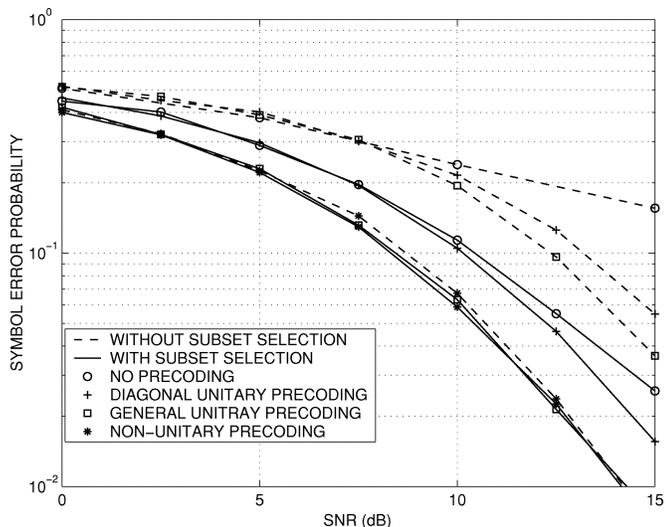
Fig. 7. Effects of unitary, nonunitary precoding, and subset selection on the 4×4 virtual channel in (24).

with QPSK symbols and $M_t = M_r = 4$ for the channel with the virtual representation in (24).

Consider two scenarios: with and without subset selection. In the first case, all four virtual transmit antennas are used ($M_t = 4$). In the second case, based on the weak coupling of virtual transmit antenna $[\mathbf{s}_V(1)]$, we use only $\{[\mathbf{s}_V]_2, [\mathbf{s}_V]_3, [\mathbf{s}_V]_4\}$ corresponding to $\{\mathbf{h}_V(2), \mathbf{h}_V(3), \mathbf{h}_V(4)\}$ and $M_t = 3$.

Using Search 1 with $N = 100$, we found optimal diagonal precoding matrices for $M_t = 3$ and $M_t = 4$. Using Search 2 with 10 000 candidate matrices, we found optimal general unitary precoding matrices for $M_t = 3$ and $M_t = 4$. For the nonunitary precoding scenario, $\mathbf{D}$ was determined by the power allocation scheme (26). The error-rate curves for all these scenarios are shown in Fig. 7. The case with no precoding provides minimal diversity gain as expected. Unitary precoding without channel statistics provides diversity gain as predicted by Theorem 2. Due to the weak coupling of the first virtual angle, the power on $[\mathbf{s}_V]_1$ is wasted. Subset selection fixes this problem, but at the expense of rate since now three symbols and not four are sent in each time period. Interestingly, nonunitary precoding with power allocation has dramatically improved performance both with and without subset selection. The reason is that in either case the power is distributed only on the three major virtual angles. Without subset selection, due to the presence of unitary precoding, the information stream that would have been transmitted on the first virtual antenna is mixed on the remaining three antennas and can still be decoded. Finally, note that the allocation (26) is obtained by maximizing the mutual information with a Gaussian codebook; it does not necessarily optimize the error performance when restricted to finite alphabet constellations. Thus, better precoder optimizations based on $\mathbf{R}_V$ may even further improve performance.

VII. CONCLUSION

We have investigated the design of spatial multiplexing techniques for correlated fading channels via the virtual channel representation. We derived bounds for the PEP in spatial mul-

tiplexing using the virtual channel matrix $\mathbf{H}_V$. The approximately uncorrelated nature of $\mathbf{H}_V$ greatly simplified this analysis. In particular, we derived exact expressions for the minimum rank of the matrix governing the PEP, as well as exact characterization of its eigenvalues. The analysis in the virtual representation yields many useful insights on the interaction between the codeword space and the channel and its effects on performance. In particular, a key insight to avoid performance degradation in correlated channels is to avoid unit error codeword vectors and to exploit virtual transmit antennas that are strongly coupled to the channels. These insights are leveraged to develop effective multiplexing techniques for correlated MIMO channels. Unitary precoding is proposed to make spatial multiplexing robust to channel correlation when channel statistics are unknown at the transmitter. When channel statistics are known at the transmitter, we showed that power allocation coupled with unitary precoding is a simple and effective means for exploiting this information. Currently, we are investigating extension of the work for space–time trellis coding and linear dispersion coding for correlated MIMO channels.

REFERENCES

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

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecomm.*, vol. 10, no. 6, 1999.

[3] 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.

[4] 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.

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

[6] A. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception (DTDR)," U. S. Patent 5345599, Sept. 6, 1994.

[7] D. S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502–513, Mar. 2000.

[8] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Processing*, vol. 50, pp. 2563–2579, Oct. 2002.

[9] A. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Mag.*, vol. 14, pp. 49–83, Nov. 1997.

[10] H. Bolcskei, R. U. Nabar, V. Erceg, D. Gesbert, and A. J. Paulray, "Performance of spatial multiplexing in the presence of polarization diversity," *IEEE Trans. Signal Processing*, vol. 50, pp. 2553–2562, Oct. 2002.

