

Precoder Design for Multiuser MIMO ISI Channels Based on Iterative LMMSE Detection

Xiaojun Yuan, *Member, IEEE*, Chongbin Xu, Li Ping, *Senior Member, IEEE*, and Xiaokang Lin

Abstract—Precoding has been widely investigated for multiple-input multiple-output (MIMO) and inter-symbol interference (ISI) channels. Most related work so far has been on uncoded systems and the impact of forward-error-correction (FEC) codes has not been well studied. In this paper, we present an optimized precoding technique that takes into account iterative joint LMMSE detection and FEC decoding. The proposed scheme has some interesting features. First, the design problem is convex. Second, the proposed scheme achieves water-filling gain when channel state information at the transmitter (CSIT) is available and diversity gain when CSIT is not available, as well as multi-user gain in multi-user environments. Third, the water-filling gain is achieved using a single code (instead of multiple codes as in conventional approaches), which greatly simplifies the design and implementation problems. Both simulation and evolution analyses are provided to show that significant performance improvement can be achieved with the proposed scheme.

Index Terms—Iterative linear minimum-mean-square-error (LMMSE) detection, multiple-input multiple-output (MIMO) inter-symbol interference (ISI) multiple-access channel, precoder.

I. INTRODUCTION

CONSIDER communication in multiple-input multiple-output (MIMO) and multipath channels. In the case of perfect channel state information at the transmitter (CSIT), a well-known approach is to apply singular value decomposition (SVD) together with orthogonal frequency division multiplexing (OFDM) to transform the channel into a set of parallel sub-channels. Then water-filling [1] can be used to achieve channel capacity, which involves multiple forward-error-correction (FEC) codes with different rates. Designing, storing, and operating these codes are complicated issues. Also, modern FEC codes, such as turbo and low-density parity-check (LDPC) codes [2], [3], may involve very long block length, for which latency is a serious problem. The use of multiple codes in parallel may worsen the problem.

Linear precoding is an alternative approach that involves a linear precoder at the transmitter side and an equalizer at the receiver side. The design objective is to maximize the signal-to-interference and noise ratio (SINR) at the output of the equalizer

(usually designed based on the linear minimum-mean-square-error (LMMSE) principle) [4]–[6], although other measures can be used instead [5]. It has been shown in [7], [8] that performance can be improved by coupling linear precoding with decision feedback detection. Most existing work [4]–[8] on linear precoding so far has been limited to systems without FEC codes.

For channels without CSIT, diversity techniques can be used to improve performance. The related space-time [9]–[12] or space-time-frequency codes [13] can be regarded as generalized nonlinear precoding techniques.

Overall, various types of treatments have been used for different scenario, e.g., SVD/OFDM/water-filling and linear precoding for channels with CSIT, and/or space-time coding for channels without CSIT.

In this paper, we develop a unified precoding scheme that is universally applicable to systems with and without CSIT. FEC coding is an integrated part in the scheme. The key is to utilize the feedback from a soft-output FEC decoder that is refined in an iterative process. The design criterion is to maximize the SINR at the end of the iterative process. Intuitively, the rationale behind this criterion is that the decoder feedback can be used to suppress interference and so SINR evolves as the quality of the decoder feedbacks improves. This process is analyzed using an SINR-variance evolution technique [14], based on which we will show that the precoder design problem is convex.

The proposed scheme involves only a single FEC code. Interestingly, it can still achieve comparable performance to the water-filling strategy. It can be shown that the proposed scheme can achieve near-capacity performance when the precoder and FEC code are jointly optimized. (This issue is discussed in a separate paper [15].)

We further extend the proposed precoding technique to multiuser environments. We show that in the MIMO case, noticeable multiuser diversity gain can be achieved in addition to the precoding gain. Again, the advantage of the proposed scheme is that only one FEC code is used by each user to explore the benefit in multiuser environments.

The proposed scheme can be implemented at a low cost of $O(\log M)$ per symbol based on the fast Hadamard transform (FHT) technique [16], with M being the length of the FHT used. Both analysis and simulation results are provided to demonstrate that the proposed scheme can bring about significant performance gains in various channels environments.

The remainder of this paper is organized as follows. We focus on single-user systems in Sections II–IV. Specifically, Section II outlines the channel model, channel diagonalization, and the water-filling principle. Section III describes the precoder design based on iterative LMMSE detection. The application of the proposed precoding technique in OFDM is described in Section IV. The extension to multiuser en-

Manuscript received September 01, 2008; revised August 28, 2009. Current version published January 13, 2010. This work was supported by the Research Grant Council of the Hong Kong SAR, China, under Project CityU 117007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Christoph Mechlénbräuer.

X. Yuan and L. Ping are with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong (e-mail: xjyuan@cityu.edu.hk, eeliping@cityu.edu.hk).

C. Xu and X. Lin are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: xcb05@mails.tsinghua.edu.cn, linxk@sz.tsinghua.edu.cn).

Digital Object Identifier 10.1109/JSTSP.2009.2035859

vironments is provided in Section V, and finally Section VI concludes this paper.

II. PRELIMINARIES

A. Channel Model

Let j be the time index. A single-user quasi-static MIMO inter-symbol interference (ISI) channel can be modeled as

$$\mathbf{r}(j) = \sum_{l=0}^{L-1} \mathbf{H}(l)\mathbf{y}(j-l) + \boldsymbol{\eta}(j) \quad (1)$$

where $\mathbf{r}(j)$ is the received signal vector from N_r receive antennas at time j , $\mathbf{y}(j)$ the transmitted signal vector over the N_t transmit antennas at time j , $\boldsymbol{\eta}(j)$ an $N_r \times 1$ vector consisting of additive-white-Gaussian-noise (AWGN) samples with variance σ^2 per dimension, $\mathbf{H}(l)$ an $N_r \times N_t$ matrix representing the MIMO channel with a delay of l taps, and L the channel memory length.

Assume that the cyclic prefix (CP) technique is applied [14], [17] to $\mathbf{y} = [\mathbf{y}(0)^T, \mathbf{y}(1)^T, \mathbf{y}(2)^T, \dots, \mathbf{y}(J-1)^T]^T$. (Note that \mathbf{y} consists of JN_t symbols.) We can then rewrite (1) with J times of channel use into a compact form as

$$\mathbf{r} = \mathbf{H}\mathbf{y} + \boldsymbol{\eta} \quad (2)$$

where \mathbf{r} and $\boldsymbol{\eta}$ are defined similarly as \mathbf{y} , and \mathbf{H} is an $N_r \times N_t$ block-circulant matrix

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}(0) & \cdots & \mathbf{H}(2) & \mathbf{H}(1) \\ \mathbf{H}(1) & \mathbf{H}(0) & & \mathbf{H}(2) \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{H}(J-1) & \cdots & \mathbf{H}(1) & \mathbf{H}(0) \end{bmatrix}_{JN_r \times JN_t} \quad (3)$$

Channel matrix \mathbf{H} is assumed to remain unchanged during the transmission of every T segments of \mathbf{y} .

Define a block-wise discrete Fourier transform (DFT) matrix

$$\mathbf{F}_N = \begin{bmatrix} \mathbf{I}_N & \mathbf{I}_N & \cdots & \mathbf{I}_N \\ \mathbf{I}_N & e^{-i2\pi\frac{1}{J}}\mathbf{I}_N & \ddots & e^{-i2\pi\frac{(J-1)}{J}}\mathbf{I}_N \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{I}_N & e^{-i2\pi\frac{J-1}{J}}\mathbf{I}_N & \cdots & e^{-i2\pi\frac{(J-1)(J-1)}{J}}\mathbf{I}_N \end{bmatrix}_{JN \times JN} \quad (4)$$

where $i \triangleq \sqrt{-1}$ and \mathbf{I}_N is an $N \times N$ identity matrix. The block-wise circulant form of \mathbf{H} allows the following decomposition [18]

$$\mathbf{H} = \mathbf{F}_{N_r}^H \mathbf{G} \mathbf{F}_{N_t} \quad (5a)$$

where \mathbf{F}_{N_r} and \mathbf{F}_{N_t} have the same form as \mathbf{F} in (4), except their dimensions are $JN_r \times JN_r$ and $JN_t \times JN_t$, respectively. Matrix \mathbf{G} in (5a) is a block diagonal matrix defined as

$$\mathbf{G} = \text{diag}\{\mathbf{G}(0), \mathbf{G}(1), \dots, \mathbf{G}(J-1)\} \quad (5b)$$

with the j th diagonal block given by

$$\mathbf{G}(j) = \sum_{l=0}^{J-1} \mathbf{H}(l) \exp(-i2\pi jl/J). \quad (5c)$$

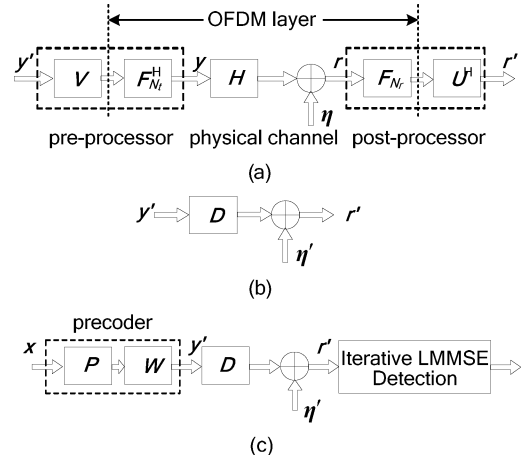


Fig. 1. (a) Transmission over a MIMO ISI channel. (b) The equivalent parallel subchannel model of the system in (a). (c) The precoding scheme based on the equivalent model in (b).

