Rate-Maximizing Power Allocation in OFDM Based on Partial Channel Knowledge

Yingwei Yao, *Member, IEEE,* and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—Power loading algorithms improve the data rates of orthogonal frequency division multiplexing (OFDM) systems. However, they require the transmitter to have perfect channel state information, which is impossible in most wireless systems. We investigate the effects of imperfect (and thus partial) channel feedback on the throughput of OFDM systems. Two channel uncertainty models are studied: 1) the ergodic model, where average rate is the figure of merit and 2) the quasi-static model, where outage rate is relevant. Rate-power allocation algorithms are developed. The throughput achieved by these algorithms and the effects of channel multipath are investigated analytically and with simulations.**

***Index Terms*—Adaptive modulation, average rate, orthogonal frequency division multiplexing (OFDM), outage rate, partial channel state information (CSI).**

I. Introduction

ORTHOGONAL frequency division multiplexing (OFDM) provides a low-complexity means of combating intersymbol interference (ISI) that arises due to the delay spread of communication channels. It has found applications in many digital communication systems, such as digital audio broadcasting (DAB), digital video broadcasting (DVB), and wireless local area networking (WLAN). In multipath channels, different subcarriers of an OFDM transmission are generally received with different channel gains. To maximize the information rate, power- and bit-loading algorithms have been derived to adaptively adjust power and data rates across subcarriers according to the channel's condition [1]. These algorithms usually assume perfect knowledge of the channel state information at the transmitter (CSIT). While this assumption is reasonable in wireline systems, where the channel remains typically invariant, wireless channels are randomly varying over time, making it impossible for the transmitter to acquire perfect CSIT.

Manuscript received February 24, 2003; revised September 29, 2003, February 24, 2004; accepted February 26, 2004. The editor coordinating the review of this paper and approving it for publication is L. I. Vandendorpe. This work was prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

Y. Yao was with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA. He is now with the Department of Electrical and Computer Engineering, University of Illinois, Chicago, IL 60607 USA (e-mail: yyao@ece.uic.edu).

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

Digital Object Identifier 10.1109/TWC.2005.847022

The problem of optimizing transmission strategies with imperfect channel feedback has been addressed in a number of recent publications in the context of single carrier multiple-antenna transmissions over multi-input multi-output (MIMO) channels [2]–[5]. Simulations indicate that substantial gain is possible with partial channel knowledge. The impact of imperfect channel information on multicarrier systems like OFDM has also been studied recently. In [6], Leke and Cioffi investigated the effects of the channel estimation error at the receiver on the uncoded bit-error rate (BER) performance of OFDM systems. In [7], Monte-Carlo simulations were conducted to study how imperfect channel feedback affects the rates achieved by the water-filling algorithm, and the successive bit allocation algorithm [8]. Bit- and power-loading algorithms were pursued in [9]–[11], where partial CSIT was utilized to adapt the constellation size and/or the power, adhering to a certain target BER per subcarrier.

In this paper, we investigate the fundamental limit imposed on the information rate of an OFDM system with partial CSIT. The system model and assumptions are presented in Section II. In Section III, we study the case when the channel uncertainty can be modeled as an ergodic process. A power loading algorithm maximizing the average mutual information is derived, and simulations comparing it to uniform power loading and water-filling approaches are presented. For the channel feedback received by the transmitter to be useful, the channel should not change much over the duration of the feedback delay, which motivates well a block-fading channel model [12]. We adopt such a channel model in Section IV, and study the impact of imperfect CSIT on the outage rate of OFDM systems. We show that the channel information error may severely reduce the outage rate of an OFDM system. Optimal and suboptimal schemes for maximizing the outage rate are obtained. The effects of channel multipath are discussed. Finally, we present our conclusions in Section V.

Notational conventions are as follows: upper and lower case bold symbols are used to denote matrices and vectors, respectively; $\mathbf{I}_N$ denotes an $N \times N$ identity matrix; and $(\cdot)^T$ and $(\cdot)^{\mathcal{H}}$ denote matrix transpose and Hermitian transpose, respectively.

II. System Model

Consider OFDM transmissions with block length N, through a frequency-selective multipath fading channel that in discrete-time baseband equivalent form is described by the taps $\{h_l\}_{l=0}^{L}$. After removing the cyclic prefix and performing fast Fourier transform (FFT) at the receiver, a received OFDM block (symbol) can be written as

$$\boldsymbol{y} = \mathbf{D}_{h_f}\boldsymbol{x} + \boldsymbol{w} \tag{1}$$

where $N \times 1$ vectors $\boldsymbol{y}$ and $\boldsymbol{x}$ denote received and transmitted blocks, respectively; the noise block $\boldsymbol{w}$ is assumed to be a circularly symmetric complex Gaussian random vector with distribution $\boldsymbol{w} \sim \mathcal{CN}\left(\mathbf{o}, \sigma_w^2 \mathbf{I}_N\right)$; and $\mathbf{D}_{h_f}$ is a diagonal matrix with diagonal elements[1] $\boldsymbol{h}_f = [h_f(0), \ldots, h_f(N-1)]^T$, where $h_f(k) = \sum_{l=0}^{L} h_l e^{-j2\pi kl/N}$

A. Partial CSIT

We consider operations under partial (imperfect) CSIT, but with full (perfect) CSI at the receiver. The transmitter acquires channel knowledge either via a feedback channel, or, by channel estimation in a time division duplex (TDD) operation. The partial CSIT includes the channel feedback $\hat{\boldsymbol{h}}_{f,0}$ that is treated as a deterministic[2] mean (or "nominal" CSIT) plus a perturbation (error) term $\tilde{\boldsymbol{h}}_f$ with known probability density function (pdf) to account for various sources of uncertainty, i.e.,

$$\check{\boldsymbol{h}}_f = \hat{\boldsymbol{h}}_{f,0} + \tilde{\boldsymbol{h}}_f \tag{2}$$

where $\check{\boldsymbol{h}}_f \sim \mathcal{CN}(\hat{\boldsymbol{h}}_{f,0}, \boldsymbol{\Sigma}_{\tilde{h}_f})$. The CSIT error $\tilde{\boldsymbol{h}}_f$ may arise due to many different reasons, e.g., channel estimation error, and/or feedback delay combined with Doppler spread and quantization error [5]. While this model looks similar to the mean feedback model adopted in many recent publications, a notable difference is that, due to the structure of OFDM systems, the error covariance matrix $\boldsymbol{\Sigma}_{\tilde{h}_f}$ in general can not be assumed to be a scalar multiple of the identity matrix. Later, we will see that this difference has important implications in OFDM systems' resilience to CSIT errors. However, for now, we compute in closed form the error covariance matrix $\boldsymbol{\Sigma}_{\tilde{h}_f}$ for an example which we will use in Sections II and III.

We assume that the channel stays invariant over the duration of an OFDM block. During the nth block, it can be modeled as an Lth-order finite impulse response (FIR) filter with coefficients $\boldsymbol{h}(n) \sim \mathcal{CN}(\mathbf{o}, (1/L+1)\mathbf{I})$; while from block to block, $\boldsymbol{h}(n)$ is slowly time-varying according to Jakes' model [13] $E[\boldsymbol{h}(m)\boldsymbol{h}^{\mathcal{H}}(n)] = J_0(2\pi f_d |m-n| T_b)\mathbf{I}_{L+1}/(L+1)$, where T_b is the block duration, and f_d is the Doppler spread. The ambient channel noise is independent identically distributed (i.i.d.) Gaussian with variance σ_w^2. CSIT is updated every frame comprising L_F OFDM blocks. For channel estimation purposes, a block of training symbols is transmitted at the beginning of every frame. Let us consider the minimum-mean-square-error (MMSE) predictor of $\boldsymbol{h}((p+Q)L_F)$, $Q > 0$, based on the ν training blocks $\boldsymbol{y}_\nu(p) = [\boldsymbol{y}^T(pL_F), \ldots, \boldsymbol{y}^T((p-\nu+1)L_F)]^T$, where

$$\boldsymbol{y}(nL_F) = \sqrt{N}\mathbf{S}_t \mathbf{W}_L \boldsymbol{h}(nL_F) + \boldsymbol{w}(nL_F) \tag{3}$$

with the diagonal matrix $\mathbf{S}_t = \mathrm{diag}(s_{1t}, \ldots, s_{Nt})$ containing the training symbols, and $\mathbf{W}_L$ denoting the truncated unit-norm FFT matrix of size $N \times (L+1)$, with entries

$$[\mathbf{W}_L]_{ml} = \frac{1}{\sqrt{N}} e^{-j(2\pi/N)ml} \quad m = 0, \ldots, N-1; \quad l = 0, \ldots, L. \tag{4}$$

[1]Subscript f will denote frequency-domain quantities.

[2]Subscript 0 will be used throughout this paper to denote the realization of the random quantity.

We will assume constant modulus training symbols, i.e., $|s_{1t}|^2 = \cdots = |s_{Nt}|^2 = P_t$. Let $\mathbf{J}$ denote the $\nu \times \nu$ matrix whose (i,j)th entry is $J_0(|i-j|\tau_F)$, and $\tau_F = 2\pi f_d L_F T_b$. Computing the MMSE of estimating $\boldsymbol{h}((p+Q)L_F)$ and applying the Fourier transform, we obtain the covariance matrix of the error $\tilde{\boldsymbol{h}}_f((p+Q)L_F)$ as follows (the details have been omitted to save space):

$$\boldsymbol{\Sigma}_{\tilde{h}_f} = E\left[\tilde{\boldsymbol{h}}_f \tilde{\boldsymbol{h}}_f^{\mathcal{H}}\right] = \frac{N\sigma_H^2}{L+1} \mathbf{W}_L \mathbf{W}_L^{\mathcal{H}}. \tag{5}$$

The value of σ_H^2 depends on the system parameters

$$\sigma_H^2 = 1 - \boldsymbol{x}_J^{\mathcal{H}} \left[\frac{(L+1)\sigma_w^2}{NP_t}\mathbf{I}_\nu + \mathbf{J}\right]^{-1} \boldsymbol{x}_J \tag{6}$$

where $\boldsymbol{x}_J = [J_0(Q\tau_F) \ldots J_0((Q+\nu-1)\tau_F)]^T$. In practical OFDM systems, we always select $N > L+1$. From (5), we can verify that the channel estimation errors of different subcarriers are correlated.

