Adaptive MIMO-OFDM Based on Partial Channel State Information

Pengfei Xia, *Student Member, IEEE*, Shengli Zhou, *Member, IEEE*, and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—Relative to designs assuming no channel knowledge at the transmitter, considerably improved communications become possible when adapting the transmitter to the intended propagation channel. As perfect knowledge is rarely available, transmitter designs based on partial (statistical) channel state information (CSI) are of paramount importance not only because they are more practical but also because they encompass the perfect- and no-knowledge paradigms. In this paper, we first provide a partial CSI model for orthogonal frequency division multiplexed (OFDM) transmissions over multi-input multi-output (MIMO) frequency-selective fading channels. We then develop an adaptive MIMO-OFDM transmitter by applying an adaptive two-dimensional (2-D) coder-beamformer we derived recently on each OFDM subcarrier, along with an adaptive power and bit loading scheme across OFDM subcarriers. Relying on the available partial CSI at the transmitter, our objective is to maximize the transmission rate, while guaranteeing a prescribed error performance, under the constraint of fixed transmit-power. Numerical results confirm that the adaptive 2-D space-time coder-beamformer (with two basis beams as the two "strongest" eigenvectors of the channel's correlation matrix perceived at the transmitter) combined with adaptive OFDM (power and bit loaded with M-ary quadrature amplitude modulated (QAM) constellations) improves the transmission rate considerably.**

***Index Terms*—Adaptive modulation, beamforming, MIMO OFDM, partial channel state information.**

I. INTRODUCTION

TRANSMITTER designs adapted to the intended propagation channel are capable of improving both performance and rate of communication links. The resulting channel-adaptive transmissions adjust parameters such as power levels, constellation sizes, coding schemes, and modulation types, depending on the channel state information (CSI) that is assumed available to the transmitter [2], [9], [14]. The potential improvement increases considerably when multiple transmit- and receive-antennas are deployed [7], [25]. However, as symbol rates increase in broadband wireless applications, the underlying multi-input multi-output (MIMO) channels exhibit strong frequency selectivity. By transforming frequency-selective channels to an equivalent set of frequency-flat subchannels, orthogonal frequency division multiplexing (OFDM) has emerged as an attractive transmission modality because it comes with low-complexity (de)modulation, equalization, and decoding to mitigate frequency-selective fading effects [3], [27].

These considerations motivate well adaptive MIMO OFDM, but the challenge is on whether and what type of CSI can be made practically available to the transmitter in a wireless setting, where fading channels are randomly varying. Certainly, this is less of an issue in wireline links, where the counterpart of adaptive OFDM, known as discrete multi-tone (DMT), has been standardized for digital subscriber line modems. Both single-input single-output (SISO) and MIMO versions of DMT [5], [20] assume that *perfect CSI* is available at the transmitter. Although it is reasonable for wireline links, perfect-CSI-based adaptive transmissions developed for SISO [14] and MIMO OFDM wireless systems [28] can be justified only when the fading is sufficiently slow. On the other hand, the proliferation of space-time coding research we have witnessed lately testifies to the efforts put toward the other extreme: nonadaptive (and thus conservative) designs requiring *no CSI* to be available at the transmitter.

As no-CSI leads to robust but rather pessimistic designs, and perfect CSI is probably a utopia for most wireless links, recent efforts geared toward quantification and exploitation of *partial (or statistical) CSI* promise to have great practical value because they are capable of offering the "jack of both trades." As with perfect CSI, partial CSI is made available either through a feedback channel from the receiver to the transmitter or when the transmitter acts also as receiver in a time- or frequency-division duplex operation. The difference is that outdated CSI (caused e.g., by feedback delay), uncertain CSI (induced e.g., by channel estimation or prediction errors), and limited CSI (appearing e.g., with quantized feedback), are all accounted for statistically under partial CSI but are ignored when perfect CSI is assumed. Using terms such as mean or covariance feedback to specify the type of partial CSI, existing designs have focused on MIMO transmissions over flat fading channels and adapt transmitter parameters based either on capacity-based [17], [18], [26] or performance-based criteria [12], [32]. Very recently, partial CSI has also been considered for SISO OFDM systems over frequency-selective channels [21], [29].

Building on our recent work on adaptive modulation over MIMO *flat-fading* channels [31], [33], we design in this paper adaptive MIMO-OFDM transmissions over *frequency-selective*

Manuscript received December 22, 2002; revised April 30, 2003. This work was supported by the ARL/CTA under Grant DAAD19-01-2-011 and the National Science Foundation under Grant 01-0516. This work was presented in part at the IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Rome, Italy, June 15–18, 2003. The associate editor coordinating the review of this paper and approving it for publication was Dr. Rick S. Blum.

P. Xia and G. B. Giannakis are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (email: pfxia@ece.umn.edu; georgios@ece.umn.edu).

S. Zhou was with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA. He is now with Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269 USA (e-mail: shengli@engr.uconn.edu).

Digital Object Identifier 10.1109/TSP.2003.819986

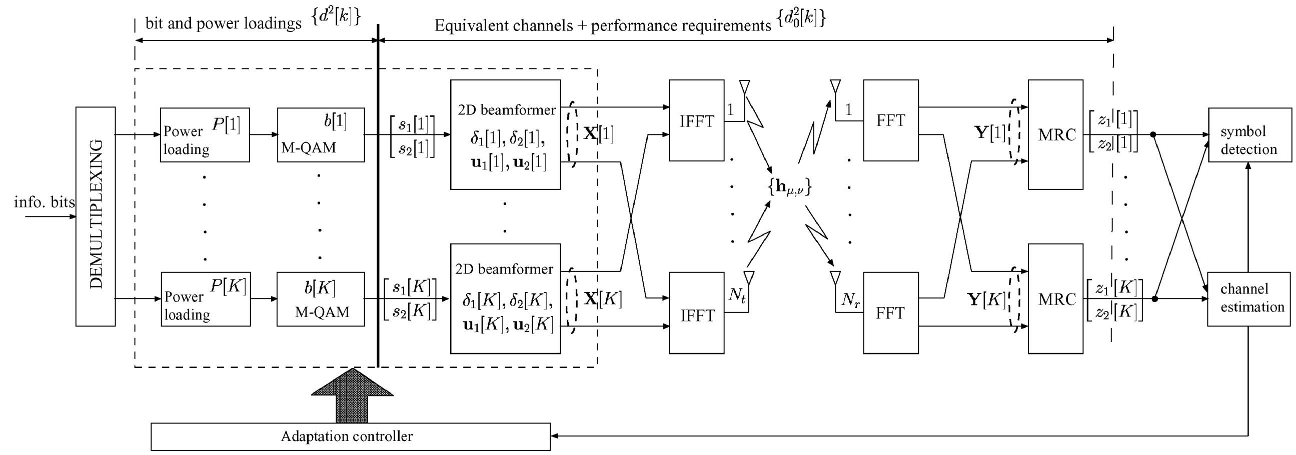

Fig. 1. Discrete-time equivalent baseband system model.

fading channels, based on partial CSI. The transmitter here applies the adaptive 2-D space-time coder-beamformer of [31]–[33] on each OFDM subcarrier, with the power and bits adaptively loaded across subcarriers, to maximize transmission rate under performance and power constraints. This problem is challenging because information bits and power should be optimally allocated over space *and* frequency, but its solution is equally rewarding because high-performance high-rate transmissions can be enabled over MIMO frequency-selective channels. Our novelties include the following:

- quantification of partial CSI for frequency-selective MIMO channels and formulation of a constrained optimization problem with the goal of maximizing rate for a given power budget and a prescribed BER performance (Section II);
- design of an optimal adaptive MIMO OFDM transmitter as a concatenation of an adaptive modulator and an adaptive 2-D coder-beamformer (Section III-A);
- identification of a suitable threshold metric that encapsulates the allowable power and bit combinations and enables joint optimization of the adaptive modulator-beamformer (Section III-A);
- unification under the umbrella of adaptive MIMO OFDM based on partial CSI of many existing SISO and MIMO designs based on partial- or perfect-CSI (Section III-A).
- incorporation of existing algorithms for joint power and bit loading across MIMO OFDM subcarriers, based on partial CSI (Section III-B);
- illustration of the tradeoffs emerging among rate, complexity, and the reliability of partial CSI, using simulated examples (Section IV).

Throughout the paper, we will adopt the following notational conventions.

Bold upper and lower case letters denote matrices and column vectors, respectively; $(\cdot)^T$ and $(\cdot)^{\mathcal{H}}$ denote transpose and Hermitian transpose, respectively; $\|\cdot\|_F$ stands for the Frobenius norm; $\mathrm{E}[\cdot]$ denotes the ensemble average; $[\cdot]_p$ denotes the pth entry of a vector, and $[\cdot]_{p,q}$ denotes the (p,q)th entry of a matrix; $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

II. System Model and Problem Statement

We deal with an OFDM system equipped with K subcarriers and N_t transmit and N_r receive antennas signaling over a MIMO frequency-selective fading channel. Per OFDM subcarrier, we deploy the adaptive two-dimensional (2-D) coder-beamformer developed in [31]–[33] that combines Alamouti's space time block coding (STBC) [1] with transmit beamforming.[1] We should state at the outset that higher dimensional coder beamformers based on orthogonal STBC with $N_t > 2$ [24] can be also applied, as we detail in Appendix C. However, we will demonstrate in Section IV that the 2-D coder-beamformer strikes desirable performance-rate-complexity tradeoffs, and for this reason, we focus our design on the 2-D case, which being simpler hits the "sweet spot" also in practice.

To apply the 2-D coder-beamformer per subcarrier, we pair two consecutive OFDM symbols to form one space-time-coded OFDM block. Due to frequency selectivity, different subcarriers experience generally different channel attenuation. Hence, in addition to adapting the 2-D coder-beamformer on each subcarrier, the total transmit-power should also be judiciously allocated to different subcarriers, based on the available CSI at the transmitter. Fig. 1 depicts the equivalent discrete-time baseband model of the system under consideration, which we will specify next.