As special cases, (2) reduces to a single-path MIMO channel when $L = 1$, and to a single-input single-output (SISO) ISI channel when $N_r = N_t = 1$.

B. Channel Diagonalization

The SVD technique can be applied to diagonalize the system in (2) [19]. Let the SVD of \mathbf{G} in (5b) be

$$\mathbf{G} = \mathbf{U} \mathbf{D} \mathbf{V}^H. \quad (6)$$

Then (5a) suggests the following efficient decomposition of \mathbf{H} :

$$\mathbf{H} = \mathbf{F}_{N_r}^H \mathbf{U} \mathbf{D} \mathbf{V}^H \mathbf{F}_{N_t}. \quad (7)$$

With perfect CSIT, we can diagonalize a MIMO ISI channel based on (7), as shown in Fig. 1(a). It involves a pre-processor and a post-processor defined as follows:

$$\text{Pre-processor: } \mathbf{y}' = \mathbf{V}^H \mathbf{F}_{N_t} \mathbf{y} \quad (8a)$$

$$\text{Post-processor: } \mathbf{r}' = \mathbf{U}^H \mathbf{F}_{N_r} \mathbf{r}. \quad (8b)$$

In this way, (2) is transformed into a set of J' parallel sub-channels represented by

$$\mathbf{r}' = \mathbf{D} \mathbf{y}' + \boldsymbol{\eta}' \quad (8c)$$

where $\boldsymbol{\eta}' = \mathbf{U}^H \mathbf{F}_{N_r} \boldsymbol{\eta}$ and \mathbf{D} is a $J' \times J'$ diagonal matrix¹ with

$$J' \equiv J \cdot \min(N_r, N_t). \quad (8d)$$

The equivalent system based on (8c) is illustrated in Fig. 1(b).

C. OFDM Systems

We can combine $\mathbf{F}_{N_t}^H$ at the transmitter, the channel \mathbf{H} , and \mathbf{F}_{N_r} at the receiver to form an underlying OFDM layer, as illustrated in Fig. 1(a). Then \mathbf{V} and \mathbf{U} represent additional operations to diagonalize the MIMO channel at each subcarrier. The related complexity, besides the standard fast Fourier transform

¹The size of \mathbf{D} is $JN_r \times JN_t$ after SVD. We can reduce its size to $J' \times J'$ by removing all-zero rows or columns. The size of \mathbf{r}' , \mathbf{y}' , and $\boldsymbol{\eta}'$ are adjusted accordingly. With some abuse of notation, we still denote the reduced variables correspondingly by \mathbf{D} , \mathbf{r}' , \mathbf{y}' , and $\boldsymbol{\eta}'$.

(FFT) operations for OFDM, is $O(N_r \cdot N_t \cdot \min(N_r, N_t))$ per sub-carrier [19]. With CSIT, the water-filling technique can be applied to MIMO OFDM. (See, e.g., LTE Advance [20].)

Due to the equivalence illustrated in Fig. 1(a), the discussion in this paper applies to both single-carrier and OFDM systems. We will return to this in Section IV.

III. PRECODING BASED ON ITERATIVE LMMSE DETECTION

A. Precoder Structure

Let \mathbf{x} be a coded vector generated by a generic encoder (consisting of an FEC encoder followed by random interleaving). Without loss of generality, the average power per entry of \mathbf{x} is normalized to 1. We will treat \mathbf{x} as a sequence of independent and identically distributed (i.i.d.) symbols, which is approximately ensured by random interleaving [21]–[24].

Assume that the channel has been diagonalized as in (8) and Fig. 1(a). We adopt the precoding scheme illustrated in Fig. 1(c) defined by linear transform

$$\mathbf{y}' = \mathbf{W}\mathbf{P}\mathbf{x} \quad (9)$$

where \mathbf{y}' is the transmitted signal vector over the parallel channel in (8c) and

$$\mathbf{W} = \text{diag}\{W(0,0), W(1,1), \dots, W(J'-1, J'-1)\}$$

that is used for power allocation among sub-channels. We will discuss the optimized selection of \mathbf{W} later in Section III-F.

The matrix \mathbf{P} in (9) is the rotation matrix introduced in ([5], (15)), which can be either a DFT matrix or a Hadamard matrix. In either case, \mathbf{P} has the following two properties.

- 1) \mathbf{P} is unitary.
- 2) The entries of \mathbf{P} have a constant modulus of $1/\sqrt{J'}$.

These two properties are useful to our discussion below.

Incidentally, using a Hadamard matrix as \mathbf{P} is more cost-effective as then $\mathbf{P}\mathbf{x}$ involves additions only.

B. Detection Principles

Return to the diagonalized channel in (8c). The use of the precoder in (9) will reintroduce inter-symbol interference. We now consider applying iterative LMMSE detection [14] to suppress this interference. A key in the discussion below is to take into account the feedback information. This brings about significant performance improvement, as will be seen later.

Combining (8c) and (9), we have

$$\mathbf{r}' = \mathbf{A}\mathbf{x} + \boldsymbol{\eta}' \quad (10a)$$

where

$$\mathbf{A} = \mathbf{D}\mathbf{W}\mathbf{P} \quad (10b)$$

and $\boldsymbol{\eta}'$ is a vector of AWGN samples with variance σ^2 . The optimal estimation for \mathbf{x} usually involves prohibitively high complexity. We will take a suboptimal alternative shown in Fig. 2. The receiver consists of two local operators, namely, the elementary signal estimator (ESE) and the decoder (DEC). The coding constraint is ignored in the ESE and the impact of the channel matrix \mathbf{A} is ignored in the DEC, which implies reduced

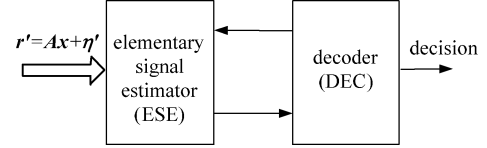


Fig. 2. Structure of an iterative receiver.

complexity at the cost of certain performance loss. To compensate for the performance loss, an iterative procedure is applied to the two local operators, as explained in Section III-E.

For the ESE, even after ignoring the coding constraint, the optimal detection of \mathbf{x} is still a very difficult task when \mathbf{x} is discrete (as in most practical systems). To overcome the difficulty, we take a two-step approach. In the first step, we treat \mathbf{x} as approximately Gaussian distributed. We assume that the *a priori* mean (denoted by $\mathbb{E}(\mathbf{x})$) and *a priori* auto-covariance of \mathbf{x} are available. To simplify the problem, we further assume that the *a priori* autocorrelation of \mathbf{x} can be written as $v\mathbf{I}$. In practice, the *a priori* information can be initialized to $\mathbb{E}(\mathbf{x}) = \mathbf{0}$ and $v = 1$, and updated using the feedbacks from the DEC during the iterative process. (We will return to this later in Section III-E.) Then the standard LMMSE estimation of \mathbf{x} is given by [25]

$$\hat{\mathbf{x}} = \mathbb{E}[\mathbf{x}] + v\mathbf{A}^H\mathbf{R}^{-1}(\mathbf{r}' - \mathbf{A}\mathbb{E}[\mathbf{x}]) \quad (11a)$$

where $\mathbf{R} \equiv \text{Cov}(\mathbf{r}', \mathbf{r}') = v\mathbf{A}\mathbf{A}^H + \sigma^2\mathbf{I}$. In the second step, we write the i th entry of $\hat{\mathbf{x}}$ in the following form [21]

$$\hat{x}(i) = \varphi(i)x(i) + \xi(i) \quad (11b)$$

In (11b)

$$\varphi(i) = v\Omega(i, i) \quad (12a)$$

$$\xi(i) = \mathbb{E}(x(i)) + v\mathbf{a}^H(i)\mathbf{R}^{-1}(\mathbf{r}' - \mathbf{A}\mathbb{E}[\mathbf{x}]) + v\Omega(i, i)x(i) \quad (12b)$$

where $\mathbf{a}(i)$ is the i th row of \mathbf{A} , and $\Omega(i, i)$ is the i th diagonal entry in the following matrix:

$$\boldsymbol{\Omega} \equiv (\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A})_{\text{diag}}. \quad (12c)$$

Here, $\xi(i)$ represents an unknown interference-and-noise term that is independent of $x(i)$. Following [21], we treat $\xi(i)$ as an AWGN sample. Its variance is given by [[26], (18)]

$$\text{Var}(\xi(i)) = \varphi(i)(1 - \varphi(i))v. \quad (12d)$$

Then we can detect $x(i)$ based on (11b) in a symbol-by-symbol manner using the true distribution of $x(i)$.