Remark: While we have assumed a uniform power delay profile in the previous derivation, almost all the analytical results in the rest of this paper can be applied to nonuniform power delay profiles such as exponential power delay profile.

B. Capacity of OFDM Systems

When the transmitter has deterministically perfect knowledge of the channel, i.e., $\hat{\boldsymbol{h}}_{f,0} = \boldsymbol{h}_{f,0}$, it is well known that the maximum mutual information between the input and the output of an OFDM system $I(\boldsymbol{x}; \boldsymbol{y} | \boldsymbol{h}_f = \boldsymbol{h}_{f,0})$ is attained when $\boldsymbol{x} \sim \mathcal{CN}(\mathbf{o}, \boldsymbol{\Lambda})$, and is given by [14]

$$I(\boldsymbol{x}; \boldsymbol{y} | \boldsymbol{h}_f) = \log\det\left(\mathbf{I}_N + \frac{\mathbf{D}_{h_f} \boldsymbol{\Lambda} \mathbf{D}_{h_f}^{\mathcal{H}}}{\sigma_w^2}\right) \tag{7}$$

which is maximized when $\boldsymbol{\Lambda}$ is diagonal, according to Hadamard's inequality [15].

With channel uncertainty at the transmitter, I is random, and our objective is to find the covariance matrix $\boldsymbol{\Lambda}_o$ which maximizes the expected mutual information

$$\boldsymbol{\Lambda}_o = \arg\max_{\boldsymbol{\Lambda}} E_{\check{\boldsymbol{h}}_f}\left[I(\boldsymbol{x}; \boldsymbol{y} | \check{\boldsymbol{h}}_f)\right] \tag{8}$$

or, maximizes the outage rate for a given outage probability P_{out}

$$\boldsymbol{\Lambda}_o = \arg\max_{\boldsymbol{\Lambda}} \sup\left\{R \,\middle|\, P_{\check{\boldsymbol{h}}_f}\left(I(\boldsymbol{x}; \boldsymbol{y} | \check{\boldsymbol{h}}_f) < R\right) \le P_{\text{out}}\right\} \tag{9}$$

subject to a power constraint $\mathrm{Trace}\{\boldsymbol{\Lambda}\} = NP_0$. For each of these two problems, we can always find a solution $\boldsymbol{\Lambda}$ which is diagonal: Suppose the matrix $\boldsymbol{\Lambda}_o$ which maximizes the outage rate is not diagonal. Construct a diagonal matrix $\mathbf{D}(\boldsymbol{\Lambda}_o)$ whose diagonal elements are the same as those of $\boldsymbol{\Lambda}_o$. From Hadamard's inequality, we can see that for any realization $\check{\boldsymbol{h}}_{f,0}$ of $\check{\boldsymbol{h}}_f$, $\det\left(\mathbf{I}_N + \mathbf{D}_{\check{h}_{f,0}} \mathbf{D}(\boldsymbol{\Lambda}_o) \mathbf{D}_{\check{h}_{f,0}}^{\mathcal{H}} / \sigma_w^2\right) \ge \det\left(\mathbf{I}_N + \mathbf{D}_{\check{h}_{f,0}} \boldsymbol{\Lambda}_o \mathbf{D}_{\check{h}_{f,0}}^{\mathcal{H}} / \sigma_w^2\right)$. So $\mathbf{D}(\boldsymbol{\Lambda}_o)$ also maximizes the outage rate.[3] Similar arguments hold for the maximization of the expected mutual information. So, we can assume

[3]For random variables X and Y, it holds that $X \ge Y\, a.s. \Rightarrow P(X < R) \le P(Y < R)$

without loss of generality that $\mathbf{\Lambda}$ is diagonal. Supposing that $\mathbf{\Lambda} = \mathrm{diag}(\gamma_0, \ldots, \gamma_{N-1})$, we have

$$I(\boldsymbol{x}; \boldsymbol{y}|\check{\boldsymbol{h}}_f) = \sum_{i=0}^{N-1} \log\left(\frac{1+\gamma_i|\check{h}_f(i)|^2}{\sigma_w^2}\right). \quad (10)$$

Based on (10), we will next determine the loadings $\{\gamma_i\}_{i=0}^{N-1}$ to maximize the average and outage mutual information, starting with the former.

Remark: While we will focus on OFDM in this paper, many results herein can also be applied to single-carrier systems with frequency-selective fading by using the framework provided by Hirt and Massey [16].

III. Maximizing Average Rate

In existing works relying on partial CSIT, the expected value of the mutual information in (10) has been adopted as the performance criterion [4]. One reason for the popularity of the resulting average capacity is its relative simplicity (compared with the outage capacity); another reason being that when the channel information error is an ergodic process (more precisely, if $\boldsymbol{h}_f$ and $\hat{\boldsymbol{h}}_f$ are jointly asymptotically stationary and ergodic), the maximum expected mutual information is the Shannon capacity [17].

With *perfect* CSIT, the optimal power allocation has been shown to be water-filling; see e.g., [18]

$$\gamma_i = \left(\frac{1}{a} - \frac{1}{|h_f(i)|^2}\right)^+ \quad (11)$$

where $(x)^+ = \max(x, 0)$ and a is determined by the power constraint $\sum_{i=0}^{N-1}\gamma_i = NP_0$.

When only *partial* CSIT is available, we want to find the set of $\{\gamma_i\}_{i=0}^{N-1}$ that maximizes the expected mutual information

$$E_{\check{\boldsymbol{h}}_f}\left[\sum_{i=0}^{N-1}\log\left(\frac{1+\gamma_i|\check{h}_f(i)|^2}{\sigma_w^2}\right)\right] = \sum_{i=0}^{N-1} E_{\check{h}_f(i)}\left[\log\left(\frac{1+\gamma_i|\check{h}_f(i)|^2}{\sigma_w^2}\right)\right] \quad (12)$$

subject to the constraint $\sum_{i=0}^{N-1}\gamma_i = NP_0$. Using the standard Lagrange multiplier method, we can obtain that the set of $\{\gamma_i\}_{i=0}^{N-1}$ maximizing (12) should satisfy

$$E_{\check{h}_f(i)}\left[\frac{\frac{|\check{h}_f(i)|^2}{\sigma_w^2}}{\frac{1+\gamma_i|\check{h}_f(i)|^2}{\sigma_w^2}}\right] = \min\left(a, E_{\check{h}_f(i)}\left[\frac{|\check{h}_f(i)|^2}{\sigma_w^2}\right]\right) \quad (13)$$

where a is determined by the power constraint, and in order to arrive at (13), we have used the fact that differentiation with respect to γ_i and expectation in the right-hand side (r.h.s.) of (12) are interchangeable since $\partial\log\left(1+\gamma_i|\check{h}_f(i)|^2/\sigma_w^2\right)/\partial\gamma_i$ is continuous both in γ_i and in $\check{h}_f(i)$. There is no closed form solution to this problem, but it can be easily solved numerically if the pdf of $\check{\boldsymbol{h}}_f$ is known; e.g., if in the partial CSIT model (2), $\tilde{\boldsymbol{h}}_f$ is complex normal, then $|\check{h}_f(i)|$ is Ricean. Noticing that the left-hand side of (13) decreases monotonically as γ_i increases, while the total transmit power $\sum_{i=0}^{N-1}\gamma_i$ decreases monotonically as a increases, we can find $\{\gamma_i\}_{i=0}^{N-1}$ and a satisfying (13) by one-dimensional (1–D) search.

```
1) Given an estimate â of a, find {γ̂_i}_{i=0}^{N-1}
that satisfy (13).
2) If Σ_{i=0}^{N-1} γ̂_i is larger (smaller) than
NP_0, increase (decrease) â by a certain
amount.
3) Repeat steps 1 and 2 until the algo-
rithm converges.
```

Example 1: Consider $\check{\boldsymbol{h}}_f \sim \mathcal{CN}\left(\hat{\boldsymbol{h}}_{f,0}, \sigma_H^2\mathbf{I}_N\right)$, perceived as Ricean at the transmitter, where $\hat{\boldsymbol{h}}_{f,0}$ is a channel with zeros at 0.8, $\pm 0.8j$ and $0.8e^{\pm j(\pi/4)}$. The number of subcarriers N is set to 16 in all our simulations. The transmitter has only knowledge of the nominal channel $\hat{\boldsymbol{h}}_{f,0}$, and the perturbation variance σ_H^2. Note that we can assume without loss of generality that the CSIT error distribution across different subcarriers is independent because according to (12), the expected mutual information depends only on the marginal distribution of the CSIT error on each subcarrier. When the signal-to-noise ratio (SNR) is $P_0/\sigma_w^2 = 10$ dB, the optimal power allocation for different values of σ_H^2 is depicted in Fig. 1. We observe that when σ_H^2 is small, the optimal allocation is water-filling according to $\hat{\boldsymbol{h}}_{f,0}$. As σ_H^2 grows larger, the optimal power allocation comes closer to a uniform allocation over all subcarriers. The expected mutual information attained by different power allocation strategies is plotted in Fig. 2 for $\sigma_H^2 = 0.1$. For comparison, we also plot the rate achieved by water-filling under perfect CSIT (according to each realization of $\check{\boldsymbol{h}}_f$). We observe that water-filling under partial CSIT performs well only when the SNR is small, while the converse is true for the uniform power allocation scheme. The optimal power allocation found by solving (13) numerically achieves higher rate than both water-filling (according to the nominal $\hat{\boldsymbol{h}}_{f,0}$), and uniform allocation across the SNR range, and suffers only negligible SNR loss compared with the perfect CSIT benchmark.