We reserve n to index space-time-coded OFDM blocks (pairs of OFDM symbols), and let k denote the subcarrier index, i.e., $k \in \{0, 1, \ldots, K-1\}$. If $P[n;k]$ stands for the power allocated to the kth subcarrier of the nth block, then depending on $P[n;k]$, we will select a constellation (alphabet) $\mathcal{A}[n;k]$, consisting of $M[n;k]$ constellation points. In addition to square QAMs with $M[n;k] = 2^{2i}$ that have been used extensively in adaptive modulation [9], we will also consider rectangular QAMs with $M[n;k] = 2^{2i+1}$. We will focus on those rectangular QAMs that can be implemented with two independent PAMs: one for the In-phase branch with size $\sqrt{2M[n;k]}$ and the other for the Quadrature-phase branch with size $\sqrt{M[n;k]/2}$,

[1]In our context, a beam toward a particular direction is formed by a set of steering weights (coefficients) multiplying the information symbols to be transmitted per antenna; see also [12], [18], [26], and [32].

as those studied in [30]. Thanks to the independence between I-Q branches, this type of rectangular QAM incurs modulation and demodulation complexity similar to square QAM. This will facilitate the adaptive transmitter implementation.

For each block time-slot n, the input to each 2-D coder-beamformer used per subcarrier entails two information symbols ($s_1[n;k]$ and $s_2[n;k]$) drawn from $\mathcal{A}[n;k]$, with each one conveying

$$b[n;k] = \log_2(M[n;k]) \tag{1}$$

bits of information. These two information symbols will be space-time coded, power-loaded, and multiplexed by the 2-D beamformer to generate an $N_t \times 2$ space-time (ST) matrix as

$$\mathbf{X}[n;k] = \underbrace{[\mathbf{u}_1^*[n;k], \mathbf{u}_2^*[n;k]]}_{:=\mathbf{U}^*[n;k]} \cdot \underbrace{\begin{bmatrix} \sqrt{\delta_1[n;k]} & 0 \\ 0 & \sqrt{\delta_2[n;k]} \end{bmatrix}}_{:=\mathbf{D}[n;k]} \cdot \underbrace{\begin{bmatrix} s_1[n;k] & -s_2^*[n;k] \\ s_2[n;k] & s_1^*[n;k] \end{bmatrix}}_{:=\mathbf{S}[n;k]} \tag{2}$$

where $\mathbf{S}[n;k]$ is the well-known Alamouti ST code matrix; $\mathbf{U}[n;k]$ is the multiplexing matrix formed by two $N_t \times 1$ basis-beam vectors $\mathbf{u}_1[n;k]$ and $\mathbf{u}_2[n;k]$; and $\mathbf{D}[n;k]$ is the corresponding power allocation matrix on these two basis-beams with $0 \le \delta_1[n;k], \delta_2[n;k] \le 1$, and $\delta_1[n;k] + \delta_2[n;k] = 1$. In the two time slots corresponding to the two OFDM symbols involved in the nth ST coded block, the two columns of $\mathbf{X}[n;k]$ are transmitted on the kth subcarrier over N_t transmit antennas.

We suppose that the MIMO channel is invariant during each space-time coded block but is allowed to vary from block to block. Let $\mathbf{h}_{\mu\nu}[n] := [h_{\mu,\nu}[n;0], \ldots, h_{\mu,\nu}[n;L]]^T$ be the baseband equivalent FIR channel between the μth transmit and the νth receive antenna during the nth block, where $1 \le \mu \le N_t$, $1 \le \nu \le N_r$, and L is the maximum channel order of all $N_t N_r$ channels. With $\mathbf{f}_k := [1, e^{j2\pi k/N}, \ldots, e^{j2\pi kL/N}]^T$, the frequency response of $\mathbf{h}_{\mu\nu}[n]$ on the kth subcarrier is

$$H_{\mu,\nu}[n;k] = \sum_{l=0}^{L} h_{\mu\nu}[n;l] e^{\frac{-j2\pi kl}{N}} = \mathbf{f}_k^{\mathcal{H}} \mathbf{h}_{\mu\nu}[n]. \tag{3}$$

Let $\mathbf{H}[n;k]$ be the $N_t \times N_r$ matrix having $H_{\mu\nu}[n;k]$ as its (μ,ν)th entry. To isolate the transmitter design from channel estimation issues at the receiver, we suppose the following.

AS0) The receiver has perfect knowledge of the channel $\mathbf{H}[n;k]$, $\forall n, k$.

With $\mathbf{Y}[n;k]$ denoting the nth received block on the kth subcarrier, we can express the input-output relationship per subcarrier and ST coded OFDM block as

$$\begin{aligned}\mathbf{Y}[n;k] &= \mathbf{H}^T[n;k]\mathbf{X}[n;k] + \mathbf{W}[n;k] \\ &= \mathbf{H}^T[n;k]\mathbf{U}^*[n;k]\mathbf{D}[n;k]\mathbf{S}[n;k] + \mathbf{W}[n;k]\end{aligned} \tag{4}$$

where $\mathbf{W}[n;k]$ stands for the additive white Gaussian noise (AWGN) at the receiver with each entry having variance $N_0/2$ per real and imaginary dimension. Based on (4), one can view our coded-beamformed MIMO OFDM transmissions per subcarrier as an Alamouti transmission with ST matrix $\mathbf{S}[n;k]$ passing through an equivalent channel matrix $\mathbf{B}^T[n;k] := \mathbf{H}^T[n;k]\mathbf{U}^*[n;k]\mathbf{D}[n;k]$. With knowledge of this equivalent channel and maximum ratio combining (MRC) at the receiver, it is not difficult to verify that each information symbol is thus passing through an equivalent scalar channel with I/O relationship

$$z_i[n;k] = h_{\text{eqv}}[n;k] s_i[n;k] + w_i[n;k], \quad i = 1, 2 \tag{5}$$

where in our case, the equivalent channel is (c.f. [1])

$$\begin{aligned}h_{\text{eqv}}[n;k] = \|\mathbf{B}[n;k]\|_F = \Big[\delta_1[n;k] \left\|\mathbf{H}^{\mathcal{H}}[n;k]\mathbf{u}_1[n;k]\right\|_F^2 \\ + \delta_2[n;k] \left\|\mathbf{H}^{\mathcal{H}}[n;k]\mathbf{u}_2[n;k]\right\|_F^2\Big]^{\frac{1}{2}}.\end{aligned} \tag{6}$$

Having introduced the system model, we next specify the partial CSI at the transmitter.

A. *Partial CSI for Frequency-Selective MIMO Channels*

For flat-fading multiantenna channels, the notion of mean feedback has been introduced in [26] to account for channel uncertainty at the transmitter, where the fading channels are modeled as Gaussian random variables with nonzero mean, and white covariance. In this paper, we will adopt this mean feedback model on each OFDM subcarrier. Specifically, we adopt the following model.

AS1) On each subcarrier k, the transmitter obtains an unbiased channel estimate $\overline{\mathbf{H}}[n;k]$ either through a feedback channel, during a duplex mode operation, or by predicting the channel from past blocks. The transmitter treats this "nominal channel" $\overline{\mathbf{H}}[n;k]$ as deterministic, and in order to account for CSI uncertainty, it adds a "perturbation" term. The partial CSI of the true $N_t \times N_r$ MIMO channel $\mathbf{H}[n;k]$ at the transmitter is thus perceived as

$$\check{\mathbf{H}}[n;k] = \overline{\mathbf{H}}[n;k] + \mathbf{\Xi}[n;k], \quad k = 0, 1, \ldots, K-1 \tag{7}$$

where $\mathbf{\Xi}[n;k]$ is a random matrix Gaussian distributed according to $\mathcal{CN}(\mathbf{0}_{N_t \times N_r}, N_r \sigma_\epsilon^2[n;k]\mathbf{I}_{N_t})$. The variance $\sigma_\epsilon^2[n;k]$ encapsulates the CSI reliability on the kth subcarrier.

The "nominal-plus-perturbation" or "mean feedback" model of AS1) has been documented for flat-fading channels [18], [26], [32]. However, it has not been considered for frequency-selective fading channels. We next illustrate a practical scenario that justifies AS1).

Motivating Example (delayed channel feedback): Suppose that the FIR channel taps have been acquired perfectly at the receiver and are fed back to the transmitter with a certain delay but without errors thanks to powerful error control codes used in the feedback. Let us also assume that the following conditions hold true.

i) The $L+1$ taps $\{h_{\mu\nu}[n;l]\}_{l=0}^{L}$ in $\mathbf{h}_{\mu\nu}[n]$ are uncorrelated but not necessarily identically distributed (to account for e.g., exponentially decaying power profiles). Each tap is zero-mean Gaussian with variance $\sigma_{\mu\nu}^2[l]$. Hence, $\mathbf{h}_{\mu\nu}[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_{\mu\nu})$, where

$\mathbf{\Sigma}_{\mu\nu} := \mathrm{diag}(\sigma^2_{\mu\nu}[0], \ldots, \sigma^2_{\mu\nu}[l])$. This assumption has also been adopted in, e.g., [16].

ii) The FIR channels $\{\mathbf{h}_{\mu\nu}[n]\}_{\mu=1,\nu=1}^{N_t,N_r}$ between different transmit- and receive-antenna pairs are independent. This requires antennas to be spaced sufficiently far apart from each other.

iii) All FIR channels have the same total energy on the average: $\sigma_h^2 = \mathrm{tr}\{\mathbf{\Sigma}_{\mu\nu}\}, \forall \mu, \nu$. This is reasonable in practice since the multiantenna transmissions experience the same scattering environment.

iv) All channel taps are time varying according to Jakes' model with Doppler frequency f_d.