The rationale behind the above two-step approach is that $\hat{x}(i)$ is statistically close to $x(i)$ after the LMME operation in (11a). Thus, the symbol-by-symbol detection in (11b) is relatively reliable.

C. Complexity

We now briefly discuss the complexity of the above approach. The cost involved in the two-step approach discussed in Section III-B is mainly for the matrix inversion \mathbf{R}^{-1} . Using a Hadamard matrix for \mathbf{P} is a cost effective option. To see

this, substitute $\mathbf{A} = \mathbf{DWP}$ into (12) and note that \mathbf{P} is unitary (property (i) in Section III-A). Then

$$\mathbf{R} = v\mathbf{DWW}^H\mathbf{D}^H + \sigma^2\mathbf{I}. \quad (13a)$$

Clearly, \mathbf{R} is diagonal and so \mathbf{R}^{-1} can be computed easily. Furthermore, note that

$$\begin{aligned} \mathbf{\Omega} &= ((\mathbf{DWP})^H(v(\mathbf{DWP})(\mathbf{DWP})^H \\ &\quad + \sigma^2\mathbf{I})^{-1}(\mathbf{DWP}))_{\text{diag}} \\ &= (\mathbf{P}^H\mathbf{\Sigma}\mathbf{P})_{\text{diag}} \end{aligned} \quad (13b)$$

where

$$\mathbf{\Sigma} = (\mathbf{DW})^H(v(\mathbf{DW})(\mathbf{DW})^H + \sigma^2\mathbf{I})^{-1}(\mathbf{DW}). \quad (13c)$$

From property (ii) in Section III-A, $|P(j, n)|^2 = 1/J', \forall j, n$. We can thus write the j th diagonal element of $(\mathbf{P}^H\mathbf{\Sigma}\mathbf{P})_{\text{diag}}$ as

$$\sum_{n=0}^{J'-1} |P(j, n)|^2 \Sigma(n, n) = \frac{1}{J'} \sum_{n=0}^{J'-1} \Sigma(n, n) \quad (14)$$

where $\Sigma(n, n)$ is the n th diagonal entry of $\mathbf{\Sigma}$. Since (14) is invariant to j , we can rewrite (13b) as

$$\mathbf{\Omega} = \omega\mathbf{I} \quad (15a)$$

where

$$\begin{aligned} \omega &= \frac{1}{J'} \sum_{j=0}^{J'-1} \Sigma(j, j) \\ &= \frac{1}{J'} \sum_{j=0}^{J'-1} \left(\frac{|W(j, j)|^2 |D(j, j)|^2}{v|W(j, j)|^2 |D(j, j)|^2 + \sigma^2} \right). \end{aligned} \quad (15b)$$

This provides an efficient way to evaluate $\mathbf{\Omega}$ in (12c).

All the other operations in (11) and (12) involving $\mathbf{A} = \mathbf{DWP}$ can be efficiently implemented using FHT. The complexity involved in (11) and (12) is thus dominated by FHT at $O(J' \log J')$ per block of \mathbf{x} [14] or $O(\log J')$ per symbol, which is modest for practical use.

D. SINR-Variance Transfer Function

The performance of the ESE can be characterized using the SINR in its output. Let $\rho(j)$ be the SINR for $x(j)$. Based on (12a) and (12d)

$$\rho(j) = \frac{|\varphi(i)|^2}{\text{Var}(\xi(i))} = \frac{\Omega(j, j)}{1 - v\Omega(j, j)}. \quad (16a)$$

From (15a)

$$\rho(j) = \rho = \frac{\omega}{1 - v\omega}, \quad \forall j. \quad (16b)$$

Substituting (15b) into (16b), we have²

$$\begin{aligned} \rho &= \phi(v) \\ &\equiv \left(\frac{1}{J'} \sum_{j=0}^{J'-1} \left(v^{-1} + \frac{|W(j, j)|^2 |D(j, j)|^2}{\sigma^2} \right)^{-1} \right)^{-1} \\ &\quad - v^{-1} \end{aligned} \quad (16c)$$

where $|W(j, j)|^2$ is the power allocated to sub-channel j . Thus, the behavior of the ESE can be characterized by a transfer function $\rho = \phi(v)$.

Note that (15a) and so (16b) are the consequences of the two properties of \mathbf{P} mentioned in Section III-A.

E. Iterative Process

Recall that the *a priori* information is assumed in Section III-B. Reliable *a priori* information is crucial to the proposed scheme. Now we consider the following iterative process to refine the *a priori* information. We forward the ESE output (i.e., $\hat{\mathbf{x}}$) to the DEC. After decoding, the DEC produces updated $\mathbf{E}(\mathbf{x})$ and v . (The details on computing $\mathbf{E}(\mathbf{x})$ and v can be found in [14].) We then re-perform the ESE operation. This process continues iteratively.

We can view $\mathbf{E}(\mathbf{x})$ as an estimate of \mathbf{x} and v as the reliability of $\mathbf{E}(\mathbf{x})$. Also, ρ in (16c) can be viewed as a reliability measure for the input to the DEC. Thus, we can characterize the DEC operation by a transfer function $v = \psi(\rho)$. We obtain this function as follows. We approximately treat $\hat{\mathbf{x}}$ as the observations of an AWGN channel following [21], [22]. This can be justified by approximating $\xi(i)$ in (11b) as an AWGN sample and noting that the SINR for (11b) is a constant according to (16b). We then obtain $\psi(\rho)$ by simulating the DEC in an AWGN channel with $\text{SNR} = \rho$.

With $\phi(v)$ and $\psi(\rho)$ available, we can predict the performance of the iterative process using a recursion of ρ and v . Let the superscript “ (q) ” be the iteration number. We initialize $v^{(0)} = 1$, representing the situation of no feedback from the DEC. The SINR at the output of the ESE can be computed using the following recursion:

$$\rho^{(q)} = \phi\left(v^{(q-1)}\right) \quad \text{and} \quad v^{(q)} = \psi\left(\rho^{(q)}\right) \quad \text{for } q = 1, 2, \dots, Q \quad (17)$$

where Q is assumed to be sufficiently large to ensure convergence. The bit error rate (BER) and frame error rate (FER) of the system can be predicted by two functions $\text{BER}(v^{(Q)})$ and

²To gain insight into (16c), define

$$\text{MMSE} = J^{-1} \min_{\hat{\mathbf{x}}} \mathbf{E}((\mathbf{x} - \hat{\mathbf{x}})^H (\mathbf{x} - \hat{\mathbf{x}})).$$

Under the assumption that \mathbf{x} is Gaussian, it can be shown that [15]

$$\text{MMSE} = v - v^2\omega.$$

Substituting this into (16c), we have

$$\rho = \frac{1}{\text{MMSE}} - \frac{1}{v}.$$

Thus, ρ is the difference between the reciprocals of the *a posteriori* and *a priori* variances (as MMSE is the *a posteriori* variance after observing \mathbf{r}^j).

$\text{FER}(v^{(Q)})$ that can be obtained by simulating the DEC in an AWGN channel [14], [23].

Note that, in the above, we have assumed that the inputs to the ESE and DEC are i.i.d., based on which the autocorrelation of \mathbf{x} is expressed as $v\mathbf{I}$. (See the discussion above (11a).) This is a standard assumption following the turbo principle [2], [3]. To ensure this assumption, we can employ random interleaving and sufficiently long coding length. This implies that \mathbf{y} defined in (2) is usually a segment in a long coded frame.