Example 2: Now, consider a channel with $L+1 = 2$ taps. The filter coefficients are *i.i.d.* complex circularly symmetric Gaussian random variables with zero mean and variance $1/(L+1)$. An estimate of the channel is sent to the transmitter through a feedback channel. Unlike Example 1, where the nominal channel is deterministic, here, the nominal channel is a random variable depending on the channel realization. The channel estimation/prediction error is $\tilde{\boldsymbol{h}} \sim \mathcal{CN}\left(\mathbf{0}, \sigma_H^2\mathbf{I}_{L+1}/(L+1)\right)$. The expected mutual information of different power allocation methods is plotted in Fig. 3 for $\sigma_H^2 = 0.001$. Again, we use water-filling with perfect CSIT for comparison. We infer that the water-filling (with imperfect CSIT), and the optimal power allocation achieve almost identical information rates. Simulation results in [18] confirmed that for an i.i.d. Rayleigh-flat-fading channel, channel information at the transmitter brings only minimal gain in ergodic capacity. In our case, the fading of different subcarriers is correlated, but the gap between the optimal power allocation and the uniform power allocation is also very small. This can be explained as follows. Consider the ideal case where we have perfect CSIT,

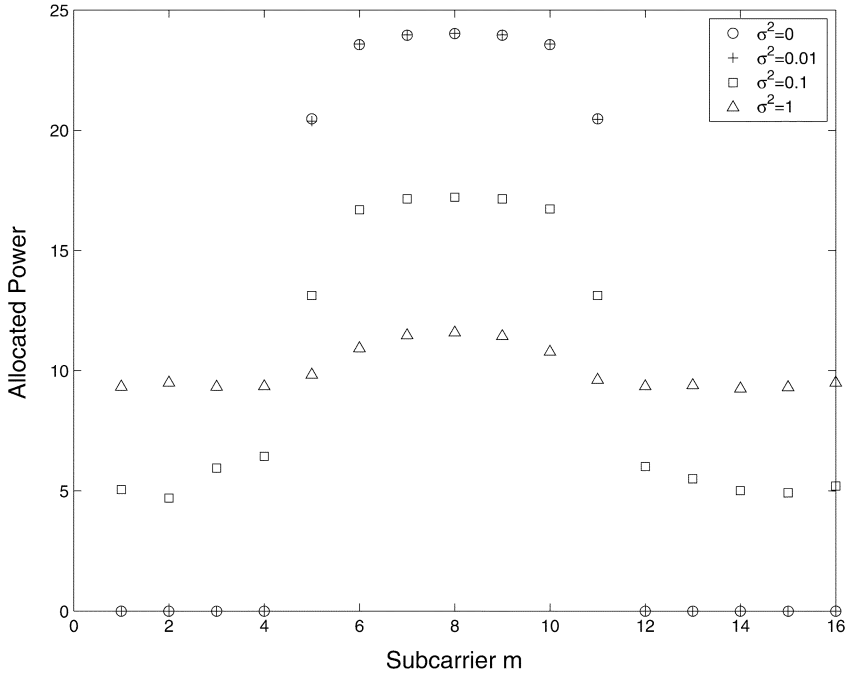

Fig. 1. Optimal power allocation under variable channel uncertainty at the transmitter ($N = 16$ and $L = 5$).

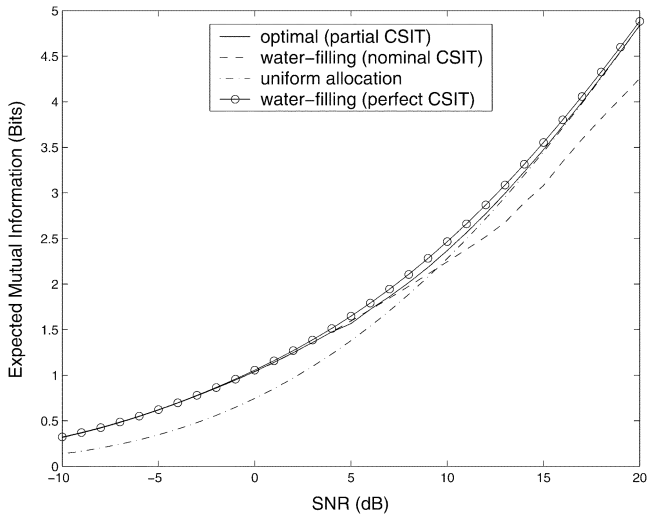

Fig. 2. Expected mutual information ($N = 16$, $L = 5$, and $\sigma_H^2 = 0.1$).

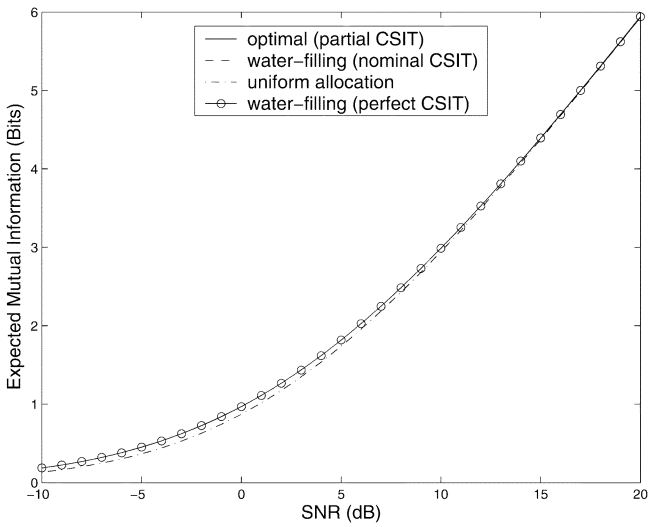

Fig. 3. Expected mutual information ($N = 16$, $L = 1$, and $\sigma_H^2 = 0.001$).

and suppose the $L + 1$ channel taps experience independent (not necessarily identical) Rayleigh fading. It follows that all subcarriers experience identical (not necessarily independent) Rayleigh fading, and the ergodic capacity can be written as

$$C = \max_{\gamma \epsilon \mathcal{A}} \left[E_{\boldsymbol{h}_f} \sum_{i=0}^{N-1} \log\left(1 + \gamma_i(\boldsymbol{h}_f) \frac{|h_f(i)|^2}{\sigma_w^2}\right)\right] \leq \max_{\gamma \epsilon \mathcal{B}} \sum_{i=0}^{N-1} E_{\boldsymbol{h}_f} \left[\log\left(1 + \gamma_i(\boldsymbol{h}_f) \frac{|h_f(i)|^2}{\sigma_w^2}\right)\right]. \quad (14)$$

where $\mathcal{A}$ is the set of power allocations satisfying $\sum_{i=0}^{N-1} \gamma_i(\boldsymbol{h}_f) = NP_0$ for any $\boldsymbol{h}_f$ and $\mathcal{B}$ is the set of power allocations satisfying $\sum_{i=0}^{N-1} E_{\boldsymbol{h}_f}[\gamma_i(\boldsymbol{h}_f)] = NP_0$. If the power allocation $\{\gamma_i^o(\boldsymbol{h}_f)\}_{i=0}^{N-1}$ maximizes the r.h.s. of (14), then $\{\gamma_i^*(h_f(i)) = E[\gamma_i^o(\boldsymbol{h}_f)|h_f(i)]\}_{i=0}^{N-1}$ is also optimal, because

$$\sum_i E[\gamma_i^*(h_f(i))] = \sum_i E[E[\gamma_i^o(\boldsymbol{h}_f)|h_f(i)]] = \sum_i E[\gamma_i^o(\boldsymbol{h}_f)] = NP_0 \quad (15)$$

and

$$E_{\boldsymbol{h}_f} \log\left(\frac{1 + \gamma_i^o(\boldsymbol{h}_f)|h_f(i)|^2}{\sigma_w^2}\right) = E_{h_f(i)}\left[E\left[\log\left(\frac{1 + \gamma_i^o(\boldsymbol{h}_f)|h_f(i)|^2}{\sigma_w^2}\right)|h_f(i)\right]\right] \leq E_{h_f(i)}\left[\log\left(\frac{1 + E[\gamma_i^o(\boldsymbol{h}_f)|h_f(i)]\,|h_f(i)|^2}{\sigma_w^2}\right)\right]. \quad (16)$$

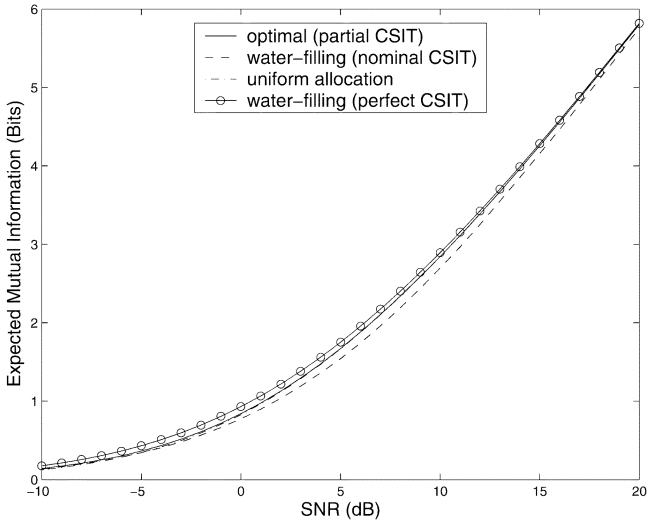

Fig. 4. Expected mutual information ($N = 16$, $L = 1$, and $\sigma_H^2 = 0.5$).

So, we need only to consider power allocation schemes such that γ_i only depends on $h_f(i)$ for each i. From (14), we have

$$C \le \max_{\Omega} \sum_{i=0}^{N-1} E_{h_f(i)} \log\left(\frac{1+\gamma_i(h_f(i))|h_f(i)|^2}{\sigma_w^2}\right) = N \max_{\gamma_1: E[\gamma_1]=P_0} \log\left(\frac{1+\gamma_1|h_f(1)|^2}{\sigma_w^2}\right) \quad (17)$$

where $\Omega = \{\gamma_i(h_f(i)) : \sum_i E[\gamma_i(h_f(i))] = NP_0\}$, and the last equality comes from the symmetry between subcarriers. The r.h.s. of (17) is maximized by time-domain water-filling, and as shown in [18], is only slightly larger than $NE_{h_f(1)} \log\left(1+P_0|h_f(1)|^2/\sigma_w^2\right)$, unless the SNR is very low.