At the nth block, suppose we obtain the channel feedback $\{\mathbf{h}^f_{\mu\nu}[n]\}_{\mu=1,\nu=1}^{N_t,N_r}$ that corresponds to the true channels N_b blocks earlier, i.e., $\mathbf{h}^f_{\mu\nu}[n] = \mathbf{h}_{\mu\nu}[n - N_b]$. Suppose each space time coded block has time duration T_b seconds. Then, $\mathbf{h}^f_{\mu\nu}[n]$ is drawn from the same Gaussian distribution as $\mathbf{h}_{\mu\nu}[n]$ but $N_b T_b$ seconds ahead. Let $\rho := J_0(2\pi f_d N_b T_b)$ denote the correlation coefficient specified by the Jakes' model, where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind. The MMSE predictor of $\mathbf{h}_{\mu\nu}[n]$, based on $\mathbf{h}^f_{\mu\nu}[n]$ and i), is $\overline{\mathbf{h}}_{\mu\nu}[n] = \rho \mathbf{h}^f_{\mu\nu}[n]$. To account for the prediction imperfections, the transmitter forms an estimate $\check{\mathbf{h}}_{\mu\nu}[n]$ as

$$\check{\mathbf{h}}_{\mu\nu}[n] = \overline{\mathbf{h}}_{\mu\nu}[n] + \boldsymbol{\xi}_{\mu\nu}[n] \tag{8}$$

where $\boldsymbol{\xi}_{\mu\nu}[n]$ is the prediction error. Under i), it is easy to verify that

$$\boldsymbol{\xi}_{\mu\nu}[n] \sim \mathcal{CN}\left(\mathbf{0}, \left(1 - |\rho|^2\right) \mathbf{\Sigma}_{\mu\nu}\right), \quad \forall \mu, \nu. \tag{9}$$

The mean feedback model (8) on the channel taps is easily translated to the CSI in (7) on the channel frequency response per subcarrier. Based on (8), we obtain the matrices in (7) with (μ, ν)th entries: $[\check{\mathbf{H}}[n;k]]_{\mu\nu} = \mathbf{f}_k^{\mathcal{H}} \check{\mathbf{h}}_{\mu\nu}[n]$, $[\overline{\mathbf{H}}[n;k]]_{\mu\nu} = \mathbf{f}_k^{\mathcal{H}} \overline{\mathbf{h}}_{\mu\nu}$, and $[\mathbf{\Xi}[n;k]]_{\mu\nu} = \mathbf{f}_k^{\mathcal{H}} \boldsymbol{\xi}_{\mu\nu}[n]$. Using i), ii), and (9), we can also verify that $\mathbf{\Xi}[n;k]$ has covariance matrix $N_r(1 - |\rho|^2)\sigma_h^2 \mathbf{I}_{N_t}$. Notice that in this case, the uncertainty indicators $\sigma_\epsilon^2[n;k] = (1 - |\rho|^2)\sigma_h^2$ are common to all subcarriers. ■

Notwithstanding, the *partial* CSI described by AS1) has also unifying value. When $K = 1$, it boils down to the partial CSI for flat fading channels [18], [26], [31]–[33]. With $\sigma_\epsilon^2 = 0$, it reduces to the *perfect* CSI of the MIMO setup considered in [20], [28]. When $N_t = N_r = 1$, it simplifies to the partial CSI feedback used for SISO FIR channels [21], [29]. Furthermore, with $N_t = N_r = 1$ and $\sigma_\epsilon^2 = 0$, it is analogous to perfect CSI feedback for wireline DMT channels [3], [5].

B. Formulation of a Constrained Optimization Problem

Our objective in this paper is to optimize the MIMO-OFDM transmissions in Fig. 1 based on partial CSI available at the transmitter. Specifically, we want to maximize the transmission rate subject to a power constraint while maintaining a target BER performance on each subcarrier. Let $\overline{\mathrm{BER}}[n;k]$ denote the perceived average BER at the transmitter on the kth subcarrier of the nth block, and let $\overline{\mathrm{BER}}_0[k]$ stand for the prescribed target BER on the kth subcarrier. The target BERs can be identical or different across subcarriers, depending on system specifications. Recall that each space-time-coded block conveys two symbols $s_1[n;k]$ and $s_2[n;k]$ and, thus, $2b[n;k]$ bits of information on the kth subcarrier. Our goal is thus formulated as the following constrained optimization problem:

$$\begin{aligned} \text{maximize} \quad & 2\sum_{k=0}^{K-1} b[n;k] \\ \text{subject to} \quad & \text{C1. } \overline{\mathrm{BER}}[n;k] = \overline{\mathrm{BER}}_0[k], \quad \forall k \\ & \text{C2. } \sum_{k=0}^{K-1} P[n;k] = P_{\mathrm{total}} \\ & \quad\;\; \text{and} \quad P[n;k] \geq 0, \quad \forall k \\ & \text{C3. } b[n;k] \in \{0, 1, 2, 3, 4, 5, 6, \ldots\} \end{aligned} \tag{10}$$

where P_{total} is the total power available to the transmitter per block.

The constrained optimization problem in (10) calls for joint adaptation of the following parameters:

- power and bit loadings $\{P[n;k], b[n;k]\}_{k=0}^{K-1}$ across subcarriers;
- basis-beams per subcarrier $\{\mathbf{u}_1[n;k], \mathbf{u}_2[n;k]\}_{k=0}^{K-1}$;
- power splitting between the two basis-beams per subcarrier $\{\delta_1[n;k], \delta_2[n;k]\}_{k=0}^{K-1}$.

Compared with the constant-power transmissions over flat-fading MIMO channels [31], [33], the problem here is more challenging, due to the needed power loading across OFDM subcarriers, which in turn depends on the 2-D beamformer optimization per subcarrier. Intuitively speaking, our problem amounts to loading power and bits optimally across space and frequency, based on partial CSI.

III. Adaptive MIMO-OFDM With 2-D Beamforming

For notational brevity, we drop the block index n since our transmitter optimization is going to be performed on a per block basis. Our transmitter includes an inner stage (adaptive beamforming) and an outer stage (adaptive modulation). Instrumental to both stages is a threshold metric $d_0^2[k]$, which determines *allowable combinations*[2] of ($P[k]$, $b[k]$) so that the prescribed $\overline{\mathrm{BER}}_0[k]$ is guaranteed.

A. Adaptive 2-D Beamforming Based on Partial CSI

In this subsection, we determine the basis beams $\mathbf{u}_1[k], \mathbf{u}_2[k]$, and the corresponding percentages $\delta_1[k]$, $\delta_2[k]$ of the power $P[k]$, for a fixed (but allowable) combination of ($P[k]$, $b[k]$). Let T_s be the OFDM symbol duration with the cyclic prefix removed, and without loss of generality, let us set $T_s = 1$. With this normalization, the constellation chosen for the kth subcarrier has average energy $\mathcal{E}_s[k] = P[k]T_s = P[k]$ and contains $M[k] = 2^{b[k]}$ signaling points. If $d^2_{\min}[k]$ denotes the minimum square Euclidean distance for this constellation, we will find it convenient to work with the scaled distance metric $d^2[k] := d^2_{\min}[k]/4$ because for QAM constellations, it holds that [31], [33]

$$d^2_{\min}[k] = 4d^2[k] = 4g\left(b[k]\right)\mathcal{E}_s[k] = 4g\left(b[k]\right)P[k] \tag{11}$$

[2]It should be intuitively clear at this point that with finite power, one cannot allow arbitrarily large constellation sizes.

where the constant $g(b)$ depends on whether the chosen constellation is rectangular or square QAM:

$$g(b) := \begin{cases} \frac{6}{5\cdot 2^b - 4}, & b = 1, 3, 5, \ldots \\ \frac{6}{4\cdot 2^b - 4}, & b = 2, 4, 6, \ldots. \end{cases} \tag{12}$$

Notice that $d^2[k]$ summarizes the power and constellation (bit) loading information that the adaptive modulator passes on to the coder beamformer. The latter relies on $d^2[k]$ and the partial CSI to adapt its design to meet constraint C1. To proceed with the adaptive beamformer design, we therefore need to analyze the BER performance of the scalar equivalent channel per subcarrier, with input $s_i[k]$ and output $z_i[k]$, as described by (5). For each (deterministic) realization of $h_{\mathrm{eqv}}[k]$, the BER when detecting $s_i[k]$ in the presence of AWGN in (5) can be approximated as

$$\mathrm{BER}[k] \approx 0.2 \exp\left(-\frac{h_{\mathrm{eqv}}^2[k] d^2[k]}{N_0}\right) \tag{13}$$

where the validity of the approximation has also been confirmed in [31] and [33]. Based on our partial CSI model in AS1), the transmitter perceives $h_{\mathrm{eqv}}[k]$ as a random variable and evaluates the *average* BER performance on the kth subcarrier as

$$\overline{\mathrm{BER}}[k] \approx 0.2 \mathrm{E}\left[\exp\left(-\frac{h_{\mathrm{eqv}}^2[k] d^2[k]}{N_0}\right)\right]. \tag{14}$$

We will adapt our basis beams $\mathbf{u}_1[k]$, $\mathbf{u}_2[k]$ to minimize $\overline{\mathrm{BER}}[k]$ for a given $d^2[k]$ based on partial CSI. To this end, we consider the eigen decomposition of the "nominal channel" per subcarrier (here, the kth)

$$\begin{aligned} \overline{\mathbf{H}}[k]\overline{\mathbf{H}}^{\mathcal{H}}[k] &= \overline{\mathbf{U}}_{\mathrm{H}}[k]\mathbf{\Lambda}_{\mathrm{H}}[k]\overline{\mathbf{U}}_{\mathrm{H}}^{\mathcal{H}}[k], \quad \text{with} \\ \overline{\mathbf{U}}_{\mathrm{H}}[k] &:= [\overline{\mathbf{u}}_{\mathrm{H},1}[k], \ldots, \overline{\mathbf{u}}_{\mathrm{H},N_t}[k]] \\ \mathbf{\Lambda}_{\mathrm{H}}[k] &:= \mathrm{diag}\left(\lambda_1[k], \ldots, \lambda_{N_t}[k]\right) \end{aligned} \tag{15}$$

where $\overline{\mathbf{U}}_{\mathrm{H}}[k]$ is unitary, and $\mathbf{\Lambda}_{\mathrm{H}}[k]$ contains on its diagonal the eigenvalues in a nonincreasing order: $\lambda_1[k] \geq \ldots \geq \lambda_{N_t}[k] \geq 0$. As proved in [31]–[33], the optimal $\mathbf{u}_1[k]$ and $\mathbf{u}_2[k]$ minimizing the $\overline{\mathrm{BER}}[k]$ in (14) are

$$\mathbf{u}_1[k] = \overline{\mathbf{u}}_{\mathrm{H},1}[k], \quad \mathbf{u}_2[k] = \overline{\mathbf{u}}_{\mathrm{H},2}[k]. \tag{16}$$

Notice that the columns of $\overline{\mathbf{U}}_{\mathrm{H}}[k]$ are also the eigenvectors of the channel correlation matrix $\mathrm{E}\{\check{\mathbf{H}}[k]\check{\mathbf{H}}^{\mathcal{H}}[k]\} = \overline{\mathbf{H}}[k]\overline{\mathbf{H}}^{\mathcal{H}}[k] + N_r\sigma_\epsilon^2[k]\mathbf{I}_{N_t}$ that is perceived by the transmitter based on partial CSI [32]. Hence, the basis beams $\mathbf{u}_1[k]$ and $\mathbf{u}_2[k]$ adapt to the two eigenvectors of the perceived channel correlation matrix, corresponding to the two largest eigenvalues.