F. Optimization for \mathbf{W}

We now consider the construction of \mathbf{W} based on the evolution technique in Sections III-D and III-E. Assume that the target performance of the system is given in the form of BER or FER. We can then compute the required v (denoted by v^{target}) according to $\text{BER}(v)$ or $\text{FER}(v)$ mentioned in Section III-E. We want to minimize the total transmission power to achieve this target after the recursion in (17). The problem can be formulated as

$$\min_{\mathbf{W}} \sum_{j=0}^{J'-1} |W(j, j)|^2 \quad \text{subject to } v^{(Q)} \leq v^{\text{target}} \quad (18)$$

where $v^{(Q)}$ is obtained by the recursion in (17). In practice, ϕ and ψ are both monotonic functions and so

$$\phi(v) > \psi^{-1}(v). \quad (19)$$

Thus, the problem in (18) can be expressed equivalently as

$$\min_{\mathbf{W}} \sum_{j=0}^{J'-1} |W(j, j)|^2 \quad (20a)$$

$$\text{subject to } \phi(v) > \psi^{-1}(v), \quad \text{for } 1 \geq v > v^{\text{target}}. \quad (20b)$$

The following lemma is helpful in solving problem (20). A similar result for SISO channels is derived in [27]. For completeness, a proof is included in Appendix A.

Lemma 1: ϕ in (16c) is concave in $\{|W(j, j)|^2\}$.

From Lemma 1, (20) is convex, which can be solved as follows. Note that the constraint in (20b) is over a continuous interval of v . Following the treatment in some similar problems (e.g., in the design of Raptor codes [28]), we can discretize v in $(v^{\text{target}}, 1]$ with a small interval Δv and let the constraint in (20b) hold on the discretized values of v . Then (20) can be solved efficiently using standard convex optimization tools [29]. Since ϕ defined in (16c) is continuous in v , the above technique has sufficient precision provided that Δv is sufficiently small.

G. Water-Filling

Here is a simpler alternative. When $\mathbf{P}\mathbf{x}$ is i.i.d., the optimal choice of \mathbf{W} achieving the channel capacity under sum power constraint is the water-filling solution [1]. We borrow this principle for a practically coded system and construct \mathbf{W} by setting its diagonal entries to the water-filling solution scaled by a proper constant. (The scaling factor is necessary as the water-filling solution specifies the theoretically minimum power.) This

is a rather heuristic method under the assumption that the optimized power allocation for a practically coded system differs from the water-filling solution only by a scaling factor, which is not necessarily true. However, this water-filling technique is simple and easy to implement. It also provides reasonably good (though suboptimal) performance. (See Fig. 6.)

H. Examples

We now present some examples to demonstrate the efficiency of the proposed precoding technique. We assume quasi-static Rayleigh-fading $N_r \times N_t$ MIMO ISI channels with N_t transmit antennas, N_r receive antennas, and L delay taps. The fading coefficients $\{h_{m,n}(l)\}$ are i.i.d., where $h_{m,n}(l)$ is the (m, n) th entry of $\mathbf{H}(l)$ in (3). The total gain of each user over the L paths between each transmit-receive antenna pair is normalized to 1, i.e.,

$$\mathbb{E} \left[\sum_{l=0}^{L-1} |h_{m,n}(l)|^2 \right] = 1, \quad \forall m, n.$$

The channel remains constant during the transmission of each coded frame (see the definition in Section II-A), but changes independently from frame to frame. The FEC code is a rate 1/2 convolutional code with generator polynomials $(7, 5)_8$ followed by random interleaving and QPSK modulation. An outage probability $\varepsilon = 0.01$ is allowed.³ Each point on the simulation curves is obtained by averaging over at least 10 000 channel realizations. The average transmission power is measured in bit energy to noise density ratio, denoted by E_b/N_0 . The above simulation settings are assumed throughout this paper unless explicitly specified.

We first show the BER performance for a single-path 4×4 MIMO channel (with $L = 1$) in Fig. 3. The system with flat precoding (in which \mathbf{W} is set to a scaled identity matrix) is included as a reference.

We can see from Fig. 3 that significant gains can be achieved by optimization on \mathbf{W} . The simulated performance is quite close to the evolution prediction when frame length is set to 32 768 information bits. The difference between the simulation results and evolution predictions becomes more noticeable for a small frame length of 4096 information bits. This is expected as the basic assumption of the evolution analysis states that the ESE and DEC inputs can be approximately modeled as i.i.d. sequences. This assumption holds only when the frame length is sufficiently large [14]. (Recall from Section III-E that \mathbf{y} defined in (2) is a segment in a long coded frame.)

It is also interesting to compare the results for different targeted BER values in solving (20). In Fig. 3, there are cross points between the two sets of results with target BER = 10^{-3} and 10^{-5} . Although setting a lower target BER can achieve slightly better asymptotical performance in the high E_b/N_0 region, the related gain is limited and is usually not achievable in practical systems with a small frame length. For this reason, we always set the target BER of the optimized precoder to 10^{-3} and frame length to 4096 information bits in our following simulations.

³If the required transmission power to achieve a target BER/FER exceeds a predetermined threshold, we set this frame in outage, i.e., no transmission will happen during this frame period. The threshold is adjustable so as to meet the outage requirement (that can be obtained by a pre-simulation). The errors due to outage are not counted in the BER performance.

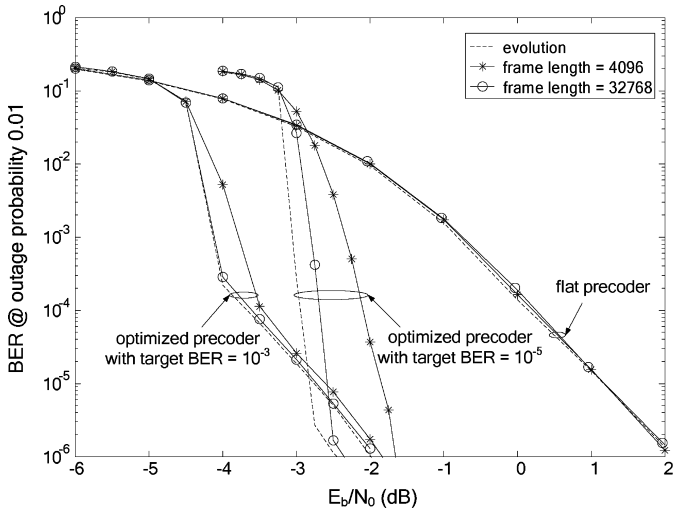


Fig. 3. Performance of the optimized precoder with various target BERs and frame lengths in a 4×4 MIMO channel with independent quasi-static Rayleigh-fading channel coefficients. The system throughput is 4 bits/channel use, $J = 128$, $L = 1$ and the number of iterations = 20. The performance of the flat precoder is also included for reference.

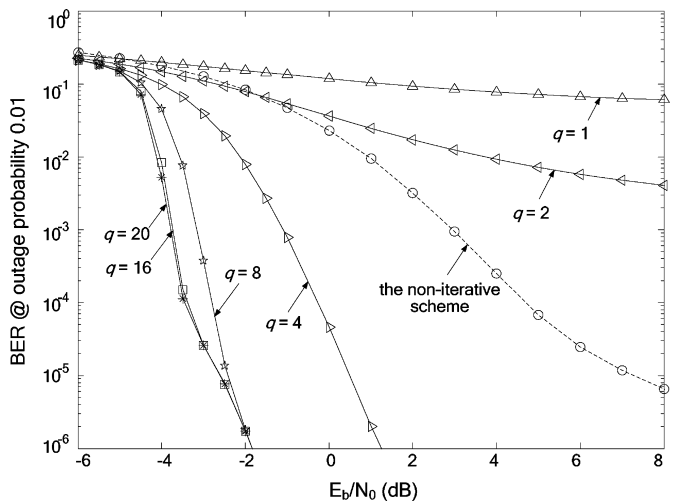


Fig. 4. Impact of the number of iterations on the performance of the optimized precoding system in Fig. 3. The target BER in precoder optimization is set to 10^{-3} and frame length = 4096. Iteration number (q) = 1, 2, 4, 8, 16, and 20, respectively. The solid lines are for the iterative scheme and dashed line for the non-iterative scheme.

Fig. 4 shows the effect of the number of iterations on the optimized precoding system considered in Fig. 3. The target BER in the precoder optimization is set to 10^{-3} . We can see that the system performance nearly converges in 20 iterations. This can also be verified using the SINR-variance transfer chart [14] (similar to the EXIT chart in [39]) shown in Fig. 5, in which the $\phi(v)$ curve is averaged over 10 000 channel realizations. We always set the number of iterations to 20. In principle, the optimization of the precoder is to adjust \mathbf{W} so as to match ϕ with ψ (as ϕ is a function of \mathbf{W}). Good matching potentially leads to improved performance at the cost of more iterations. This performance-cost tradeoff has been observed extensively for turbo codes, LDPC codes, and other iterative detection techniques [3].