In Fig. 4, we plot the performance of different power allocation schemes when the CSIT error is large ($\sigma_H^2 = 0.5$). As expected, the average capacity of the uniform power allocation is higher than that of water-filling according to the nominal channel, and is almost identical to the optimal scheme. Surprisingly, even with such a large CSIT error, the gap between the water-filling (with imperfect CSIT), and the optimal scheme remains small.

Remark: While in Example 1, the optimal power loading brings substantial gain over both water-filling and uniform allocation, in Example 2, the channel feedback to the transmitter seems to have very little impact. An important question is when to apply power loading, and when not to. It appears that utilizing CSIT will pay off if the probability distributions of different subcarriers are sufficiently different from each other, as in Example 1. When the pdfs of different subcarriers are identical, as in Example 2, the cost of the optimal power loading algorithm outweighs its gain.

IV. Maximizing Outage Rate

When the channel uncertainty can be described as an ergodic process, the expected mutual information described in the previous section closely bounds the rate of an OFDM system. In most wireless communication systems, however, the channel is slowly varying over time, and the channel estimation/prediction error in general can not be modeled as an ergodic process. This is especially relevant for systems relying on partial CSIT, where the transmitter can not obtain perfect channel estimates due to the delay of the feedback channel. Block-fading additive white Gaussian noise (BF-AWGN) channel models are, thus, more pragmatic, and using outage rate (instead of average mutual information) as a figure of merit, is more appropriate [12].

The problem of selecting power loading coefficients to minimize the outage probability given a target data rate is studied in [19], where perfect channel information is available to both transmitter and receiver. Minimization of the outage probability using dynamic programming is pursued in [20]. In this section, we will develop loading algorithms for an OFDM system with imperfect CSIT in order to maximize system throughput given a target outage probability. Note that, unlike [12] and [19], where outage is caused by channel-induced fading, here the outage is caused by overly aggressive rate-power allocation due to CSIT errors. We will study two approaches: 1) independent loading for different subcarriers and 2) joint loading across all subcarriers.

A. Independent Loading

We will first focus on the data rate supportable by one subchannel, namely the one corresponding to subcarrier i. Assume that the power assigned to this subcarrier is γ_i. Given the imperfect CSIT $\check{h}_f(i) \sim \mathcal{CN}\left(\hat{h}_{f,0}(i), \sigma_H^2\right)$, the mutual information $I_i = \log\left(1+\gamma_i|\check{h}_f(i)|^2/\sigma_w^2\right)$ is a random variable. Since the loading algorithm must assure that the rate on this subband satisfies $\Pr(I_i < R_i|\hat{h}_f(i) = \hat{h}_{f,0}(i)) \le P_{\text{out}}$, we need to express first the outage rate in terms of the partial CSIT $\hat{h}_{f,0}(i)$. For notational brevity, we will drop the subcarrier index i when this will not cause confusion.

Since $\check{h}_f \sim \mathcal{CN}\left(\hat{h}_{f,0}, \sigma_H^2\right)$, $|\check{h}_f|^2$ follows a noncentral chi-square distribution

$$p(|\check{h}_f|^2 = x) = \frac{1}{\sigma_H^2} e^{-\left((|\hat{h}_{f,0}|^2+x)/\sigma_H^2\right)} I_0\left(2\sqrt{\frac{|\hat{h}_{f,0}|^2 x}{\sigma_H^4}}\right) \quad (18)$$

where $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind. With a transformation of random variables, it is straightforward to obtain the pdf of $I = \log\left(1+\gamma|\check{h}_f|^2/\sigma_w^2\right)$ as

$$p(I|\hat{h}_f = \hat{h}_{f,0}) = b \cdot e^{I-a-b(e^I-1)} I_0\left(2\sqrt{ab(e^I-1)}\right) \quad (19)$$

where $a = |\hat{h}_{f,0}|^2/\sigma_H^2$ and $b = \sigma_w^2/\left(\gamma\sigma_H^2\right)$. Our objective is to find $R_{\text{out},0}$ such that

$$\int_0^{R_{\text{out},0}} p(I|\hat{h}_f = \hat{h}_{f,0}) dI = P_{\text{out}}. \quad (20)$$

If a is not very large, (20) can be computed using the following series expansion [21]:

$$\int_0^R p(I|\hat{h}_f = \hat{h}_{f,0}) dI = \int_0^R b \cdot e^{I-a-b(e^I-1)} I_0\left(2\sqrt{ab(e^I-1)}\right) dI$$

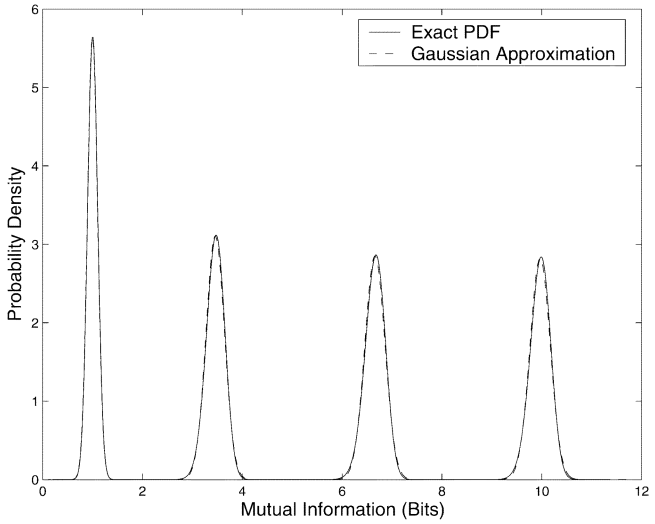

Fig. 5. PDF of $I(x, y|\hat{h}_f = \hat{h}_{f,0})$ $(a = 100)$.

$$= \int_0^{e^R-1} be^{-a-bx} \sum_{n=0}^{\infty} \frac{(abx)^n}{(n!)^2} dx$$
$$= e^{-a-b(e^R-1)} \sum_{k=0}^{K} \sum_{n=0}^{k-1} \frac{a^n}{n!} \cdot \frac{[b(e^R-1)]^k}{k!} + o\left(\frac{a^K}{K!}\right). \quad (21)$$

If a is large, then the series expansion in (21) will converge very slowly. However, we notice that for large a, $p(I|\hat{h}_f = \hat{h}_{f,0})$ in (19) can be well approximated by a Gaussian pdf with the same mean and covariance. In Fig. 5, we compare the exact $p(I|\hat{h}_f = \hat{h}_{f,0})$ with its Gaussian approximation for $a = 100$, and $b = 0.1$, 1, 10, and 100. We verify that the Gaussian approximation is accurate for a wide range of b values. This can be explained using the following arguments: When a is large, $p(I)$ is nonzero only in a small interval around $\log(1 + a/b)$. Let us set $x = I - \log(1 + a/b)$, and examine the behavior of $p(x)$ when $|x|$ is small

$$p(x) = b\left(1 + \frac{a}{b}\right) e^{x-a-t(x)} I_0\left(2\sqrt{a \cdot t(x)}\right)$$
$$\cong (a+b) e^{x-\left(\sqrt{a}-\sqrt{t(x)}\right)^2} \frac{1}{\sqrt{4\pi}} (a \cdot t(x))^{-(1/4)}$$
$$\cong \frac{a+b}{\sqrt{4\pi a}} e^{x-(a/4)(1+(b/a))^2 x^2} \quad (22)$$

where $t(x) = (b + a)\, e^x - b$. The next lemma follows readily from (22).

Lemma 1: When a is large, the probability distribution of the mutual information I per subcarrier can be approximated by the Gaussian distribution

$$I \sim \mathcal{N}\left(\log\left(1 + \frac{a}{b}\right), \frac{2a}{(a+b)^2}\right). \quad (23)$$

From the definition of $R_{\text{out},0}$ in (20) and Lemma 1, we obtain

$$R_{\text{out},0} \cong \log\left(1 + \frac{a}{b}\right) - \frac{\sqrt{2a}}{a+b} Q^{-1}(P_{\text{out}}) \quad (24)$$

where $Q(\cdot)$ is the complementary Gaussian cumulative distribution function.

To gain insight on how partial CSIT affects the average rate on subcarrier i, and also to check how accurate the approximation given by (24) is, let us examine a Rayleigh flat-fading channel. Supposing that the channel estimation error for the ith subcarrier is $\tilde{h}_f(i) = h_f(i) - \hat{h}_f(i) \sim \mathcal{CN}\left(0, \sigma_H^2\right)$, and it is independent of $\hat{h}_f(i)$ (which is satisfied when using the MMSE estimator), then the channel estimator satisfies $\hat{h}_f(i) \sim \mathcal{CN}\left(0, 1 - \sigma_H^2\right)$. The average throughput of subcarrier i given the outage probability P_{out} is

$$\bar{R}_{\text{out},i} = E_{\hat{h}_f(i)}[R_{\text{out},i}]$$
$$\cong E_{\hat{h}_f(i)}\left[\log\left(1 + \frac{a_i}{b_i}\right) - \frac{\sqrt{2a_i}}{a_i + b_i} Q^{-1}(P_{\text{out}})\right]$$
$$= \int_0^{\infty} \left[\log\left(1 + \frac{\gamma_i y}{\sigma_w^2}\right) - \frac{\sqrt{2\sigma_H^2 y}}{y + \frac{\sigma_w^2}{\gamma_i}} Q^{-1}(P_{\text{out}})\right]$$
$$\times \frac{1}{1-\sigma_H^2} e^{-\left(y/\left(1-\sigma_H^2\right)\right)} dy$$
$$= e^{\left(\sigma_w^2/\left(\gamma_i\left(1-\sigma_H^2\right)\right)\right)} Ei\left(\frac{\sigma_w^2}{\gamma_i(1-\sigma_H^2)}\right)$$
$$- \left[\sqrt{\frac{2\pi\sigma_H^2}{1-\sigma_H^2}} - \frac{2\pi\sqrt{2\sigma_H^2\sigma_w^2}}{\sqrt{\gamma_i}(1-\sigma_H^2)} e^{\left(\sigma_w^2/\left(\gamma_i\left(1-\sigma_H^2\right)\right)\right)}\right.$$
$$\left. \times Q\left(\sqrt{\frac{2\sigma_w^2}{\gamma_i\left(1-\sigma_H^2\right)}}\right)\right] Q^{-1}(P_{\text{out}}) \quad (25)$$

where $Ei(\cdot)$ is the exponential-integral function.