Having obtained the optimal basis beams, to complete our beamformer design, we have to decide how to split the power $P[k]$ between these two basis beams.

With the optimal basis beams in (16), the equivalent scalar channel is [c.f. (5)]

$$h_{\mathrm{eqv}}^2[k] = \delta_1[k]\left\|\check{\mathbf{H}}^{\mathcal{H}}[k]\overline{\mathbf{u}}_{\mathrm{H},1}[k]\right\|^2 + \delta_2[k]\left\|\check{\mathbf{H}}^{\mathcal{H}}[k]\overline{\mathbf{u}}_{\mathrm{H},2}[k]\right\|^2. \tag{17}$$

For $i = 1, 2$, the vector $\check{\mathbf{H}}^{\mathcal{H}}[k]\overline{\mathbf{u}}_{\mathrm{H},i}[k]$ in (17) is Gaussian distributed with $\mathcal{CN}(\overline{\mathbf{H}}^{\mathcal{H}}[k]\overline{\mathbf{u}}_{\mathrm{H},i}[k], \sigma_\epsilon^2[k]\mathbf{I}_{N_r})$. Furthermore, we have that $\|\overline{\mathbf{H}}^{\mathcal{H}}[k]\overline{\mathbf{u}}_{\mathrm{H},i}[k]\|^2 = \lambda_i[k]$. For an arbitrary vector $\mathbf{a} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{\Sigma})$, the following identity holds true [23, eq. (15)]:

$$\mathrm{E}\left\{\exp(-\mathbf{a}^{\mathcal{H}}\mathbf{a})\right\} = \frac{\exp\left(-\boldsymbol{\mu}^{\mathcal{H}}(\mathbf{I}+\mathbf{\Sigma})^{-1}\boldsymbol{\mu}\right)}{\det(\mathbf{I}+\mathbf{\Sigma})}. \tag{18}$$

Substituting (17) into (14) and applying (18), we obtain

$$\overline{\mathrm{BER}}[k] \approx 0.2 \prod_{\mu=1}^{2}\left[\left(\frac{1}{1+\frac{\delta_\mu[k]d^2[k]\sigma_\epsilon^2[k]}{N_0}}\right)^{N_r} \cdot \exp\left(-\frac{\frac{\lambda_\mu[k]\delta_\mu[k]d^2[k]}{N_0}}{1+\frac{\delta_\mu[k]d^2[k]\sigma_\epsilon^2[k]}{N_0}}\right)\right]. \tag{19}$$

Equation (19) shows that the power splitting percentages $\delta_1[k]$, $\delta_2[k]$ depend on $\lambda_1[k]$, $\lambda_2[k]$, and $d^2[k]$. Their optimum values can be found by minimizing (19) to obtain, as in [32, eq. (54)]

$$\delta_1[k] = \min\left(\bar{\delta}_1[k], 1\right), \quad \delta_2[k] = \max\left(\bar{\delta}_2[k], 0\right) \tag{20}$$

where, with $\mathcal{K}_\mu[k] := \lambda_\mu[k]/(N_r\sigma_\epsilon^2[k])$ and $m_\mu[k] := (1+\mathcal{K}_\mu[k])^2/(1+2\mathcal{K}_\mu[k])$, $\mu = 1, 2$, we have

$$\bar{\delta}_\mu[k] = \frac{m_\mu[k]}{\sum_i m_i[k]} + \frac{m_\mu[k]}{\frac{d^2[k]\sigma_\epsilon^2[k]}{N_0}} \times \left(\frac{\sum_i \frac{m_i[k]}{1+\mathcal{K}_i[k]}}{\sum_i m_i[k]} - \frac{1}{1+\mathcal{K}_\mu[k]}\right), \quad \mu = 1, 2. \tag{21}$$

The solution in (20) guarantees that $0 \leq \delta_2[k] \leq \delta_1[k] \leq 1$, and $\delta_1[k] + \delta_2[k] = 1$. Based on the partial CSI ($\overline{\mathbf{H}}[k]$, $\sigma_\epsilon^2[k]$), (16) and (20) provide the 2-D coder-beamformer design with the minimum $\overline{\mathrm{BER}}[k]$ that is adapted to a given $d^2[k]$ output of the adaptive modulator. Because this minimum $\overline{\mathrm{BER}}[k]$ depends on $d^2[k]$, the natural question at this point is the following: For which value of $d^2[k]$, call it $d_0^2[k]$, will the minimum $\overline{\mathrm{BER}}[k]$ reach the target $\overline{\mathrm{BER}}_0[k]$?

1) Determining the Threshold Metrics $\{d_0^2[k]\}_{k=0}^{K-1}$: Before addressing this question, we first establish that $\overline{\mathrm{BER}}[k]$ in (19), with $\{\delta_i[k]\}_{i=1}^2$ specified in (20), is a monotonically decreasing function of $d^2[k]$.

Lemma: Given partial CSI as in AS1), the $\overline{\mathrm{BER}}[k]$ in (19) is a monotonically decreasing function of $d^2[k]$. Hence, there exists a threshold $d_0^2[k]$ for which $\overline{\mathrm{BER}}[k] \leq \overline{\mathrm{BER}}_0[k]$ if and only if $d^2[k] \geq d_0^2[k]$. The threshold $d_0^2[k]$ is found by solving (19) with respect to $d^2[k]$ when $\overline{\mathrm{BER}}[k] = \overline{\mathrm{BER}}_0[k]$.

Proof: A detailed proof requires the derivative of $\overline{\mathrm{BER}}[k]$ with respect to $d^2[k]$ over two possible scenarios: $\delta_2[k] = 0$, and $\delta_2[k] > 0$, as indicated by (20). We have verified that this derivative is always less than zero for any given $d^2[k]$. However, we will skip the lengthy derivation and provide an intuitive justification instead. Suppose that $\delta_1[k]$ and $\delta_2[k]$ are optimized as in (20) for a given $d^2[k]$. Now, let us increase $d^2[k]$ by an amount Δ_d. Even when $\delta_1[k]$ and $\delta_2[k]$ are fixed to previously optimized values (i.e., even if the 2-D coder-beamformer is nonadaptive), the corresponding BER decreases since signaling with larger minimum distance always leads to better performance. With the minimum constellation distance $d^2[k] + \Delta_d$, optimizing $\delta_1[k]$

and $\delta_2[k]$ will further decrease the BER. Hence, increasing $d^2[k]$ decreases $\overline{\text{BER}}[k]$ monotonically. ■

This lemma implies that we can obtain the desirable $d_0^2[k]$ by solving $\overline{\text{BER}}[k] = \overline{\text{BER}}_0[k]$ with respect to $d^2[k]$. However, since no closed-form solution appears possible, we have to rely on a 1-D numerical search.

To avoid the numerical search, we next propose a simple, albeit *approximate*, solution for $d_0^2[k]$. Notice that (19) is nothing but the average BER of an $2N_r$-branch diversity combining system, with N_r branches undergoing Rician fading with Rician factor $\mathcal{K}_1[k] = \lambda_1[k]/(N_r\sigma_\epsilon^2[k])$, whereas the other N_r branches are experiencing Rician fading with Rician factor $\mathcal{K}_2[k] = \lambda_2[k]/(N_r\sigma_\epsilon^2[k])$. Approximating a Rician distribution by a Nakagami-m distribution [22, pp. 49–50], we can approximate the $\overline{\text{BER}}[k]$ in (19) by (see also [32, eq. (38)])

$$\overline{\text{BER}}'[k] \approx \frac{1}{5}\prod_{\mu=1}^{2}\left(1+\delta_\mu[k]\frac{(1+\mathcal{K}_\mu[k])\,d^2[k]\sigma_\epsilon^2[k]}{m_\mu[k]\cdot N_0}\right)^{-m_\mu[k]N_r} \tag{22}$$

where m_μ is defined after (20). It can be easily verified that $\overline{\text{BER}}'[k]$ in (22) is also monotonically decreasing as $d^2[k]$ increases. Setting $\overline{\text{BER}}'[k] = \overline{\text{BER}}_0[k]$, we can solve for $d_0^2[k]$ using the following two-step approach.

Step 1) Suppose that $d_0^2[k]$ can be found with $\delta_2[k] > 0$. Substituting (21) into (22), we obtain

$$d_0^2[k] = \left[\frac{A_0[k]\cdot\left(5\overline{\text{BER}}_0[k]\right)^{\frac{-1}{(A_0[k]N_r)}}}{\prod_{\mu=1}^{2}\left(1+\mathcal{K}_\mu[k]\right)^{\frac{m_\mu[k]}{A_0[k]}}} - B_0[k]\right]\cdot\frac{N_0}{\sigma_\epsilon^2[k]} \tag{23}$$

where

$$A_0[k] := \sum_{i=1}^{2} m_i[k], \quad B_0[k] := \sum_{i=1}^{2}\frac{m_i[k]}{1+\mathcal{K}_i[k]}. \tag{24}$$

To verify the validity of the solution in (23), let us substitute $d_0^2[k]$ into (21). If $\bar{\delta}_2[k] > 0$ is satisfied, then (23) yields the desired solution. Otherwise, we go to Step 2.