We have also included the results for the precoder optimized for one iteration (i.e., $Q = 1$ in Section III-F) in Figs. 4 and

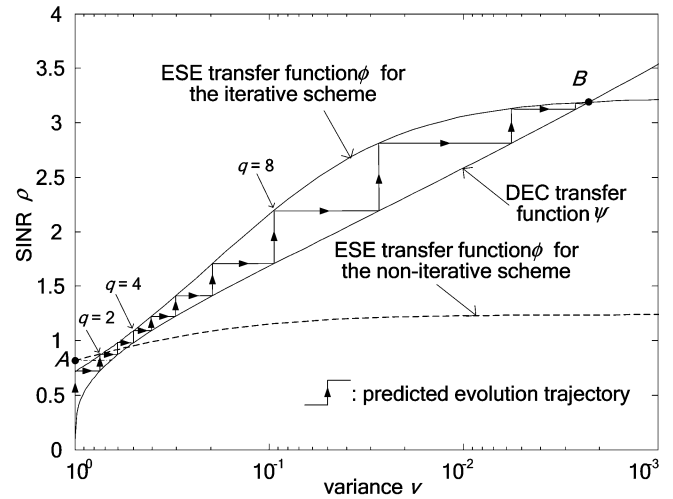


Fig. 5. SINR-variance transfer chart for the optimized precoding system in Fig. 4. The ESE transfer function is obtained by averaging over 10 000 realizations of the channel at $E_b/N_0 = -3.5$ dB. The solid lines are for the iterative scheme and dashed line for the non-iterative scheme.

5, which is equivalent to a conventional non-iterative scheme (e.g., the precoder defined by (15) in [5]). We can see from Fig. 4 that significant performance gain can be achieved by introducing iteration. The cause of this gain is that interference level reduces (and so SINR increases) when more reliable feedback information is available. The proposed scheme adapts to this SINR change, which can be seen more clearly from Fig. 5. The ϕ -curve in Fig. 5 for the non-iterative scheme has a relatively high value at $v = 1$ than the proposed scheme, indicating a better performance for the former in the first iteration. However, when more iterations are involved, the ϕ -curve of the proposed scheme leads to better performance. Comparing the two ϕ -curves in Fig. 5, we can see that the SINR in the ESE output is significantly increased from about 0.8 at point A for the conventional approach to about 3.2 at point B for the proposed scheme. This explains the significant performance difference between the two methods at $E_b/N_0 = -3.5$ dB in Fig. 4. (Note: E_b/N_0 is fixed at -3.5 dB in Fig. 5.)

Fig. 5 suggests that further performance improvement is possible by shaping $\phi(v)$ and $\psi(\rho)$ together. This is indeed the case and, in fact, near capacity performance can be achieved by doing so. The topic is beyond the scope of this paper and is presented in [15].

Fig. 6 shows the impact of N_r , N_t , and L on the BER performance of the proposed precoding scheme. For comparison, the performance of the water-filling precoder is also included. It is seen that the system performance improves as N_r , N_t , and/or L increases, thanks to multi-path and multi-antenna diversities. The channel capacities are also included for reference. There is a gap of about 4 dB between the simulated performance and its corresponding capacity. This gap is due to the use of the convolutional code. The general trend for different settings roughly agrees with what is promised by capacity analysis. For example, about 3 dB power gain is obtained when the dimension of the MIMO channels increases from 2×2 to 4×4 , which is well expected as the capacity of a MIMO channel increases linearly with $\min(N_r, N_t)$ [30]–[32]. The above observations demonstrate that our proposed precoding scheme can indeed achieve

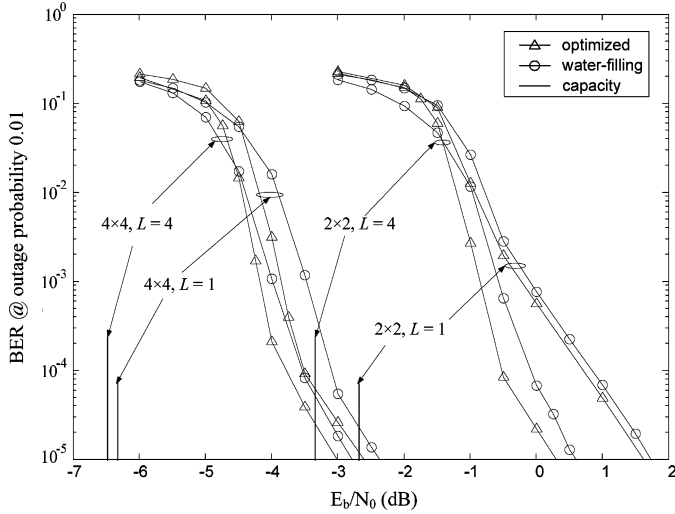


Fig. 6. Performance of the precoding scheme in the quasi-static Rayleigh fading MIMO ISI channels with various N_r, N_t and L . $J = 128$. The throughput is 2 bits per channel use for the 2×2 MIMO system, and 4 bits per channel use for the 4×4 MIMO system.

the gain provided by the MIMO and ISI environments. We can also see from Fig. 6 that the water-filling principle provides a low-cost solution with slight performance loss.

IV. PRECODING FOR SINGLE-USER MIMO OFDM SYSTEMS

A. Precoding for OFDM Systems With CSIT

As mentioned in Section II-C and illustrated in Fig. 1(a), the $\mathbf{F}_{N_t}^H$ at the transmitter, the channel \mathbf{H} , and \mathbf{F}_{N_r} at the receiver can be combined to form an underlying OFDM layer. Therefore, the above precoding scheme can also be used in OFDM systems. The transceiver principle and performance analysis are similar to the discussion in Section III with perfect CSIT. The potential advantages are also similar to those demonstrated in Figs. 3–6. For simplicity, we omit details here.

B. Precoding for OFDM Systems Without CSIT

We now consider systems without CSIT. It is known that the optimal scheme in this case should generate the transmitted signals with a flat spectrum [33], [34]. Accordingly, we can simply select a precoder with \mathbf{P} being a Hadamard matrix and

$$\mathbf{W} = \mathbf{I}. \quad (21)$$

We will call this scheme as a *flat precoder*. Assume that (7) still holds: $\mathbf{H} = \mathbf{F}_{N_r}^H \mathbf{U} \mathbf{D} \mathbf{V}^H \mathbf{F}_{N_t}$. As CSIT is not available, we skip \mathbf{V} in Fig. 1. We can still apply the operation related to \mathbf{U}^H at the receiver in Fig. 1 [assuming perfect channel state information at the receiver (CSIR)]. Then we can modify (10) to

$$\mathbf{r}' = \mathbf{A} \mathbf{x} + \boldsymbol{\eta}', \quad (22a)$$

$$\mathbf{A} = \mathbf{D} \mathbf{V}^H \mathbf{P}. \quad (22b)$$

Now apply the LMMSE operation in (11) to (22). The involved matrixes are given as follows:

$$\mathbf{R} = v \mathbf{D} \mathbf{V}^H \mathbf{P} \mathbf{P}^H \mathbf{V} \mathbf{D}^H + \sigma^2 \mathbf{I} = v \mathbf{D} \mathbf{D}^H + \sigma^2 \mathbf{I}. \quad (23)$$

Clearly, \mathbf{R} is still diagonal (and we thus still enjoy low receiver complexity). We also have

$$\boldsymbol{\Omega} = (\mathbf{A}^H \mathbf{R}^{-1} \mathbf{A})_{\text{diag}} = (\mathbf{P}^H \mathbf{V} \boldsymbol{\Sigma} \mathbf{V}^H \mathbf{P})_{\text{diag}} \quad (24a)$$

where

$$\boldsymbol{\Sigma} = \mathbf{D}^H (v \mathbf{D} \mathbf{D}^H + \sigma^2 \mathbf{I})^{-1} \mathbf{D}. \quad (24b)$$

The diagonal entries of $\boldsymbol{\Omega}$ in (24a) are usually not equal in this case, (as opposed to $\boldsymbol{\Omega}$ in (13)). Therefore, the evolution technique discussed in Section III-E (based on (15)) cannot be directly applied here. However, as shown in Appendix B, when L and J are sufficiently large, $\boldsymbol{\Omega}$ can still be approximated by a scaled identity matrix. We can then again devise an evolution procedure to approximately assess the performance of the iterative receiver.

The flat precoder effectively spreads each coded symbol evenly over all subcarriers in transmission and so this is effectively a space–frequency coding scheme. It can be verified that the above flat precoder is, in spirit, similar to the diversity scheme introduced in [35] for SISO fading channels based on DFT matrixes. (The use of Hadamard matrixes here is more cost-effective than the use of DFT matrixes.) We would also like to mention the contribution in [10] on the space-time codes based on Hadamard matrixes. The flat precoder in (21) is more general as it covers the frequency selective fading effect of the subcarriers.