In Fig. 6, we plot the dependence of the average data rate $\bar{R}_{\text{out},i}$ on the partial channel knowledge versus SNR. We observe that the approximation in (25) fits the simulation results accurately except when the channel information error is large, which corresponds to the case that $a_i = |\hat{h}_{f,0}(i)|^2/\sigma_H^2$ is small.

In an OFDM system with loading performed for each subcarrier independently, the average system throughput is the mean of all individual subcarriers' throughput. When the CSIT uncertainty is small, it follows from (24) that the outage throughput averaged across subcarriers, $\bar{R}_{\text{out}} = \sum_{i=0}^{N-1} R_i/N$, can be approximated by

$$\bar{R}_{\text{out}} \cong \frac{1}{N} \sum_{i=0}^{N-1} \left[\log\left(1 + \frac{a_i}{b_i}\right) - \frac{\sqrt{2a_i}}{a_i + b_i} Q^{-1}(P_{\text{out}})\right] \quad (26)$$

where $a_i = |\hat{h}_{f,0}(i)|^2/\sigma_H^2$, and $b_i = \sigma_w^2/\left(\gamma_i \sigma_H^2\right)$. Note that, in an OFDM system, it is possible that for some subcarriers the a_i's are relatively small, even when σ_H^2 is small. For these subcarriers, the outage rate can be computed using (21). Since these subcarriers have small channel gains, and their contribution to the total system throughput is insignificant unless the SNR is high, we assume for simplicity that a_i is sufficiently large for all i. Simulation results verify that this assumption is reasonable.

Finding the power allocation $\boldsymbol{\gamma} = \{\gamma_i\}_{i=0}^{N-1}$ maximizing $\bar{R}_{\text{out}}$ is nontrivial, because $\bar{R}_{\text{out}}$ is, in general, not a concave function of $\boldsymbol{\gamma}$. In the Appendix, we show that the number of local maxima is finite, and develop an algorithm that goes through these local

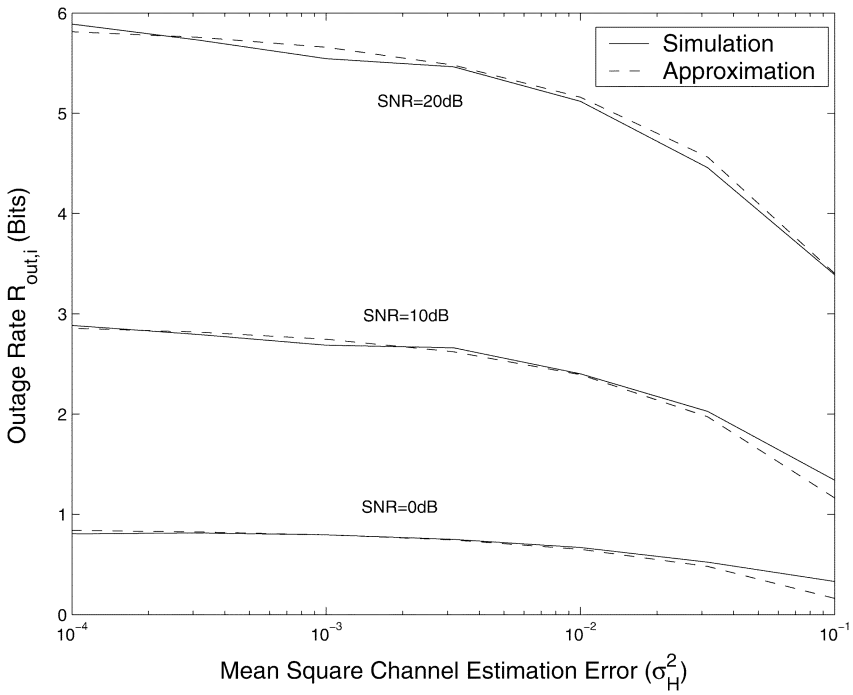

Fig. 6. Impact of σ_H^2 on $\bar{R}_{\text{out},i}$ ($P_{\text{out}} = 0.01$).

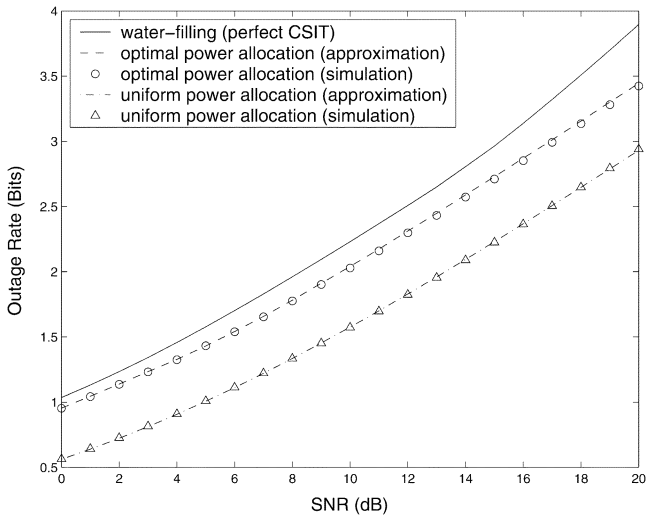

Fig. 7. Performance comparison of different power loading schemes ($N = 16$, $L = 1$, $\sigma_h^2 = 0.01$, and $P_{\text{out}} = 0.01$).

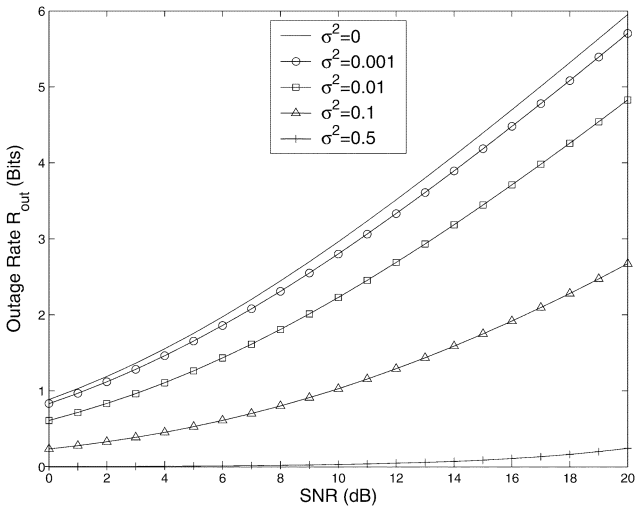

Fig. 8. $\bar{R}_{\text{out}}$ versus SNR ($N = 16$, $L = 1$, $P_{\text{out}} = 10^{-3}$).

maxima systematically to find the global maximum. In most applications, however, close to optimal rate can be achieved using the following suboptimal power allocation scheme:

$$\gamma_i = \begin{cases} \left[\frac{2\sqrt{2}\sigma_H\sigma_w^2 Q^{-1}(P_{\text{out}})}{|\hat{h}_{f,0}(i)|^3(1-\sqrt{\beta_i(\xi)})} - \frac{\sigma_w^2}{|\hat{h}_{f,0}(i)|^2}\right]^+, & \beta_i(\xi) \geq 0 \\ 0, & \beta_i(\xi) < 0, \end{cases} \tag{27}$$

where $\beta_i(\xi) = 1 - 4\sqrt{2}\xi\sigma_H\sigma_w^2 Q^{-1}(P_{\text{out}})/|\hat{h}_{f,0}(i)|^3$, and ξ is determined by the power constraint. Similar to (13), for a given P_{out}, the rate-maximizing γ_i's in (27) are obtained using a 1-D search.

In Fig. 7, we compare $\bar{R}_{\text{out}}$ achieved by the optimal power allocation and the uniform power allocation. We adopt the setting of Example 1 in Section III, and set the target $P_{\text{out}} = 0.01$ and $\sigma_h^2 = 0.01$. We observe that optimal power allocation brings significant gain over uniform power allocation. For comparison, we also plot the approximation given by (26). Again, we verify that this approximation is accurate.

The average system throughput $\bar{R}_{\text{out}}$ versus SNR under different CSIT errors σ_h^2 is plotted in Figs. 8 and 9 for $P_{\text{out}} = 10^{-3}$ and $P_{\text{out}} = 0.1$, respectively. We simulate a two-tap channel with *i.i.d.* channel coefficients distributed as $\mathcal{CN}(0, 1/2)$. We observe that, unlike the expected mutual information, the system throughput under the outage probability constraint degrades quickly as the CSIT error grows large. This is not unexpected, since the outage behavior of a system depends on the tail of the channel pdf, which is sensitive to the CSIT error covariance.