Step 2) When Step 1 fails to find the desired $d_0^2[k]$ with $\delta_2[k] > 0$, we set $\delta_2[k] = 0$. Substituting $\delta_1[k] = 1$ and $\delta_2[k] = 0$ into (22), we have

$$d_0^2[k] = \frac{\left(5\overline{\text{BER}}_0[k]\right)^{\frac{-1}{(m_1[k]N_r)}} - 1}{\frac{(1+\mathcal{K}_1[k])}{m_1[k]}}\cdot\frac{N_0}{\sigma_\epsilon^2[k]}. \tag{25}$$

This approximate solution for $d_0^2[k]$ avoids numerical search, thus reducing the transmitter complexity. We will compare the approximate solution with the exact solution found via numerical search in Section IV.

2) Special Cases: We next detail some important special cases.

Special Case 1—MIMO OFDM with 1-D beamforming based on partial CSI: The 1-D beamforming is subsumed by the 2-D beamforming if one fixes *a priori* the power percentages to $\delta_1[k] = 1$, and $\delta_2[k] = 0$. In this case, $d_0^2[k]$ can be found in closed-form as in (25).

Special Case 2—SISO-OFDM based on partial CSI: The single-antenna OFDM based on partial CSI [21], [29] can be obtained from (7) by setting $N_t = N_r = 1$. In this case, $\lambda_1[k] = |\overline{H}[k]|^2$, where $\overline{H}[k]$ is the "nominal channel" on the kth subcarrier. Hence, (25) yields $d_0^2[k]$ in this case too, after setting $N_r = 1$, and $\mathcal{K}_1 := |\overline{H}[k]|^2/\sigma_\epsilon^2[k]$.

Special Case 3—MIMO-OFDM based on perfect CSI: With $\sigma_\epsilon^2[k] = 0$, the adaptive beamformer on each OFDM subcarrier reduces to the 1-D beamformer with $\delta_2[k] = 0$. This corresponds to the MIMO-OFDM system studied in [28], when cochannel interference (CCI) is absent. In this special case, no Nakagami approximation is needed, and the BER performance in (19) simplifies to

$$\overline{\text{BER}}[k] = 0.2\exp\left(\frac{-d^2[k]\lambda_1[k]}{N_0}\right) \tag{26}$$

which leads to a simpler calculation of the threshold metrics as

$$d_0^2[k] = \left[-\ln\left(5\overline{\text{BER}}_0[k]\right)\right]\frac{N_0}{\lambda_1[k]}. \tag{27}$$

Special Case 4—Wireline DMT systems: The conventional wireline channel in DMT systems [5], [13] can be incorporated in our partial CSI model[3] by setting $N_t = 1, N_r = 1$, and $\sigma_\epsilon^2[k] = 0$. In this case, the threshold metric $d_0^2[k]$ is given by (27) with $\lambda_1[k] = |H[k]|^2$.

B. Adaptive Modulation Based on Partial CSI

With $d_0^2[k]$ encapsulating the allowable $(P[k], b[k])$ pairs per subcarrier, we are ready to pursue joint power and bit loading across OFDM subcarriers to maximize the data rate. It turns out that after suitable interpretations, many existing power and bit loading algorithms developed for DMT systems [4], [11], [15] can be applied to our adaptive MIMO-OFDM system based on partial CSI. We first show how the classical Hughes-Hartogs algorithm (HHA) [11] can be utilized to obtain the optimal power and bit loadings.

1) Optimal Power and Bit Loading: As the loaded bits in (10) assume finite (nonnegative integer) values, a globally optimal power and bit allocation exists. Given any allocation of bits on all subcarriers, we can construct it in a step-by-step bit loading manner, with each step adding a single bit on a certain subcarrier, and incurring a cost quantified by the additional power needed to maintain the target BER performance. This hints toward the idea behind the Hughes Hartogs algorithm (HHA) [11]: At each step, it tries to find which subcarrier supports one additional bit with the least required additional power. Notice that the HHA belongs to the class of greedy algorithms (see, e.g., [6, ch. 16]) that have found many applications such as the minimum spanning tree and Huffman encoding.

Recalling our results in Section III-A, the minimum required power to maintain i bits on the kth subcarrier with threshold

[3]One major difference is that our additive noise is white, whereas in DMT systems, the noise variance is subcarrier dependent.

metric $d_0^2[k]$ is $d_0^2[k]/g(i)$. Therefore, the power cost incurred when loading the ith bit to the kth subcarrier is

$$c(k,i) = \frac{d_0^2[k]}{g(i)} - \frac{d_0^2[k]}{g(i-1)}, \quad i \geq 1, \forall k. \tag{28}$$

For $i=1$, we set $g(i-1)=\infty$, and thus, $c(k,1) = d_0^2[k]/g(1)$. In the following algorithm, we will use $P_{\rm rem}$ to record the remaining power after each bit loading step, $b_c[k]$ to store the number of bits already loaded on the kth subcarrier, and $P_c[k]$ to denote the amount of power currently loaded on the kth subcarrier. Now, we are ready to describe the greedy algorithm for joint power and bit loading of our adaptive MIMO-OFDM based on partial CSI.

The Greedy Algorithm

1) Initialization: Set $P_{\rm rem} = P_{\rm total}$. For each subcarrier, set $b_c[k] = P_c[k] = 0$, and compute $d_0^2[k]$.
2) Choose the subcarrier that requires the least power to load one additional bit, i.e., select

$$k_0 = \arg\min_k c\left(k, b_c[k]+1\right). \tag{29}$$

3) If the remaining power cannot accommodate it, i.e., if $P_{\rm rem} < c(k_0, b_c[k_0]+1)$, then exit with $P[k] = P_c[k]$, and $b[k] = b_c[k]$. Otherwise, load one bit to subcarrier k_0, and update state variables as

$$P_{\rm rem} = P_{\rm rem} - c\left(k_0, b_c[k_0]+1\right) \tag{30}$$
$$P_c[k_0] = P_c[k_0] + c\left(k_0, b_c[k_0]+1\right) \tag{31}$$
$$b_c[k_0] = b_c[k_0] + 1. \tag{32}$$

4) Loop back to step 2.

The greedy algorithm yields a "1-bit optimal" solution since it offers the optimal strategy at each step when only a single bit is considered. In general, the 1-bit optimal solution obtained by a greedy algorithm may not be overall optimal [6]. However, for our problem at hand, we establish, in Appendix A, the following.

Proposition 1: The power and bit loading solution $\{P[k], b[k]_{k=0}^{K-1}\}$ to which the greedy algorithm converges in a finite number of steps is overall optimal.

Notice that the optimal bit loading solution may not be unique. This happens when two or more subcarriers have identical $d_0^2[k]$ under their respective (and possibly different) performance requirements. However, a unique solution can be always obtained, after establishing simple rules to break possible ties arising in (29).

Allowing for both rectangular and square QAM constellations, the greedy algorithm loads one bit at a time. However, only square QAMs are used in many adaptive systems. If only square QAMs are selected during our adaptive modulation stage, we can then load two bits in each step of the greedy algorithm and thereby halve the total number of iterations. It is natural to wonder whether restricting the class to square QAMs has a major impact on performance. Fortunately, as the following proposition establishes, limiting ourselves to square QAMs only incurs marginal loss (see Appendix B for a proof).

Proposition 2: Relative to allowing for both rectangular and square QAMs, the adaptive MIMO-OFDM with only square QAMs incurs up to one bit loss (on the average) per transmitted space-time coded block that contains two OFDM symbols.

Compared with the total number of bits conveyed by two OFDM symbols, the one bit loss is negligible when using only square QAM constellations. However, reducing the number of possible constellations by 50% simplifies the practical adaptive transmitter design. These considerations advocate only square QAM constellations for adaptive MIMO-OFDM modulation (this excludes also the popular BPSK choice).

The reason behind Proposition 2 is that square QAMs are more power efficient than rectangular QAMs [c.f. (12)]. With K subcarriers at our disposal, it is always possible to avoid usage of less efficient rectangular QAMs and save the remaining power for other subcarriers to use power-efficient square QAMs. Interestingly, this is different from the adaptive modulation over flat fading channels considered in [31] and [33], where the transmit power is constant and considerable loss (one bit every two symbols on average) is involved, if only square QAM constellations are adopted.

2) Practical Considerations: The complexity of the optimal greedy algorithm is on the order of $\mathcal{O}(N_{bits}K)$, where N_{bits} is the total number of bits loaded, and K is the number of subcarriers. In addition, it is considerable when N_{bits} and K are large. Alternative low-complexity power and bit loading algorithms have been developed for DMT applications; see, e.g., [4], [15], and [19]. Notice that [4] and [19] study a dual problem: optimal allocation of power and bits to minimize the total transmission power with a target number of bits. Interestingly, the truncated water-filling solution in [4] can be modified and used in our transmitter design, whereas the fast algorithm of [19] cannot, since it requires knowledge of the total number of bits to start with. In spite of low-complexity, the algorithm in [4] is suboptimal and may result in a considerable rate loss due to the truncation operation. The most interesting algorithm is the fast Lagrange bi-sectional search proposed in [15] that provides an optimal solution, while having complexity as low as [4]. Hence, for our adaptive transmissions, we recommend [15] in practice.

Before we conclude this section, we briefly summarize the overall adaptation procedure for our adaptive MIMO-OFDM design based on partial CSI.

1) Basis beams per subcarrier $\{\mathbf{u}_1[k], \mathbf{u}_2[k]\}_{k=0}^{K-1}$ are adapted first using (16) to obtain an adaptive 2-D coder-beamformer for each subcarrier.
2) Power and bit loading $\{b[k], P[k]\}_{k=0}^{K-1}$ is then jointly performed across all subcarriers, using the algorithm in [15] that offers optimality at complexity lower than the greedy algorithm.
3) Finally, power splitting between the two basis beams on each subcarrier $\{\delta_1[k], \delta_2[k]\}_{k=1}^{K}$ is decided using (20).