C. Examples

We first present some numerical examples by assuming no CSIT. Fig. 7 shows the BER performance of the flat precoder in SISO and 2×2 MIMO OFDM systems. (Outage is not considered since the transmitter does not know the channel.) For comparison, the performance without precoding in a Rayleigh fading channel is also included. The number of subcarriers $J = 128$. The coefficients of different channel delay taps are i.i.d. (performance with $L = 8, 16, 32, 128$ is provided).

From Fig. 7, the flat precoder significantly outperforms the system without precoding in the SISO case, thanks to the diversity gain provided by the precoding matrix \mathbf{P} .

We can also see a considerable performance gain for the 2×2 MIMO OFDM system relative to the SISO system. The evolution result for the MIMO system based on (32) is also included, which provides reasonably accurate prediction for a large L (e.g., ≥ 16).

Note that the simulated performance in Fig. 7 for fading SISO channel with flat precoding agrees well with the result in Fig. 7 of [35] using a DFT precoder. The scheme discussed in [35] is equivalent to the flat precoder, except that a DFT matrix is used in [35]. A DFT matrix also meets the two criteria introduced in Section III-A. Thus, the scheme in [35] can be treated as a special case of the precoder developed in this paper.

The performance improvement in Fig. 7 results from diversity gain. We now show that an extra gain called water-filling gain can be achieved when CSIT is available. Fig. 8 illustrates the performance of a 2×2 MIMO OFDM system with $L = 4$. At $\text{BER} = 10^{-5}$, the optimized precoder can provide about 4-dB gain over the system with flat precoding. The gap between the

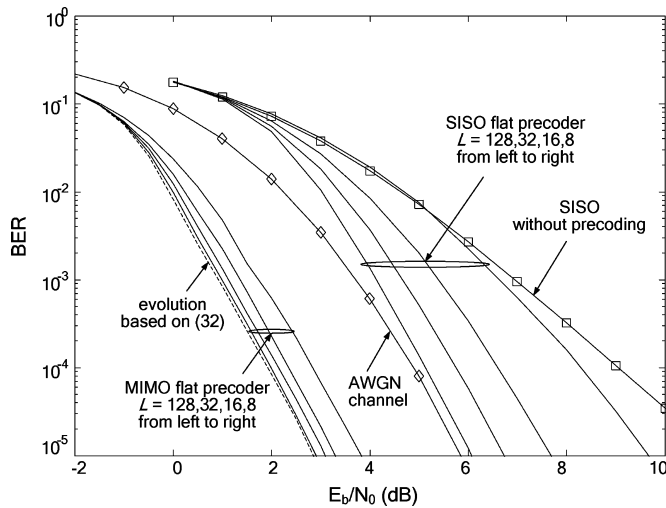


Fig. 7. Performance of the proposed precoder in SISO and 2×2 MIMO OFDM with $J = 128$ subcarriers and $L = 8, 16, 32, 128$ taps. The throughput is 1 bits per channel use for the SISO system, and 2 bits per channel use for the 2×2 MIMO system.

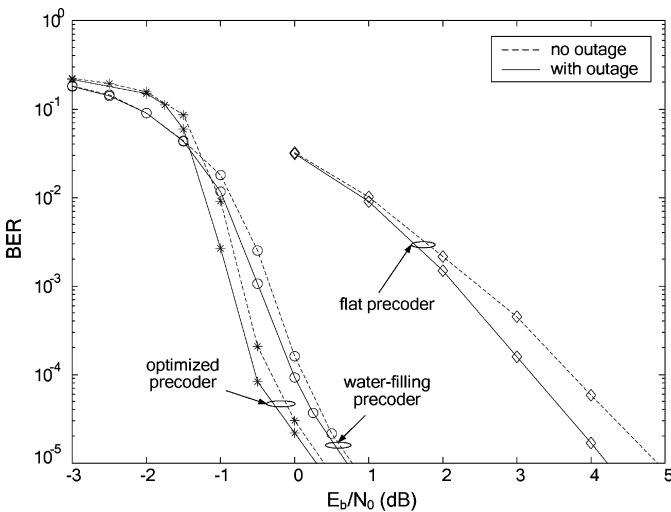


Fig. 8. Performance comparison of different precoding strategies in an OFDM system on a 2×2 MIMO ISI channel with $L = 4$ and $J = 128$. The system throughput is 2 bits per channel use. The capacity is $E_b/N_0 \approx -3.3$ dB. We have included both the results with and without outage in Fig. 8 for comparison.

performance in Fig. 8 and the capacity is about 4 dB at $\text{BER} = 10^{-5}$, which is expected for the convolutional code used. Thus, the proposed scheme can almost fully explore the advantage of CSIT.

We have also included the results for the water-filling precoding discussed in Section III-G. This scheme is optimal assuming ideal codes, but not necessarily so for practical codes. Consequently, we can see in Fig. 8 that, although worse than the optimized precoders by about 0.4 dB, the water-filling precoder still demonstrates an impressive 3.5-dB gain compared with the flat precoder. The water-filling precoder is easier to design than the optimized one and so it can be used as a low-cost alternative.

We will not provide simulation results for the conventional approach using multiple codes. This is due to the difficulty in designing a large number of codes with the rates required. Instead, we approximately estimate the performance of the multi-code

scheme as follows. The performance of the $(7, 5)_8$ convolutional code used in Fig. 8 is $\text{BER} = 10^{-5}$ at $E_b/N_0 \approx 6$ dB in an AWGN channel, which represents a gap of about 6 dB from capacity (i.e., $E_b/N_0 \approx 0.2$ dB at rate = 0.5 bits). The binary convolutional codes with other rates less than 1 have a similar performance gap (at $\text{BER} = 10^{-5}$) of about 6 dB from capacity. For rate higher than 1, trellis coded modulation (TCM) [36] with similar detection complexity also has a performance gap (at $\text{BER} = 10^{-5}$) of about 6 dB from capacity. Thus the performance obtained after water-filling by mixing binary and TCM schemes of various rates are expected to be also roughly 6 dB away from capacity. For comparison, the performance shown in Fig. 8 for the optimized precoder is only 4 dB away from the capacity. (i.e., $E_b/N_0 \approx -3.3$ dB at rate = 2 bits, as shown in Fig. 6.)

Thus, we can see that the proposed scheme provides an attractive alternative to the traditional water-filling method involving multiple codes. The new method is much simpler, more straightforward, and has competitive performance.

V. PRECODING FOR MULTI-USER MIMO OFDM SYSTEMS

We now extend the precoding technique to MIMO orthogonal frequency-division multiple access (OFDMA) systems. The presence of multiple users brings about a new form of diversity gain, namely, multiuser gain. This is in addition to the multi-path and multi-antenna diversity gains discussed in the previous sections. We will show that our precoding technique can help to achieve this gain.

A. MIMO OFDMA

Consider a MIMO OFDM system, in which each subcarrier is assigned exclusively to one user (i.e., in the so-called OFDMA form). Let K be the number of active users, and S_k be the set of subcarriers assigned to user k . For each user k , the single-user precoding technique described in Section IV can be directly applied to the set of subcarriers in S_k . We will briefly discuss below how to find $\{S_k\}$ so as to minimize the total transmission power.

B. Subcarrier Assignment

The optimal subcarrier assignment is, in general, a complicated problem [37], [38]. We will take a suboptimal approach based on the rate-craving greedy (RCG) algorithm proposed in [38].

We assume that the channels seen by different users are i.i.d. For simplicity, we also assume that every user supports the same information rate of R/K , where R is referred as the sum-rate of the system. For convenience, assume that all $\{S_k\}$ have the same size J/K . (More general resource allocation algorithms can be found in, e.g., [38].) For simplicity, assume $N_t = N_r = N$. In a multiuser MIMO OFDM system, each subcarrier seen by a particular user can be modeled by an MIMO channel and SVD can be applied to decompose it into N eigenmodes. We use R as the target rate per subcarrier. Applying water-filling to the N eigenmodes of each subcarrier seen by each user (and assuming ideal codes), we can find the required power for each subcarrier. In this way, we obtain a total of $J \times K$ estimates denoted by

$\{P_{k,j}\}$. Here, $\{P_{k,j}\}$ can be used as a rough indication⁴ of the qualities of the subcarriers seen by different users.

The subcarrier assignment procedure is as follows.