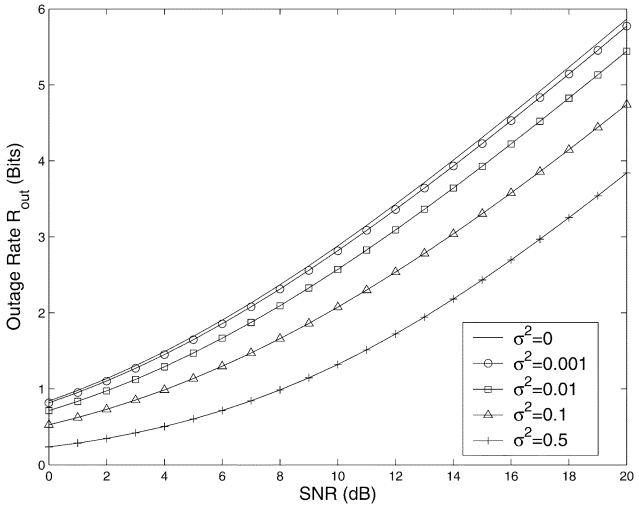

Fig. 9. $\bar{R}_{\text{out}}$ versus SNR ($N = 16$, $L = 1$, and $P_{\text{out}} = 0.1$).

B. Joint Loading Across Subcarriers

1) Outage Rate Under Joint Loading: Performing joint loading across all subcarriers brings gain over individual loading by taking advantage of the multipath diversity. We assume that given the mean channel feedback $\hat{\boldsymbol{h}}_{f,0}$, the partial CSIT is distributed according to $\check{\boldsymbol{h}}_f \sim \mathcal{CN}(\hat{\boldsymbol{h}}_{f,0}, \boldsymbol{\Sigma}_{h_f})$. The first step of the joint loading algorithm is to find $\bar{R}_{\text{out}}$ that satisfies

$$\frac{1}{\det(\pi\boldsymbol{\Sigma}_{h_f})}\int \exp[-(\check{\boldsymbol{h}}_f-\hat{\boldsymbol{h}}_{f,0})^{\mathcal{H}}\boldsymbol{\Sigma}_{h_f}^{-1}(\check{\boldsymbol{h}}_f-\hat{\boldsymbol{h}}_{f,0})] \times 1_{\left\{(1/N)\sum_{i=0}^{N-1}\log\left(1+\left((\gamma_i|\check{h}_f(i)|^2)/\sigma_w^2\right)\right)<\bar{R}_{\text{out}}\right\}} \times d\check{\boldsymbol{h}}_f = P_{\text{out}} \tag{28}$$

where $1_{\{\cdot\}}$ is the indicator function. In general, there is no closed-form solution to this problem, and we have to rely on Monte-Carlo simulation. When σ_H^2 is small, however, we can make use of the Gaussian approximation as follows. For a given feedback $\hat{\boldsymbol{h}}_{f,0}$, the mutual information for the ith subcarrier is Gaussian with $I_i \sim \mathcal{N}(\log(1+(a_i/b_i)),(2a_i/(a_i+b_i)^2))$, where $a_i = |\hat{h}_{f,0}(i)|^2/\sigma_H^2$ and $b_i = \sigma_w^2/\left(\gamma_i\sigma_H^2\right)$. Note that we have assumed $\Sigma_{ii} = \sigma_H^2$ for all i. Approximating the sum of I_i's with a Gaussian random variable, we have, $\sum_{i=0}^{N-1} I_i \sim \mathcal{N}\left(\mu_I, \sigma_I^2\right)$, where $\mu_I = \sum_{i=0}^{N-1}\log(1+a_i/b_i)$, $\sigma_I^2 = \sum_{i=0}^{N-1}\text{var}(I_i) + 2\sum_{i<j}\text{cov}(I_i,I_j)$.

Having obtained that $\text{var}(I_i) \cong 2a_i/(a_i+b_i)^2$, let us now examine the $\text{cov}(I_i,I_j)$. With a slight abuse of notation, we let $\check{\boldsymbol{h}}_{\text{ij}} = [\check{h}_f(i)\check{h}_f(j)]^T$, $\hat{\boldsymbol{h}}_{\text{ij}} = [\hat{h}_f(i)\hat{h}_f(j)]^T$, and $\boldsymbol{\Sigma}_{\text{ij}} = E[(\check{\boldsymbol{h}}_{\text{ij}} - \hat{\boldsymbol{h}}_{\text{ij}})(\check{\boldsymbol{h}}_{\text{ij}} - \hat{\boldsymbol{h}}_{\text{ij}})^{\mathcal{H}}]$, and express $\text{cov}(I_i,I_j)$ as

$$\text{cov}(I_i,I_j) = \int \frac{1}{\det(\pi\boldsymbol{\Sigma}_{\text{ij}})}\exp\left[-\boldsymbol{x}_{\text{ij}}^{\mathcal{H}}\boldsymbol{\Sigma}_{\text{ij}}^{-1}\boldsymbol{x}_{\text{ij}}\right] \times \log\left(\frac{1+\frac{\gamma_i|\check{h}_f(i)|^2}{\sigma_w^2}}{1+\frac{\gamma_i|\hat{h}_f(i)|^2}{\sigma_w^2}}\right)\log\left(\frac{1+\frac{\gamma_j|\check{h}_f(j)|^2}{\sigma_w^2}}{1+\frac{\gamma_j|\hat{h}_f(j)|^2}{\sigma_w^2}}\right)d\check{\boldsymbol{h}}_{\text{ij}} \tag{29}$$

where $\boldsymbol{x}_{\text{ij}} = \check{\boldsymbol{h}}_{\text{ij}} - \hat{\boldsymbol{h}}_{\text{ij}}$. For large (a_i,a_j), we have

$$\begin{aligned}\text{cov}(I_i,I_j) &\cong \int \frac{\exp\left(-\boldsymbol{x}_{\text{ij}}^H\boldsymbol{\Sigma}_{\text{ij}}^{-1}\boldsymbol{x}_{\text{ij}}\right)}{\det(\pi\boldsymbol{\Sigma}_{\text{ij}})} \times \frac{4\gamma_i\gamma_j\text{Re}(\hat{h}_f(i)^*x_i)\text{Re}(\hat{h}_f(j)^*x_j)}{\left(\sigma_w^2+\gamma_i|\hat{h}_f(i)|^2\right)\left(\sigma_w^2+\gamma_j|\hat{h}_f(j)|^2\right)}d\boldsymbol{x}_{\text{ij}} \\ &= \frac{2\gamma_i\gamma_j\sigma_H^2\text{Re}(\hat{h}_f(i)\hat{h}_f(j)^*\rho_{\text{ij}})}{\left(\sigma_w^2+\gamma_i|\hat{h}_f(i)|^2\right)\left(\sigma_w^2+\gamma_j|\hat{h}_f(j)|^2\right)} \\ &= \frac{2\kappa_{\text{ij}}\sqrt{a_ia_j}}{(a_i+b_i)(a_j+b_j)}\end{aligned} \tag{30}$$

where $\rho_{\text{ij}} = \left(1/\sigma_H^2\right)E[(\check{h}_f(i)-\hat{h}_f(i))^*(\check{h}_f(j)-\hat{h}_f(j))]$, and $\kappa_{\text{ij}} = \text{Re}((\rho_{\text{ij}}\hat{h}_f(i)\hat{h}_f(j)^*)/(|\hat{h}_f(i)\hat{h}_(j)|))$. So, the outage throughput averaged across subcarriers is

$$\bar{R}_{\text{out}} \cong \frac{1}{N}\sum_{i=0}^{N-1}\log\left(1+\frac{a_i}{b_i}\right) - \frac{1}{N}Q^{-1}(P_{\text{out}}) \times \left[\sum_{i=0}^{N-1}\frac{2a_i}{(a_i+b_i)^2} + \sum_{i<j}\frac{4\kappa_{\text{ij}}\sqrt{a_ia_j}}{(a_i+b_i)(a_j+b_j)}\right]^{(1/2)}. \tag{31}$$

Optimal power allocation $\{\gamma_i\}_{i=0}^{N-1}$ which maximizes (31) can be found using numerical search.

2) Effects of Multipath: The number of resolvable paths $L+1$ affects the outage rate of an OFDM system in three ways: 1) the channel estimation error σ_h^2 will increase with L, which is confirmed by (6); 2) larger delay spread will require a longer cyclic prefix; and 3) we can see from (31) that given σ_h^2, the $\bar{R}_{\text{out}}$ depends on κ_{ij}, which, in turn, depends on the covariance matrix of the partial CSIT $\boldsymbol{\Sigma}_{h_f} = E\left[\tilde{\boldsymbol{h}}_f\tilde{\boldsymbol{h}}_f^{\mathcal{H}}\right]$. Intuitively speaking, given σ_h^2, multipath introduces some degree of independence among different subcarriers and, hence, increases $\bar{R}_{\text{out}}$. In the following, we will specialize our joint loading algorithm for a Rayleigh-fading channel, and investigate the effects of multipath on κ_{ij}. Specifically, we will use the example of Section II-A.

Using $\boldsymbol{\Sigma}_{h_f}$ in (5), and denoting the ith row of $\mathbf{W}_L$ as $\boldsymbol{w}_i^T$, we can write κ_{ij} as

$$\kappa_{\text{ij}} = \text{Re}\left(\frac{N}{L+1}\boldsymbol{w}_i^T\boldsymbol{w}_j^*e^{j(\theta_i-\theta_j)}\right) \tag{32}$$

where θ_i is the phase of $\hat{h}_f(i)$. If there is no multipath, i.e., $L=0$, then $\kappa_{\text{ij}}=1$, (31) reduces to (26), and there is no gain in performing joint loading. If $L=N-1$, then $\kappa_{\text{ij}} = \delta_{\text{ij}}$. For $0<L<N-1$, the value of κ_{ij} depends on the realization of $\hat{\boldsymbol{h}}$. Usually, large $|\kappa_{\text{ij}}|$ occurs when subcarriers i and j are neighbors; i.e., $(i-j) \bmod N$ is small. For these neighboring subcarriers, $\text{Re}\left(\boldsymbol{w}_i^T\boldsymbol{w}_j^*\right) > 0$ and $\theta_i - \theta_j$ are usually small, so we would expect $\kappa_{\text{ij}} > 0$. When L gets larger, $\boldsymbol{w}_i^T\boldsymbol{w}_j^*$ comes closer to 0 for $i \neq j$. So, in general, we would expect the gain of joint loading to increase as L increases.