[11] R. Nabar, H. Bolcskei, and A. J. Paulraj, "Transmit optimization for spatial multiplexing in the presence of spatial fading correlation," in *Proc. IEEE GLOBECOM*, vol. 1, 2001, pp. 131–135.

[12] H. Sampath and A. J. Paulraj, "Linear precoding for space-time coded systems with known fading correlations," in *Proc. 35th Asilomar Conf. Signals, Systems, Computers*, vol. 1, 2001, pp. 246–251.

[13] G. B. Giannakis and S. Zhou, "Optimal transmit-diversity precoders for random fading channels," in *Proc. IEEE GLOBECOM*, vol. 3, 2000, pp. 1839–1843.

[14] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuit Syst.*, vol. 25, pp. 772–781, Sept. 1978.

[15] J. Fuhl, A. F. Molisch, and E. Bonek, "Unified channel model for mobile radio systems with smart antennas," *IEE Proc. Radar, Sonar Navig.*, vol. 145, pp. 32–41, Feb. 1998.

[16] D. Gore, R. W. Heath Jr., and A. J. Paulraj, "Performance analysis of spatial multiplexing in correlated channels," *IEEE Trans. Commun.*, Mar. 2002, submitted for publication.

[17] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Combined array processing and space-time coding," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1121–1128, May 1999.

[18] G. L. Turin, "The characteristic function of Hermitian quadratic forms in complex normal variables," *Biometrika*, vol. 47, pp. 199–201, 1960.

[19] K. Liu and A. M. Sayeed, "Space-time D-block codes via the virtual MIMO channel representation," *IEEE Trans. Wireless Commun.*, 2003, submitted for publication.

[20] 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," *IEEE Trans. Commun.*, vol. 47, pp. 527–537, Apr. 1999.

[21] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[22] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1804–1824, July 2002.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1992.

[24] D. A. Gore, R. W. Heath Jr., and A. J. Paulraj, "Transmit selection in spatial multiplexing systems," *IEEE Commun. Lett.*, vol. 6, pp. 491–494, Nov. 2002.

Zhihong Hong (S'98–M'02) received the B.S. degree from Tsinghua University, Beijing, China, in 1994, and the M.S. and Ph.D. degrees, in 1998 and 2002, respectively, from North Carolina State University, Raleigh, all in electrical engineering.

From 1999 to 2002, he was a Research Assistant at the Center for Advanced Computing and Communication (CACC), North Carolina State University. Since February 2002, he has been with the University of Wisconsin at Madison, as a Postdoctoral Research Associate. His research interests include advanced coding and modulation in wireless communications, multicarrier transmissions, and wireless networks.

Ke Liu (S'99) received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 1999, the M.S. degree in electrical engineering, and the M.A. degree in mathematics, both from the University of Wisconsin at Madison, in 2001 and 2002, respectively, where he is currently working toward the Ph.D. degree in electrical engineering.

His research interests include information theory, signal processing and coding in wireless communications and wireless networks.

Robert W. Heath, Jr. (S'96–M'01) received the B.S. and M.S. degrees from the University of Virginia, Charlottesville, in 1996 and 1997, respectively, and the Ph.D. degree from Stanford University, Stanford, CA, in 2002, all in electrical engineering.

In October 1998, he was part of the founding team of Iospan Wireless Inc., formerly Gigabit Wireless Inc., San Jose, CA. From 1999 to 2001, he served as a Senior Consultant for Iospan Wireless Inc. In January 2002, he joined the Electrical and Computer Engineering Department, The University of Texas at Austin, where he serves as an Assistant Professor as part of the Wireless Networking and Communications Group. His research lab, the Wireless Systems Innovations Laboratory, focuses on the theory, design, and practical implementation of wireless systems. His current research interests are coding, modulation, equalization, and resource allocation for MIMO wireless communication systems. He has published over 25 refereed conference and journal papers and has four patents issued, all in the area of wireless communications, signal processing, and communication theory.
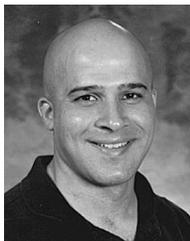
Akbar M. Sayeed (S'89–M'97–SM'02) received the B.S. degree from the University of Wisconsin at Madison, in 1991, and the M.S. and Ph.D. degrees from the University of Illinois, Urbana-Champaign, all in electrical and computer engineering, in 1993 an 1996, respectively.

While at the University of Illinois, he was a Research Assistant in the Coordinated Science Laboratory and was also the Schlumberger Fellow in signal processing from 1992 to 1995. From 1996–1997, he was a Postdoctoral Fellow at Rice University, Houston, TX. Since August 1997, he has been with the University of Wisconsin, where he is currently an Assistant Professor in electrical and computer engineering. His research interests are in wireless communications, sensor networks, statistical signal processing, wavelets, and time-frequency analysis.

Dr. Sayeed received the NSF CAREER Award in 1999 and the ONR Young Investigator Award in 2001. He served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 1999 to 2002.