1) Form $\{S_k\}$ by allocating each subcarrier j to user k with

$$k = \arg \min_{k'} \{P_{k',j}\}. \quad (25)$$

2) If $|S_k| > J/K$, we move subcarrier j from S_k to $S_{k'}$ with $|S_{k'}| < J/K$ if such a subcarrier reallocation incurs the minimum increase in the sum-power among all possible choices. (For this purpose, we search for j and k' such that $P_{k',j} - P_{k,j}$ is minimized for all j and k' with $|S_{k'}| < J/K$.) Repeat until all $\{S_k\}$ have the same size.

The key of the proposed precoding technique, assuming CSIT, is to best exploit the fluctuation in the gains of the subcarriers. However, after subcarrier allocation, only relatively good subcarriers are assigned to each user. The fluctuation among the subcarrier gains in each S_k is reduced in a SISO OFDM environment. Then the precoding technique may not be necessary due to reduced benefit.

The situation is quite different in a MIMO OFDM environment. After SVD, N eigenmodes are obtained for each sub-carrier. The fluctuation is still substantial among the gains of different eigenmodes. The proposed precoding technique can exploit this fluctuation and still bring about considerable multiuser diversity gain, as demonstrated below.

C. Examples

Consider an OFDMA system in which each user sees a Rayleigh fading MIMO ISI channel with $N_r = 2, N_t = 2, L = 4$, and $J = 128$. Each user is allocated J/K subcarriers using the strategy described in Section V-B. The channel code is still the rate-1/2 convolution code with generator polynomial $(7, 5)_8$, followed by random interleaving and QPSK modulation. Frame length is fixed at 4096 information bits per user.

Recall from Fig. 6 that the difference between the optimized precoder and the water-filling precoder is not significant. Therefore, we adopt water-filling precoding in determining \mathbf{W} , as the optimization for multiple users is quite time-consuming.

We included the performance of a single-user flat precoder as reference. We can see that, for the single user case, about 3.5-dB performance gain can be achieved by choosing an appropriate \mathbf{W} . The performance of the system further improves as K increases, which attributes to the multiuser diversity. The total power required using ideal multi-rate codes is also included for reference. From Fig. 9, the theoretical multiuser gain by increasing K from 1 to 16 is about 2 dB, which is roughly in agreement with the relative gain achieved by simulated system.

VI. CONCLUSION

In this paper, we have developed a precoding scheme with iterative detection. The precoder can be optimized using convex

⁴The indication is rough since R is only the average rate per channel. (Ignore prefix. A frame of OFDM scheme of length J involves J times channel uses. Thus, a rate of R per channel use implies a rate of R per subcarrier in average.) After water-filling, the actual transmission rate on each channel may be different from R .

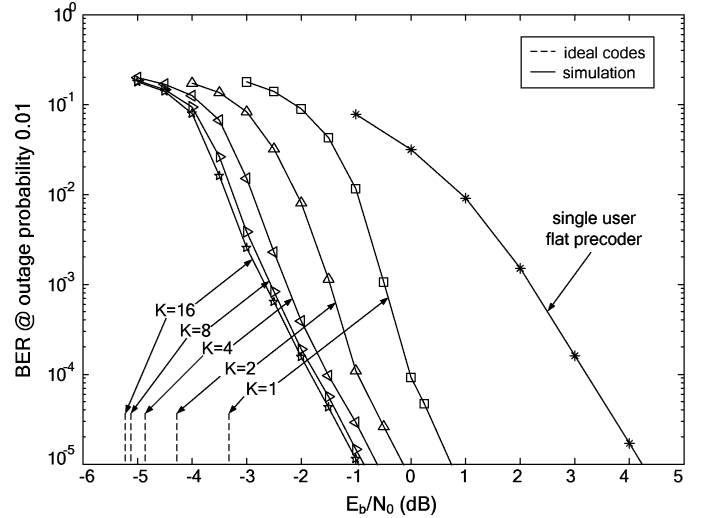


Fig. 9. Performance of the water-filling precoder in 2×2 MIMO OFDMA system with 1, 2, 4, 8, and 16 users, respectively. $L = 4, J = 128$, and total throughput = 2 bits per channel use. The performance limits using ideal codes are also included for reference.

optimization. The complexity involved in the iterative detection is relatively low based on the FHT algorithm. We have also discussed the application of the proposed precoding scheme in multi-user environments.

With both evolution prediction and simulation results, we have demonstrated that the proposed scheme can achieve significant diversity gain, water-filling gain, and multi-user gain.

It is also interesting to study the performance limit of the proposed scheme. It can be shown that near-capacity performance can be achieved by properly shaping both ϕ and ψ curves (see Fig. 5). This is discussed in detail in [15].

APPENDIX A PROOF OF LEMMA 1

We first show that the harmonic mean

$$f(z_0, \dots, z_{J-1}) \equiv \left(J^{-1} \sum_{j=0}^{J-1} 1/z_j \right)^{-1}$$

is concave in $\{z_j\}_{j=0}^{J-1}$ for $z_j > 0, 0 \leq j \leq J-1$. To see this, calculate

$$\begin{aligned} \frac{\partial^2 f}{\partial z_j^2} &= 2J \left(\sum_{j=0}^{J-1} 1/z_j \right)^{-3} \\ &\quad \times \left(1/z_j^4 - z_{J-1}^{-3} \left(\sum_{j=0}^{J-1} 1/z_j \right) \right) \\ \frac{\partial^2 f}{\partial z_i \partial z_j} &= 2J \left(\sum_{j=0}^{J-1} 1/z_j \right)^{-3} \frac{1}{z_i^2 z_j^2}, \quad \text{for } i \neq j. \end{aligned}$$

Then, the Hessian matrix $\nabla^2 f$ [29] can be expressed as

$$\nabla^2 f = 2J \left(\sum_{j=0}^{J-1} 1/z_j \right)^{-3} \times \left(\mathbf{q}\mathbf{q}^T - \text{diag} \{z_0^{-3}, z_1^{-3}, \dots, z_{J-1}^{-3}\} \left(\sum_{j=0}^{J-1} 1/z_j \right) \right)$$

where $\mathbf{q} \equiv [1/z_0^2, 1/z_1^2, \dots, 1/z_{J-1}^2]^T$. We need to show that $\nabla^2 f$ is semi-definite positive, i.e., that, for $\forall \mathbf{t}$

$$\mathbf{t}^T \nabla^2 f(\mathbf{z}) \mathbf{t} = 2J \left(\sum_{j=0}^{J-1} 1/z_j \right)^{-3} \cdot \left(\left(\sum_{j=0}^{J-1} t_j^2/z_j^2 \right)^2 - \left(\sum_{j=0}^{J-1} t_j^2/z_j^3 \right) \left(\sum_{j=0}^{J-1} 1/z_j \right) \right) \leq 0.$$

But this follows from the Cauchy–Schwarz inequality $(\mathbf{a}^T \mathbf{a})(\mathbf{b}^T \mathbf{b}) \geq (\mathbf{a}^T \mathbf{b})^2$ applied to the vectors with components

$$a_j = t_j/z_j^{3/2}, \quad \text{and } b_j = 1/z_j^{1/2}.$$

With the above fact, Lemma 1 holds immediately by letting $z_j = 1/v + |W(j, j)|^2 |D(j, j)|^2 / \sigma^2$ in (16c).

APPENDIX B

AN APPROXIMATE EXPRESSION OF (24a)

For simplicity, we only consider a special case with $N_r = N_t = 2$. (Our discussion can be readily extended to more general cases.) Define a Hadamard matrix with order 2 as

$$\mathbf{P}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix}. \quad (26a)$$

The Hadamard matrixes with other orders can be defined recursively using \mathbf{P}_2 as

$$\begin{aligned} \mathbf{P}_4 &= \frac{1}{\sqrt{2}} \begin{bmatrix} +\mathbf{P}_2 & +\mathbf{P}_2 \\ +\mathbf{P}_2 & -\mathbf{P}_2 \end{bmatrix} \\ \mathbf{P}_8 &= \frac{1}{\sqrt{2}} \begin{bmatrix} +\mathbf{P}_4 & +\mathbf{P}_4 \\ +\mathbf{P}_4 & -\mathbf{P}_4 \end{bmatrix} \quad \text{etc.} \end{aligned} \quad (26b)$$

We write

$$\mathbf{V} = \text{diag}\{\mathbf{V}_0, \mathbf{V}_1, \dots, \mathbf{V}_j, \dots, \mathbf{V}_{J-1}\} \quad (27)$$

where \mathbf{V}_j is a 2×2 matrix characterizing the MIMO effect on the j th subcarrier. Correspondingly, we can also write

$$\mathbf{\Sigma} = \text{diag}\{\mathbf{\Sigma}_0, \mathbf{\Sigma}_1, \dots, \mathbf{\Sigma}_j, \dots, \mathbf{\Sigma}_{J-1}\} \quad (28)$$

where all $\{\mathbf{\Sigma}_j\}$ are diagonal. Now assume that \mathbf{P} in (22) is a normalized Hadamard matrix with order $J' = 2J$. (Recall that $N_r = N_t = 2$ and J is the number of subcarriers.)