Summarizing, the increase of L will cause the increase of $\bar{R}_{\text{out}}$ given σ_h^2, but it will also lead to a larger σ_h^2 and a longer

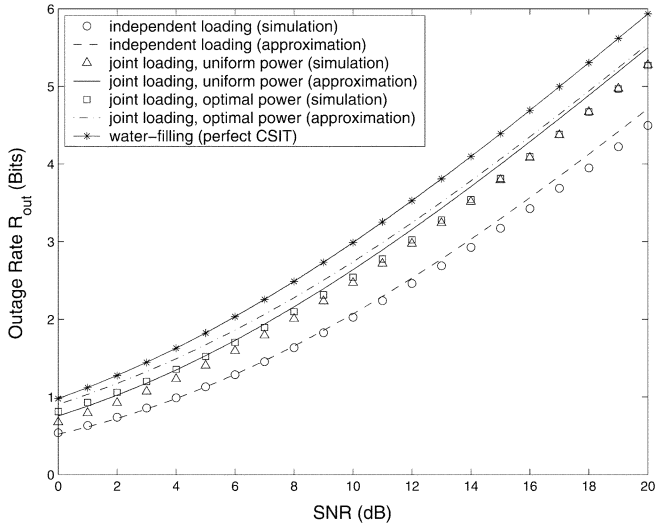

Fig. 10. Joint loading versus independent loading ($N = 16$, $L = 3$, $\sigma_h^2 = 0.01$, and $P_{\text{out}} = 10^{-4}$).

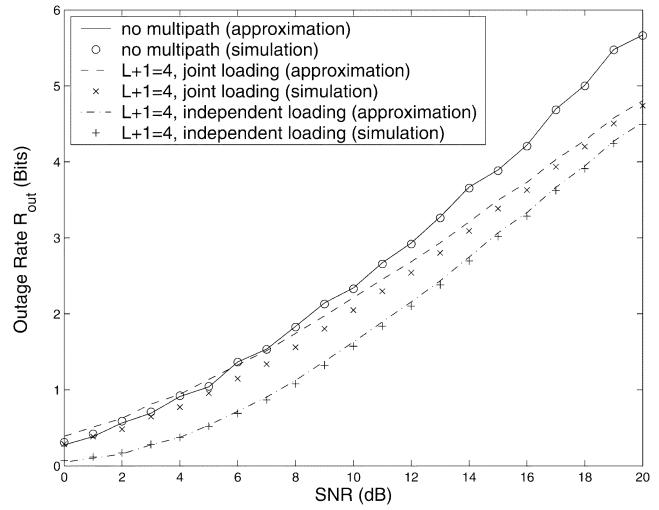

Fig. 11. Effects of channel multipath ($N = 16$, $P_{\text{out}} = 10^{-3}$).

cyclic prefix. The combined effects will be investigated through simulation.

3) Numerical Results: We simulate an OFDM transmission through a Rayleigh-fading channel of order $L = 3$. The CSIT has $\sigma_H^2 = 0.01$, and the target outage probability is $P_{\text{out}} = 10^{-4}$. The outage throughput achieved by independent loading, and by joint loading with uniform and optimal power allocation are all plotted in Fig. 10. We observe that joint loading brings about 2-dB gain over independent loading. We can also confirm that the approximation given by (31) fits the simulation results quite well. In this setting, optimal power allocation does not bring significant gain for the joint loading algorithm over uniform power allocation.

To see how multipath affects throughput, we simulate the outage rates ($\bar{R}_{\text{out}}$) for channels with no multipath and with $L+1 = 4$ paths, respectively. The channel is estimated from one training block (for simplicity, we assume no feedback delay). The target outage probability is set to 10^{-3}. The simulation results are plotted in Fig. 11. We observe that the rate of independent loading suffers severe degradation when there is multipath. Joint loading is able to largely offset this rate loss at low to moderate SNR, although it still suffers considerable loss at high SNR, mainly due to the use of a longer cyclic prefix.

V. Conclusion

We studied the fundamental limit imposed by imperfect channel feedback on the throughput of OFDM systems. Both ergodic and quasi-static channel uncertainty models were investigated. For the ergodic model, we derived an optimal power loading scheme, and compared it with uniform power allocation and water-filling. For quasi-static channels, we developed two different rate allocation schemes: independent loading and joint loading across subcarriers. An efficient optimal power allocation algorithm was developed for the independent loading scheme. We demonstrated that in flat-fading channels, independent loading and joint loading are equivalent, while in multipath channels, joint loading brings substantial gains to outage rates over independent loading.

In this paper, we derived optimal *power loading* algorithms for OFDM based on average and outage capacity criteria. For optimal *bit loading* schemes maximizing rate in adaptive SISO and MIMO OFDM systems, the reader is referred to [9]–[11], where partial CSIT based algorithms are derived to meet a target BER under a given transmit-power budget.

Appendix

In this appendix, we find the γ_i's that maximize $R(\boldsymbol{\gamma}) = \sum_{i=0}^{N-1}[\log(1 + a_i/b_i) - \sqrt{2a_i}/(a_i + b_i)Q^{-1}(P_{\text{out}})]$, subject to the power constraint. Alternatively, we can state the problem as follows:

$$\text{minimize} \quad C(\boldsymbol{\gamma}) = -R(\boldsymbol{\gamma}) \tag{33}$$

$$\text{subject to} \quad -\gamma_i \le 0, \quad i = 0, \ldots, N-1 \tag{34}$$

$$\sum_{i=0}^{N-1} \gamma_i = NP_0. \tag{35}$$

We first examine the convexity of the objective function. Taking first-order derivative of $C(\boldsymbol{\gamma})$, we have

$$\frac{\partial C}{\partial \gamma_i} = -\frac{|\hat{h}_f(i)|^2}{\sigma_w^2 + \gamma_i|\hat{h}_f(i)|^2} + \frac{\sqrt{2}|\hat{h}_f(i)|\sigma_H\sigma_w^2 Q^{-1}(P_{\text{out}})}{\left(\sigma_w^2 + \gamma_i|\hat{h}_f(i)|^2\right)^2} \quad i = 0, \ldots, N-1. \tag{36}$$

It is easy to see that the Hessian of $C(\boldsymbol{\gamma})$ is a diagonal matrix with diagonal elements

$$\frac{\partial^2 C}{\partial \gamma_i^2} = \frac{|\hat{h}_f(i)|^4}{\left(\sigma_w^2 + \gamma_i|\hat{h}_f(i)|^2\right)^2} \times \left[1 - \frac{2\sqrt{2}\sigma_H\sigma_w^2 Q^{-1}(P_{\text{out}})}{|\hat{h}_f(i)|\left(\sigma_w^2 + \gamma_i|\hat{h}_f(i)|^2\right)}\right]. \tag{37}$$

So, $C(\boldsymbol{\gamma})$ is a convex function if and only if $a_i \ge 8[Q^{-1}(P_{\text{out}})]^2$, for all i. For now, let us assume that these

conditions are satisfied, so the Karush–Kuhn–Tucker (KKT) conditions are sufficient conditions for a global minimum [22]

$$\frac{\partial C}{\partial \gamma_i} + \xi - \mu_i = 0 \tag{38}$$

$$\mu_i \geq 0 \tag{39}$$

$$\mu_i \gamma_i = 0, \qquad i = 0, \ldots, N-1. \tag{40}$$

Combining (36) with (38), we obtain

$$\begin{aligned}\mu_i &= -\frac{|\hat{h}_f(i)|^2}{\sigma_w^2 + \gamma_i |\hat{h}_f(i)|^2} + \frac{\sqrt{2} Q^{-1}(P_{\text{out}}) |\hat{h}_f(i)| \sigma_H \sigma_w^2}{(\sigma_w^2 + \gamma_i |\hat{h}_f(i)|^2)^2} + \xi \\ &= g_i(\gamma_i, \xi).\end{aligned} \tag{41}$$

If no feasible γ_i satisfies (41), then from (40), we have $\gamma_i = 0$. It is easy to verify that $\mu_i > 0$, in this case. Otherwise, setting $\mu_i = 0$, there are two possible solutions to (41)

$$\gamma_{i-} = \frac{2\sqrt{2}\sigma_H \sigma_w^2 Q^{-1}(P_{\text{out}})}{|\hat{h}_f(i)|^3 (1 - \sqrt{\beta_i(\xi)})} - \frac{\sigma_w^2}{|\hat{h}_f(i)|^2} \tag{42}$$

$$\gamma_{i+} = \frac{2\sqrt{2}\sigma_H \sigma_w^2 Q^{-1}(P_{\text{out}})}{|\hat{h}_f(i)|^3 (1 + \sqrt{\beta_i(\xi)})} - \frac{\sigma_w^2}{|\hat{h}_f(i)|^2} \tag{43}$$

where $\beta_i(\xi) = 1 - 4\sqrt{2}\xi\sigma_H \sigma_w^2 Q^{-1}(P_{\text{out}})/|\hat{h}_f(i)|^3$. Since

$$\frac{\frac{\partial^2 C}{\partial \gamma_i^2}\big|_{\gamma_i = \gamma_{i-}}}{\frac{|\hat{h}_f(i)|^4}{\left(\sigma_w^2 + \gamma_{i-}|\hat{h_f}(i)|^2\right)^2}} = -\frac{\frac{\partial^2 C}{\partial \gamma_i^2}\big|_{\gamma_i = \gamma_{i+}}}{\frac{|\hat{h}_f(i)|^4}{\left(\sigma_w^2 + \gamma_{i+}|\hat{h_f}(i)|^2\right)^2}} = \sqrt{\beta_i(\xi)}. \tag{44}$$

$\gamma_i = \gamma_{i-}$ is the desirable solution. Hence, the optimal power allocation is given by (27).