IV. Numerical Results

We present numerical results in this section, based on the delayed-feedback paradigm of Section II-A. We set $K = 64$ and $L = 5$ and assume that the channel taps are i.i.d. with covariance matrix $\mathbf{\Sigma}_{\mu\nu} = (1/L+1)\mathbf{I}_{L+1}$. We allow for

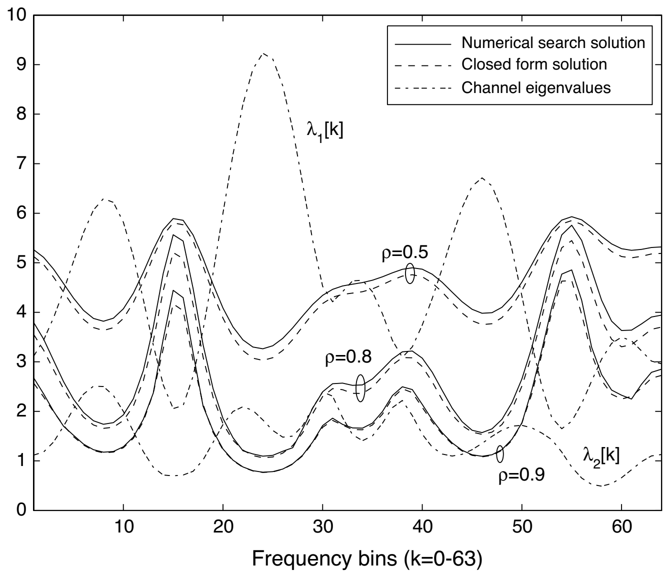

Fig. 2. Threshold distances $\{d_0^2[k]\}_{k=0}^{K-1}$ with $\overline{\mathrm{BER}}_0 = 10^{-3}$.

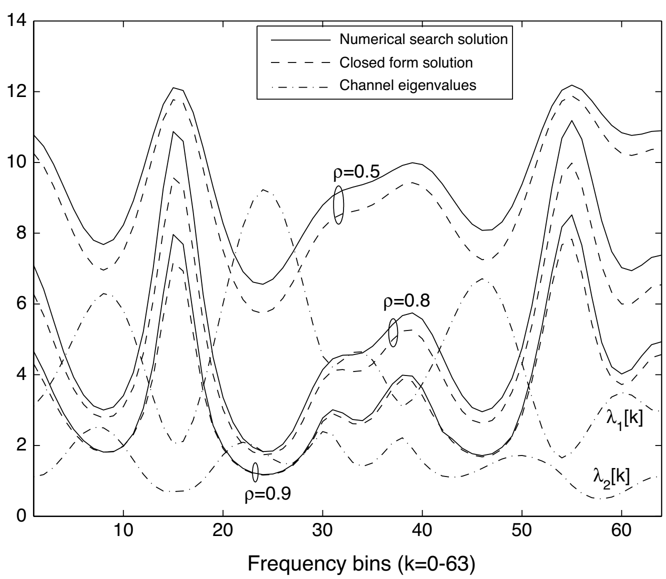

Fig. 3. Threshold distances $\{d_0^2[k]\}_{k=0}^{K-1}$ with $\overline{\mathrm{BER}}_0 = 10^{-4}$.

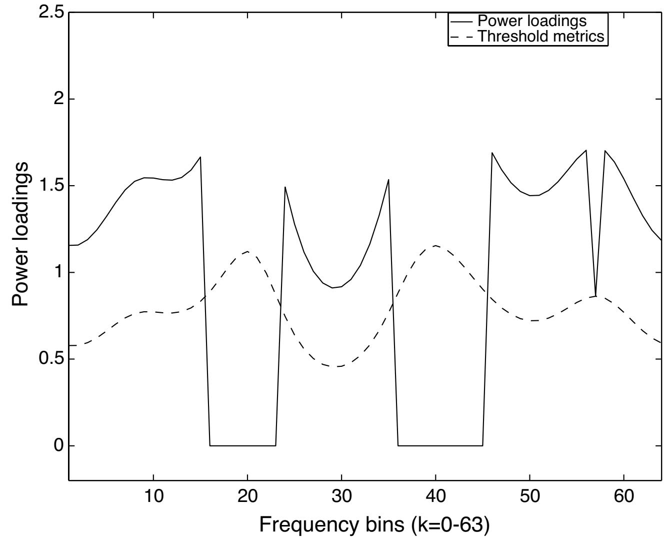

Fig. 4. Power loading snapshot for a certain channel realization.

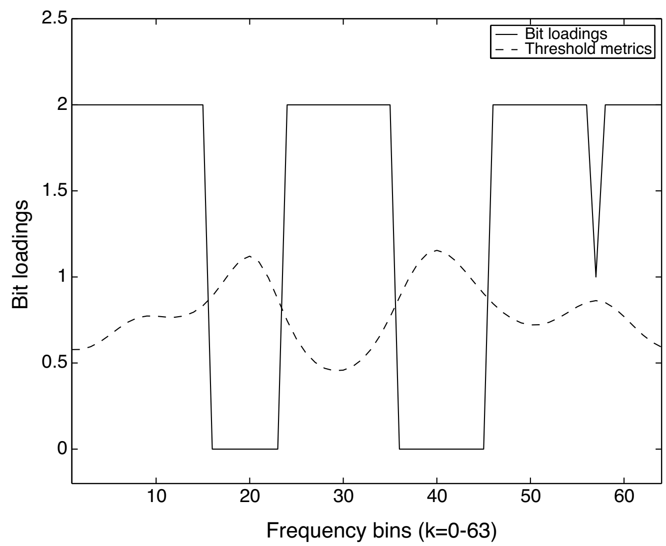

Fig. 5. Bit loading snapshot for a certain channel realization.

both rectangular and square QAM constellations in the adaptive modulation stage. Throughout this section, the average transmit-signal-to-noise ratio (SNR) across subcarriers is defined as $\mathrm{SNR} = P_{\mathrm{total}}T_s/(KN_0)$. The transmission rate (the loaded number of bits) is counted every two OFDM symbols as $\sum_{k=0}^{K-1} 2b[k]$.

Test Case 1—Comparison between exact and approximate solutions for $d_0^2[k]$: We simulate typical MIMO multipath channels with $N_t = 4$, $N_r = 2$, and $N_0 = 1$. For a certain channel realization, assuming 2-D beamforming on each subcarrier, we plot in Fig. 2 the thresholds $d_0^2[k]$ obtained via numerical search (19), and from the closed-form solution based on (22), with $\rho = 0.5$, 0.8, 0.9 and a target $\mathrm{BER} = 10^{-3}$. Fig. 3 is the counterpart of Fig. 2 but with target $\mathrm{BER} = 10^{-4}$. The non-negative eigenvalues $\lambda_1[k]$ and $\lambda_2[k]$ of the nominal channels are also plotted in dash-dotted lines for illustration purposes. We observe that the solutions of $d_0^2[k]$ obtained via these two different approaches are generally very close to each other, and the discrepancy decreases as the feedback quality ρ increases or as the target $\overline{\mathrm{BER}}_0$ increases. Notice that the suboptimal closed-form solution tends to underestimate $d_0^2[k]$. To deploy the suboptimal solution in practice, some SNR margins may be needed to ensure the target BER performance. Nevertheless, we will use the suboptimal closed-form solution for $d_0^2[k]$ in our ensuing numerical results.

Figs. 2 and 3 also reveal that on subchannels with large eigenvalues (indicating "good quality"), the resulting $d_0^2[k]$ is small; hence, large size constellations can be afforded on those subchannels.

Test Case 2—Power and bit loading with the Greedy algorithm: We set $N_t = 4$, $N_r = 2$, $\rho = 0.5$, $\mathrm{SNR} = 9$ dB, and $\overline{\mathrm{BER}}_0 = 10^{-4}$. For a certain channel realization, we plot the power and bit loading solutions obtained via the greedy algorithm in Figs. 4 and 5, respectively. For illustration purposes, we also plot the threshold metrics $d_0^2[k]$. We observe that when-

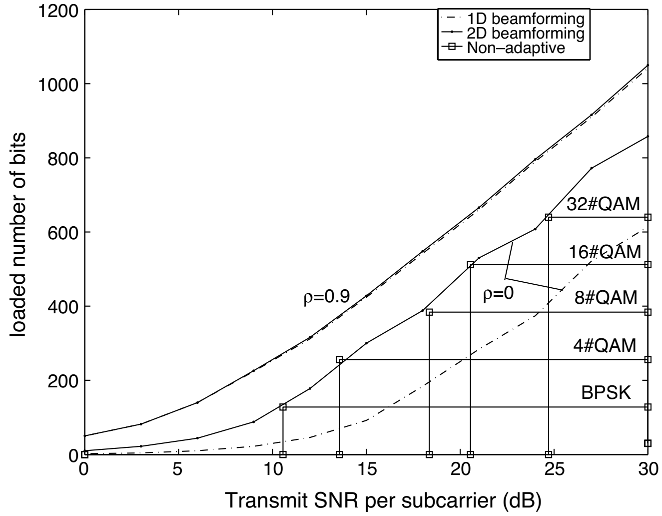

Fig. 6. Rate comparisons with $N_t = 2, N_r = 2$.

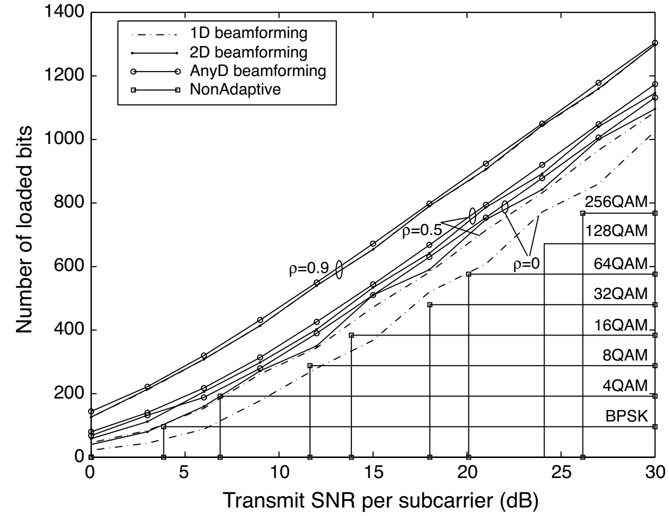

Fig. 8. Rate comparisons with $N_t = 4, N_r = 4$.