From (26b), \mathbf{P} can be expressed in a block form with each block being $\pm(1/\sqrt{J})\mathbf{P}_2$. We decompose $\mathbf{\Omega}$ into a corresponding block form. With some straightforward manipulation, we can show that each 2×2 block on the diagonal line of $\mathbf{\Omega}$ can be expressed as

$$\mathbf{P}_2 \left(\frac{1}{J} \sum_{j=0}^{J-1} \mathbf{V}_j \mathbf{\Sigma}_j \mathbf{V}_j^H \right) \mathbf{P}_2^H. \quad (29)$$

When $L = J$, the sub-channels are i.i.d. Then $\{\mathbf{V}_j\}$ are i.i.d. and so are $\{\mathbf{\Sigma}_j\}$. Note that $\{\mathbf{V}_j\}$ are unitary and $\{\mathbf{\Sigma}_j\}$ non-negative. When J is large, based on the law of large numbers, we have

$$\lim_{J \rightarrow \infty} \frac{1}{J} \sum_{j=0}^{J-1} \mathbf{V}_j \mathbf{\Sigma}_j \mathbf{V}_j^H = \mathbb{E} [\mathbf{V}_j \mathbf{\Sigma}_j \mathbf{V}_j^H] = \omega \mathbf{I}_{2 \times 2} \quad (30)$$

where $\mathbf{I}_{2 \times 2}$ is a 2×2 identity matrix and ω can be computed using (24b) as

$$\omega = \int_0^\infty \frac{|\xi|^2 p(\xi)}{v|\xi|^2 + \sigma^2} d\xi \quad (31)$$

with $p(\xi)$ the empirical density distribution of $\{D(j, j)\}$. Hence,

$$\mathbf{\Omega} = \omega \mathbf{I}. \quad (32)$$

In practice, J is finite and L is usually smaller than J . Then (32) can only be treated as an approximation, i.e., $\mathbf{\Omega} \approx \omega \mathbf{I}$.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable suggestions and comments.

REFERENCES

- [1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [2] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," in *Proc. ICC'93*, Geneva, Switzerland, May 1993.
- [3] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.
- [4] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers part I: Unification and optimal designs," *IEEE Trans. Signal Process.*, vol. 47, no. 7, pp. 1988–2006, Jul. 1999.
- [5] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.
- [6] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.
- [7] Y. Jiang, J. Li, and W. W. Hager, "Joint transceiver design for MIMO communications using geometric mean decomposition," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3791–3803, Oct. 2005.
- [8] F. Xu, T. N. Davidson, J. K. Zhang, and K. M. Wong, "Design of block transceivers with decision feedback detection," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 965–978, Mar. 2006.
- [9] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance analysis and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
- [10] P. Lusina, E. M. Gabidulin, and M. Bossert, "Efficient decoding of space-time Hadamard codes using the Hadamard transform," in *Proc. ISIT2001*, Washington, DC, Jun. 24–29, 2001.

- [11] K. Wu and L. Ping, "A quasi-random approach to space-time codes," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1073–1085, Mar. 2008.
- [12] N. Prasad, I. Berenguer, and X. Wang, "Design of spherical lattice space-time codes," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4847–4865, Nov. 2008.
- [13] H. Bölcskei and A. J. Paulraj, "Space-frequency coded broadband OFDM systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Chicago, IL, Sep. 2000, pp. 1–6.
- [14] X. Yuan, Q. Guo, X. Wang, and L. Ping, "Evolution analysis of low-cost iterative equalization in coded linear systems with cyclic prefixes," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 301–310, Feb. 2008.
- [15] X. Yuan and L. Ping, "Achievable rates of coded linear systems with iterative MMSE detection," in *Proc. IEEE Globecom'09*, Honolulu, HI, Nov. 30–Dec. 4 2009.
- [16] R. K. Yarlagadda and J. E. Hershey, *Hadamard Matrix Analysis and Synthesis: With Applications to Communications and Signal/Image Processing*. Boston, MA: Kluwer, 1996.
- [17] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eldson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Commun. Mag.*, vol. 40, no. 4, pp. 58–66, Apr. 2002.
- [18] X. Yuan, Q. Guo, X. Wang, and L. Ping, "Low-complexity iterative detection in multi-user MIMO ISI channels," *IEEE Signal Process. Lett.*, vol. 15, pp. 25–28, 2008.
- [19] L. N. Trefethen and D. Bau, III, *Numerical Linear Algebra*. Philadelphia, PA: SIAM, 1997.
- [20] [Online]. Available: <http://www.3gpp.org>
- [21] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, Jul. 1999.
- [22] M. Tüchler, R. Koetter, and A. C. Singer, "Turbo equalization: Principles and new results," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 754–767, May 2002.
- [23] L. Ping, L. Liu, K. Wu, and W. K. Leung, "Interleave division multiple-access," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 938–947, Apr. 2006.
- [24] P. A. Hoeher, H. Schoeneich, and J. C. Fricke, "Multi-layer interleave-division multiple access: Theory and practice," *Eur. Trans. Telecommun.*, vol. 19, pp. 523–536, Jan. 2008.
- [25] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall PTR, 1993.
- [26] L. Ping, J. Tong, X. Yuan, and Q. Guo, "Superposition coded modulation and iterative linear MMSE detection," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 6, pp. 995–1204, Aug. 2009.
- [27] X. Yuan, H. Li, L. Ping, and X. Lin, "Optimized spectrum-shaping strategy for coded single-carrier transmission," *IEEE Signal Process. Lett.*, vol. 15, pp. 809–812, 2008.
- [28] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2551–2567, Jun. 2006.
- [29] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [30] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, Mar. 1998.
- [31] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–598, Nov. 1999.
- [32] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.
- [33] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [34] Z. Liu, Y. Xin, and G. B. Giannakis, "Linear constellation precoding for OFDM with maximum multipath diversity and coding gains," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 416–427, Mar. 2003.
- [35] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Performance of BICM-ID with signal space diversity," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1732–1742, May 2007.
- [36] S. X. Ng, T. H. Liew, and L. Hanzo, "Comparative study of TCM, TICM, BICM and BICM-ID schemes," in *Proc. IEEE VTC'2001*, Rhodes, Greece, May 2001, pp. 2450–2454.
- [37] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1757, Oct. 1999.
- [38] D. Kivanc, G. Li, and H. Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Trans. Wireless Commun.*, vol. 2, no. 6, pp. 1150–1158, Nov. 2003.
- [39] S. Ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.



X. Yuan, (S'04–M'08) received the B.S. degree in electronic and information systems from Shanghai Jiaotong University, Shanghai, China, the M.S. degree in circuit and systems from Fudan University, Shanghai, and the Ph.D. degree at City University of Hong Kong.

He is now with the Department of Electronic Engineering, University of Hawaii, Manoa, HI. His research interests include signal processing, coding design, optimized communication systems design, and networks.



C. Xu, received the B.E. degrees in electronic engineering from Xian Jiaotong University, Xi'an, China, in 2005. He is currently working towards the Ph.D. degree at Tsinghua University, Beijing, China.

His research interests include information theory, multiple antenna transmission, and broadband wireless communications.



L. Ping, (S'87–M'91–SM'06) received the Ph.D. degree from Glasgow University, Glasgow, U.K., in 1990.

He lectured in the Department of Electronic Engineering, Melbourne University, Melbourne, Australia, from 1990 to 1992, and worked as a Member of Research Staff at Telecom Australia Research Laboratories from 1993 to 1995. He has been with the Department of Electronic Engineering, City University of Hong Kong, since January 1996, where he is now a Chair Professor. His research

interests are communications systems and coding theory.

Dr. Ping was awarded a British Telecom-Royal Society Fellowship in 1986, the IEE J. J. Thomson premium in 1993, and a Croucher Senior Research Fellowship in 2005.



X. Lin, received the B.E. degree from the Department of Electrical Engineering, Tsinghua University, Beijing, China, in 1970.

Since then, he has been with Tsinghua University as an Assistant Teacher, Lecturer, Associate Professor, and currently as a Full Professor. He was also a Visiting Scholar with Stuttgart University, Stuttgart, Germany, in 1988 and Hong Kong University, Hong Kong, in 1994. Currently, he is the Director of National Key Labs on Microwaves and Digital Communications, Tsinghua University. His

research interests include high-speed switching, MPLS, broadband networks, and IC design.