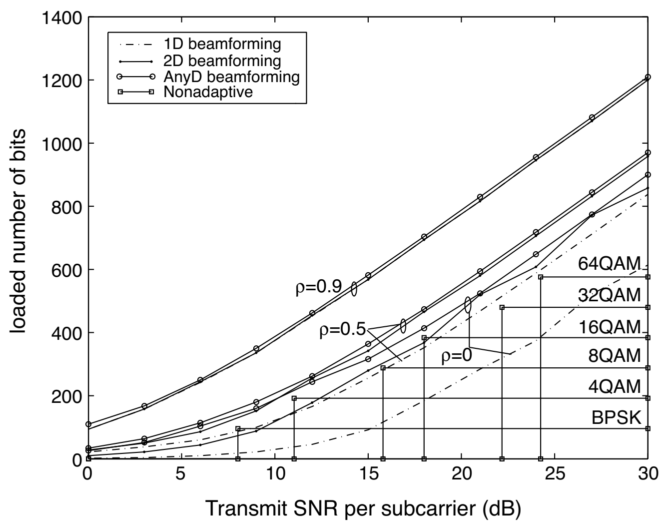

Fig. 7. Rate comparisons with $N_t = 4, N_r = 2$.

ever there is a change in the bit loading solution in Fig. 5 from one subcarrier to the next, there will be an abrupt change in the corresponding power loading in Fig. 4. Furthermore, for those subcarriers with the same number of bits, the power loaded by the greedy algorithm is proportional to the threshold metric. In addition, from the bit loading of the greedy algorithm in Fig. 5, we see that all subcarriers are loaded with an even number of bits (with the exception of one subcarrier at most), which is consistent with Proposition 2.

Test Case 3—Adaptive MIMO OFDM based on partial CSI: In addition to the adaptive MIMO-OFDM based on 1-D and 2-D coder-beamformers, we derived in Appendix C an adaptive transmitter that relies on higher dimensional beamformers on each OFDM subcarrier; we term it any-D beamformer here. With $\overline{\mathrm{BER}}_0 = 10^{-4}$, we compare nonadaptive transmission schemes (that use fixed constellations per OFDM subcarrier) and adaptive MIMO-OFDM schemes based on any-D, 2-D, and 1-D beamforming in Fig. 6 with $N_t = 2$, $N_r = 2$, in Fig. 7 with $N_t = 4$, $N_r = 2$, and in Fig. 8 with $N_t = 4$, $N_r = 4$. The Alamouti codes [1] are used when $N_t = 2$, and the rate 3/4 STBC code in [24] is used when $N_t = 4$. The transmission rates for adaptive MIMO-OFDM are averaged over 200 feedback realizations.

With $N_t = 2$ in Fig. 6, the any-D beamformer reduces to the 2-D coder-beamformer since there are at most two basis beams. With $N_t = 4$ in Figs. 7 and 8, we observe that the adaptive transmitter based on 2-D coder-beamformer achieves almost the same data rate as that based on any-D beamformer, for variable quality of the partial CSI (as ρ varies), and various size MIMO channels (as N_r varies). Thanks to its reduced complexity, 2-D beamforming is thus preferred over any-D beamforming. On the other hand, 1-D beamforming is considerably inferior to 2-D beamforming when low-quality CSI is present at the transmitter. However, as CSI quality increases (e.g., $\rho \geq 0.9$), the transmitter based on 1-D beamforming approaches the performance of that based on 2-D beamforming.

With $N_t = 2$ and $N_r = 2$ in Fig. 6, the adaptive MIMO-OFDM based on the 2-D coder-beamformer always outperforms nonadaptive alternatives. With $N_t = 4$ and $N_r = 2$ in Fig. 7, the nonadaptive transmitter could outperform the adaptive 2-D beamforming transmitter at the low SNR range, with extremely low feedback quality ($\rho = 0$). However, as the SNR increases or the feedback quality improves, the adaptive 2-D transmitter outperforms the nonadaptive transmitter considerably. As the number of receive antennas increases to $N_r = 4$ in Fig. 8, the adaptive 2-D beamforming transmitter is uniformly better than the nonadaptive transmitter, regardless of the feedback quality.

V. Concluding Summary

We designed MIMO-OFDM transmissions capable of adapting to partial (statistical) channel state information (CSI). Adaptation takes place in three (out of four) levels at the transmitter:

1) the power and (QAM) constellation size of the information symbols;
2) the power splitting among space-time coded information symbol substreams;

TABLE I
POWER REQUIRED TO LOAD THE iTH BIT ON THE kTH SUBCARRIER

i	1	2	3	4	5	6	7	8	...
$d_0^2[k]/g(i)$	$d_0^2[k]$	$2d_0^2[k]$	$6d_0^2[k]$	$10d_0^2[k]$	$26d_0^2[k]$	$42d_0^2[k]$	$106d_0^2[k]$	$170d_0^2[k]$	...
$c(k;i)$	$d_0^2[k]$	$d_0^2[k]$	$4d_0^2[k]$	$4d_0^2[k]$	$16d_0^2[k]$	$16d_0^2[k]$	$64d_0^2[k]$	$64d_0^2[k]$	...

3) the basis-beams of two- (or generally multi-) dimensional beamformers that are used (per time slot) to steer the transmission over the flat MIMO subchannels corresponding to each subcarrier.

These subchannels are created by the fourth (inner most) level that implements OFDM and enables design of our adaptive transmitter per subcarrier.

For a fixed transmit-power and a prescribed bit error rate performance per subcarrier, we maximize the transmission rate for the proposed transmitter structure over frequency-selective MIMO fading channels. The power and bits are judiciously allocated across space and subcarriers (frequency), based on partial CSI. Analogous to perfect-CSI-based DMT schemes, we established that loading in our partial-CSI-based MIMO OFDM design is controlled by a minimum distance parameter (which is analogous to the SNR-threshold used in DMT systems) that depends on the prescribed performance, the channel information, and its reliability, as those are partially (statistically) perceived by the transmitter. This analogy we have established offers two important implications: i) It unifies existing DMT metrics under the umbrella of partial CSI, and ii) it allows application of existing DMT loading algorithms from the wireline (perfect CSI) setup to the pragmatic wireless regime, where CSI is most often known only partially.

Regardless of the number of transmit antennas, our adaptive 2-D coder-beamformer should be preferred in practice over higher-dimensional alternatives since it enables desirable performance-rate-complexity tradeoffs.

APPENDIX A
PROOF OF PROPOSITION 1

Based on (28) and (12), we have

$$c(k,i) = 2^{2(j-1)} d_0^2[k] \quad \text{for } i = 2j-1, 2j, \text{ and } j = 1, 2, \ldots. \tag{33}$$

Table I lists the required power to load the ith bit on the kth subcarrier. From Table I and (33), we infer that

$$c(k, i+1) \geq c(k,i), \quad \forall i,\ k. \tag{34}$$

Although the greedy algorithm chooses always the 1-bit optimum [6], (34) reveals that all future additional bits will cost no less power. This is the key to establishing the overall optimality because no matter what the optimal final solution is, the bits on each subcarrier can be constructed in a bit-by-bit fashion, with every increment being most power efficient, as in the greedy algorithm. Hence, the greedy algorithm is overall optimal for our problem at hand. Lacking an inequality like (34), the optimality has not been formally established in [3] and [11].

APPENDIX B
PROOF OF PROPOSITION 2

An important observation from (33) is that $c(k, 2j-1) = c(k, 2j)$ holds true for any k and j. Suppose at some intermediate step of the greedy algorithm, the $(2j-1)$st bit on the kth subcarrier is the chosen bit to be loaded, which means that the associated cost $c(k, 2j-1)$ is the minimum out of all possible choices. Notice that $c(k, 2j) = c(k, 2j-1)$ has exactly the same cost, and therefore, after loading the $(2j-1)$st bit on the kth subcarrier, the next bit chosen by the optimal greedy algorithm *must* be the $(2j)$th bit on the same subcarrier, unless power insufficiency is declared. Therefore, the overall procedure effectively loads two bits at a time. As long as the power is adequate, the greedy algorithm will always load two bits in a row to each subcarrier. Let us denote the total number of bits as $R_{\text{square}} = 2\sum_{k=0}^{K-1} b_1[n;k]$ when using only square QAMs and $R_{\text{rect}} = 2\sum_{k=0}^{K-1} b_2[n;k]$ when allowing for rectangular QAMs as well. At most, on one subcarrier k', it holds that $b_2[n;k'] = b_1[n;k'] + 1$, which has probability 1/2, whereas for all other subcarriers, $b_2[n;k] = b_1[n;k]$. Hence, R_{square} is less than R_{rect} by at most one bit per space-time-coded OFDM block.

APPENDIX C
HIGHER THAN TWO-D BEAMFORMING

For practical deployment of our adaptive transmitter, we have advocated the 2-D coder-beamformer on each OFDM subcarrier. With $N_t > 2$, however, higher than 2-D coder beamformers have been developed in [12] and [32]. They are formed by concatenating higher dimensional orthogonal space-time block coding designs [24] with properly loaded space time multiplexers. Collecting more diversity through multiple basis beams, the optimal N_t-dimensional beamformer outperforms the 2-D coder beamformer from the minimum achievable $\overline{\text{BER}}$ point of view. Hence, with more than two basis beams, the threshold metric per subcarrier may improve, and the constellation size on each subcarrier may increase under the same performance constraint. However, the main disadvantage of N_t-dimensional beamforming is that the orthogonal STBC design in [24] loses rate when $N_t > 2$. The important issue in this context is how much one could lose in adaptive transmission rate by focusing only on the 2-D coder beamformer instead of allowing all possible choices of beamforming that can use up to N_t basis beams.

In the following, we use the notation n_tD to denote beamforming with n_t "strongest" basis beams. With $n_t \leq 2$, two symbols are transmitted over two time slots, as in (2). When $n_t = 3$, 4, the beamformer can be constructed based on the rate 3/4 orthogonal STBC, with three symbols transmitted over four time slots. When $5 \leq n_t \leq 8$, the beamformer can be constructed based on the rate 1/2 orthogonal STBC, with four symbols transmitted over eight time slots. Let us consider, for simplicity, a maximum of eight directions even when $N_t > 8$, i.e., $n_{t,\max} = \min(N_t, 8)$. If we take a super block with eight OFDM symbols as the adaptive modulation unit, then each super block allows for different n_tD beamformers on different subcarriers at each modulation adaptation step. Specifically, in one super block, one subcarrier could place four 2-D coder-beamformers, two 4-D beamformers, or one 8-D beamformer, depending on partial CSI. With constellation size $M[k]$, the corresponding transmission rate for the n_tD beamformer is $8 f_{n_t} \log_2(M[k])$ per subcarrier per super block, where $f_{n_t} = 1$ for $n_t = 1$, 2, $f_{n_t} = 3/4$ for $n_t = 3$, 4, and $f_{n_t} = 1/2$ for $n_t = 5$, 6, 7, 8. Furthermore, with power $P[k]$ on each subcarrier, the energy per information symbol is $d^2[k] = (1/f_{n_t}) g(b[k]) P[k]$. This includes (11) as a special case with $f_1 = f_2 = 1$.

As with 2-D beamforming, we wish to maximize the transmission rate of the MIMO-OFDM subject to the performance constraint on each subcarrier. Mimicking the steps followed in Section III, we first determine the distance threshold $d^2_{0,n_t}[k]$ on each subcarrier for the n_tD beamformer, where $1 \le n_t \le n_{t,\max}$. With the average BER expression for the n_tD beamformer [32, eq. (38)], we find $d^2_{0,n_t}[k]$ through 1-D numerical search. Hence, if the assigned constellation has $d^2[k] \ge d^2_{0,n_t}[k]$, adopting the n_tD beamformer will lead to the guaranteed BER performance thanks to the monotonicity we established in our Lemma.

Having specified $\{d^2_{0,n_t}[k]\}_{k=0}^{K-1}$ for each $n_t \in [1, 2, \ldots, n_{t,\max}]$, we can also modify our greedy algorithm to obtain the optimal power and bit loading across subcarriers. First, we define the effective number of bits $b_e := b f_{n_t}$ when 2^b-QAM is used together with n_tD beamforming. Second, we constrain the effective number of bits b_e to be integers in order to facilitate the problem solving procedure. To achieve this, noninteger QAMs are assumed to be temporarily available for any n_t (we will later on quantize them to the closet square or rectangular QAMs). This entails a certain approximation error, but our objective here is to quantify the difference between 2-D beamforming and any n_tD beamforming. The greedy algorithm can be applied as in Section III-B1 but with each step loading effectively one bit on a certain subcarrier. Specifically, we need to replace $c(k, b+1)$ in the original greedy algorithm with $c(k, b_e+1)$, where

$$c(k, b_e+1) = \min_{n_t} \left[\frac{f_{n_t} d^2_{0,n_t}[k]}{g\left(\frac{(b_e+1)}{f_{n_t}}\right)} \right] - \min_{n_t} \left[\frac{f_{n_t} d^2_{0,n_t}[k]}{g\left(\frac{b_e}{f_{n_t}}\right)} \right] \quad (35)$$

is the minimal power required to load one additional bit on top of b_e effective bits on the kth subcarrier, given that all possible n_tD beamformers can be arbitrarily chosen. Notice that the optimal beamforming, based on as many as $n_{t,\max}$ basis beams, includes 2-D beamforming as a special case with $n_{t,\max} = 2$. Numerical results demonstrate that the 2-D transmitter performs close to any higher dimensional one in most practical cases. However, the 2-D transmitter reduces the complexity considerably, which is the reason why we favor the 2-D coder beamformer in practice.

References

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

[2] M. S. Alouini and A. Goldsmith, "Adaptive modulation over Nakagami channels," *J. Wireless Commun.*, vol. 13, no. 1–2, pp. 119–143, May 2000.

[3] J. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, pp. 5–14, May 1990.

[4] P. Chow, J. Cioffi, and J. Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Trans. Commun.*, vol. 43, pp. 773–775, Feb.–Apr. 1995.

[5] J. M. Cioffi. A multicarrier primer [Online]. Available: http://www-isl.stanford.edu/cioffi/pdf/multicarrier.pdf

[6] T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Cambridge, MA: MIT Press, 2001.

[7] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[8] R. Gallager, *Information Theory and Reliable Communications*. New York: Wiley, 1968.

[9] A. Goldsmith and S. Chua, "Variable rate variable power M-QAM for fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 1218–1230, Oct. 1997.

[10] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[11] D. Hughes-Hartogs, "Ensemble Modem Structure for Imperfect Transmission Media," U.S. Patents Nos. 4 679 227, 4 731 816, 4 833 706, July 1987, Mar. 1988, May 1989.

[12] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Inform. Theory*, vol. 48, pp. 611–627, Mar. 2002.

[13] I. Kalet, "The multitone channel," *IEEE Trans. Commun.*, vol. 37, pp. 119–124, Feb. 1989.

[14] T. Keller and L. Hanzo, "Adaptive multicarrier modulation: a convenient framework for time-frequency processing in wireless communications," *Proc. IEEE*, vol. 88, pp. 611–640, May 2000.

[15] B. Krongold, K. Ramchandran, and D. Jones, "Computationally efficient optimal power allocation algorithms for multicarrier communication systems," *IEEE Trans. Commun.*, vol. 48, pp. 23–27, Jan. 2000.

[16] B. Lu, X. Wang, and Y. Li, "Iterative receivers for space-time block-coded OFDM systems in dispersive fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 213–225, Apr. 2002.

[17] A. Moustakas and S. Simon. (2002) Optimizing multi-transmitter single-receiver (MISO) antenna systems with partial channel knowledge. *Bell Lab Tech. Memo.* [Online]. Available: http://mars.bell-labs.com/cm/ms/what/mars/papers/physics/MISO.pdf

[18] 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. Select. Areas Commun.*, vol. 16, pp. 1423–1436, Oct. 1998.

[19] L. Piazzo, "Fast algorithm for power and bit allocation in OFDM systems," *Electron. Lett.*, vol. 35, no. 25, pp. 2173–2174, Dec 1999.

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

[21] Z. Song, K. Zhang, and Y. Guan, "Statistical adaptive modulation for QAM-OFDM systems," in *Proc. IEEE Globecom Conf.*, Taipei, Taiwan, R.O.C., Nov. 17–21, 2002.

[22] G. Stüber, *Principles of Mobile Communication*. Boston, MA: Kluwer, 1996.

[23] G. Taricco and E. Biglieri, "Exact pairwise error probability of space-time codes," *IEEE Trans. Inform. Theory*, vol. 48, pp. 510–513, Feb. 2002.

[24] 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.

[25] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Bell Labs. Technical Memo.*, 1995.

[26] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2632–2639, Sept. 2001.

[27] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications: Where Fourier meets Shannon," *IEEE Signal Processing Mag.*, vol. 47, pp. 29–48, May 2000.
[28] K. Wong, R. Cheng, K. Letaief, and R. Murch, "Adaptive antennas at the mobile and base stations in an OFDM/TDMA system," *IEEE Trans. Commun.*, vol. 49, pp. 195–206, Jan. 2001.
[29] S. Ye, R. Blum, and L. Cimini, "Adaptive modulation for variable rate OFDM systems with imperfect channel information," in *Proc. Veh. Technol. Conf.*, vol. 2, Birmingham, AL, May 6–9, 2002, pp. 767–771.
[30] D. Yoon and K. Cho, "General bit error probability of rectangular quadrature amplitude modulation," *Electron. Lett.*, vol. 38, no. 3, pp. 131–133, Jan. 2002.
[31] S. Zhou and G. B. Giannakis, "Adaptive modulation for multiantenna transmissions with channel mean feedback," *IEEE Trans. Wireless Commun.*, to be published.
[32] ——, "Optimal transmitter eigen-beamforming and space-time block coding based on channel mean feedback," *IEEE Trans. Signal Processing*, vol. 50, pp. 2599–2613, Oct. 2002.
[33] ——, "Adaptive modulation for multi-antenna transmissions with channel mean feedback," in *Proc. Int. Conf. Commun.*, May 2003.

Pengfei Xia (S'03) received the B.S. and M.S. degrees in electrical engineering from the University of Science and Technology of China (USTC), Hefei, China, in 1997 and 2000 respectively. He is now pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis.

His research interests lie in the areas of signal processing and communications, including multiple-input multiple-output wireless communications, transceiver designs, adaptive modulation, multicarrier transmissions, space-time codes, and iterative decoding techniques.

Shengli Zhou (M'03) received the B.S. degree in 1995 and the M.Sc. degree in 1998 from the University of Science and Technology of China (USTC), Hefei, China, all in electrical engineering and information science. He received the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, in 2002.

He joined the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, as an Assistant Professor in 2003. His research interests lie in the areas of communications and signal processing, including channel estimation and equalization, multiuser and multicarrier communications, space-time coding, adaptive modulation, and cross-layer designs.

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

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, 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 150 journal papers, 300 conference papers, and two edited books. Current research topics focus on transmitter and receiver diversity techniques for single- and multiuser fading communication channels, precoding and space-time coding for block transmissions, and multicarrier and ultra wideband wireless communication systems. He is a frequent consultant for the telecommunications industry.

Dr. Giannakis is the (co-) recipient of four best paper awards from the IEEE Signal Processing (SP) Society (in 1992, 1998, 2000, and 2001). He also received the Society's Technical Achievement Award in 2000. He co-organized three IEEE-SP Workshops and guest (co-) edited four special issues. He has 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, and as chair of the SP for Communications Technical Committee. He is a member of the Editorial Board for the PROCEEDINGS OF THE IEEE and the steering committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He is a member of the IEEE Fellows Election Committee and the IEEE-SP Society's Board of Governors.