When the convexity conditions are not satisfied, exhaustive search is usually needed to find the global minimum. However, we will show next that a lower complexity alternative is possible. Without loss of generality, let us assume that

$$|\hat{h}_f(0)| \geq |\hat{h}_f(1)| \geq \cdots \geq |\hat{h}_f(N-1)|. \tag{45}$$

It is easy to verify that the constraints in (33) satisfy the Mangasarian-Fromovitz constraint qualification [22]. So, every local minimum must satisfy the KKT conditions. In addition, a local minimum $\boldsymbol{\gamma}^*$ must satisfy

$$\left.\frac{\partial^2 C}{\partial \gamma_i^2}\right|_{\gamma_i = \gamma_i^*} + \left.\frac{\partial^2 C}{\partial \gamma_j^2}\right|_{\gamma_j = \gamma_j^*} \geq 0, \quad \forall\, i \neq j,\ \gamma_i^*, \gamma_j^* \neq 0. \tag{46}$$

Otherwise, given a small perturbation of the form $\delta\boldsymbol{\gamma} = [0, \ldots, 0, \delta\gamma_i, 0, \ldots, 0, \delta\gamma_j, 0, \ldots, 0]^T$, where $\delta\gamma_i = -\delta\gamma_j$, we have that the constraints (34) and (35) are satisfied, and

$$\begin{aligned}C(\boldsymbol{\gamma}^* + \delta\boldsymbol{\gamma}) \cong C(\boldsymbol{\gamma}^*) &+ \left.\frac{\partial^2 C}{\partial \gamma_i^2}\right|_{\gamma_i = \gamma_i^*} (\delta\gamma_i)^2 \\ &+ \left.\frac{\partial^2 C}{\partial \gamma_j^2}\right|_{\gamma_j = \gamma_j^*} (\delta\gamma_j)^2 < C(\boldsymbol{\gamma}^*).\end{aligned} \tag{47}$$

Using (42)–(46), we can show that $\xi \geq 0$ and that if $\gamma_i^* = \gamma_{i+}$ for some i, then

$$\gamma_j^* = \begin{cases} \gamma_{j-}, & |\hat{h}_f(j)| > |\hat{h}_f(i)| \\ \gamma_{j-} \text{ or } 0, & |\hat{h}_f(j)| = |\hat{h}_f(i)| \\ 0, & |\hat{h}_f(j)| < |\hat{h}_f(i)|. \end{cases} \tag{48}$$

Therefore, a local minimum must have one of the following two forms:

$$\boldsymbol{\gamma}_-^*(\xi_-^*) = [\gamma_{0-}^*, \ldots, \gamma_{K-}^*, \gamma_{(K+1)-}^*, 0, \ldots, 0]^T \tag{49}$$

$$\boldsymbol{\gamma}_+^*(\xi_+^*) = [\gamma_{0-}^*, \ldots, \gamma_{K-}^*, \gamma_{(K+1)+}^*, 0, \ldots, 0]^T. \tag{50}$$

So, the difference between local minima lies in the number of zero elements and the choice of the last nonzero element. Note that for $\xi \geq 0$, γ_{i-} in (42) is a strictly monotonically decreasing function of ξ, while γ_{i+} in (43) is a strictly monotonically increasing function of ξ. Once the number of zero elements is fixed, ξ_-^* and ξ_+^* are determined by the power constraint. Suppose that we have found a local minimum $\boldsymbol{\gamma}_-(\xi_-)$, and by forcing $\gamma'_{(K+1)-} = 0$, we can obtain another local minimum point $\boldsymbol{\gamma}'_-(\xi'_-)$. Obviously, $\xi'_- < \xi_-$ because the other subcarriers will now be allocated more power. So

$$g_{K+1}(0, \xi_-) > g_{K+1}(0, \xi'_-) = \mu'_{K+1} \geq 0 \tag{51}$$

meaning that

$$\xi_- > \frac{|\hat{h}_f(K+1)|^2}{\sigma_w^2} - \frac{\sqrt{2} Q^{-1}(P_{\text{out}}) |\hat{h}_f(K+1)| \sigma_H}{\sigma_w^2}. \tag{52}$$

Using (52), we can show that

$$\begin{aligned}&\sqrt{\beta_{K+1}(\xi_-)} \\ &\quad < \frac{1}{|\hat{h}_f(K+1)|} \left| |\hat{h}_f(K+1)| - 2\sqrt{2}\sigma_H Q^{-1}(P_{\text{out}}) \right|.\end{aligned} \tag{53}$$

Since

$$\begin{aligned}0 &\leq \frac{|\hat{h}_f(K+1)|^2}{\sigma_w^2} \gamma_{(K+1)-}(\xi_-) \\ &< \frac{2\sqrt{2}\sigma_H^2 Q^{-1}(P_{\text{out}})}{|\hat{h}_f(K+1)| - \left| |\hat{h}_f(K+1)| - 2\sqrt{2}\sigma_H Q^{-1}(P_{\text{out}}) \right|} - 1\end{aligned} \tag{54}$$

we have $|\hat{h}_f(K+1)| < 2\sqrt{2}\sigma_H Q^{-1}(P_{\text{out}})$, and, hence

$$\begin{aligned}&\frac{|\hat{h}_f(K+1)|^2}{\sigma_w^2} \gamma_{(K+1)+}(\xi_-) \\ &\quad \geq \frac{2\sqrt{2}\sigma_H^2 Q^{-1}(P_{\text{out}})}{|\hat{h}_f(K+1)| + \left| |\hat{h}_f(K+1)| - 2\sqrt{2}\sigma_H Q^{-1}(P_{\text{out}}) \right|} - 1 \\ &\quad = 0.\end{aligned} \tag{55}$$

Summarizing, the global minimum can be found as follows:

1) compute the KKT point $\boldsymbol{\gamma}_-(\xi_-)$ given by **(27)**, and the other possible KKT point

$\boldsymbol{\gamma}_+(\xi_+)$ with the same number of nonzero elements;
2) if $C(\boldsymbol{\gamma}_-) \leq C(\boldsymbol{\gamma}_+)$, then $\boldsymbol{\gamma}^* = \boldsymbol{\gamma}_-$; otherwise, $\boldsymbol{\gamma}^* = \boldsymbol{\gamma}_+$.
3) find i such that $\gamma_{i-}(\xi_-) \neq 0$, and $\hat{h}_f(i)$ is the smallest among all subcarriers with nonzero power;
4) if $\gamma_{i+}(\xi_-) \leq 0$, terminate the search; otherwise, compute $\boldsymbol{\gamma}'_-(\xi'_-)$ and $\boldsymbol{\gamma}'_+(\xi'_+)$ by forcing $\gamma'_i = 0$.
5) set $\boldsymbol{\gamma}_- = \boldsymbol{\gamma}'_-$, $\boldsymbol{\gamma}_+ = \boldsymbol{\gamma}'_+$, $\xi_- = \xi'_-$, and $\xi_+ = \xi'_+$; go to step 2.

In our simulations, we find that this algorithm usually terminates in a few iterations. Also, the suboptimal power loading scheme given by (27) achieves close to optimal rate in most scenarios.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers, whose comments helped improve the quality of this paper.

REFERENCES

[1] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Filterbank transceivers optimizing information rate in block transmissions over dispersive channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 1019–1032, Apr. 1999.
[2] A. Moustakas and S. Simon. (2002) Optimizing multi-transmitter single-receiver (MISO) Antenna systems with partial channel knowledge. Bell Labs. Tech. Memo. [Online]. Available: http://mars.bell-labs.com
[3] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct. 1998.
[4] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.
[5] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel mean feedback," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2599–2613, Oct. 2002.
[6] A. Leke and J. M. Cioffi, "Impact of imperfect channel knowledge on the performance of multicarrier systems," in *Proc. IEEE GLOBECOM*, vol. 2, Sydney, Australia, 1998, pp. 951–955.
[7] M. R. Souryal and R. L. Pickholtz, "Adaptive modulation with imperfect channel information in OFDM," in *Proc. Int. Conf. Communications*, vol. 6, Helsinki, Finland, Jun. 2001, pp. 1861–1865.
[8] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.
[9] Z. Song, K. Zhang, and Y. Guan, "Statistical adaptive modulation for QAM-OFDM systems," in *Proc. GLOBECOM*, vol. 1, Taipei, Taiwan, R.O.C., Nov. 2002, pp. 706–710.
[10] P. Xia, S. Zhou, and G. B. Giannakis, "MIMO-OFDM with ST coding and beamforming adapted to partial CSI," in *Proc. 37th Conf. Information Sciences Systems*, Baltimore, MD, Mar. 2003.
[11] S. Ye, R. Blum, and L. Cimini, "Adaptive modulation for variable rate OFDM systems with imperfect channel information," in *Proc. IEEE Vehicular Technology Conf.*, vol. 2, Birmingham, AL, May 2002, pp. 767–771.
[12] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information-theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359–378, May 1994.
[13] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.
[14] I. E. Telatar, "Capacity of multi-antenna Gaussian channel," AT&T Bell Labs. Tech. Memo, 1995.
[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[16] W. Hirt and J. L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 34, no. 3, pp. 380–388, May 1988.
[17] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, Sep. 1999.
[18] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.
[19] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1468–1489, Jul. 1999.
[20] R. Negi and J. M. Cioffi, "Delay-constrained capacity with causal feedback," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2478–2494, Sep. 2002.
[21] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6th ed. San Diego, CA: Academic, 2000.
[22] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.

Yingwei Yao (M'03) received the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 2002.

He was a Postdoctoral Researcher with the University of Minnesota, Minneapolis, from 2002 to 2004. Currently, he is an Assistant Professor of electrical and computer engineering with the University of Illinois, Chicago. His research interests include statistical signal processing, wireless communication theory, and wireless networks.

Georgios B. Giannakis (S'84–M'86–SM'91–F'97) received the Diploma in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1981, and the M.Sc. degree in electrical engineering, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering, all from the University of Southern California (USC), Los Angeles, in 1983 and 1986, respectively.

After lecturing for one year at USC, he joined the University of Virginia, Charlottesville, in 1987, where he became a Professor of electrical engineering in 1997. Since 1999, he has been a Professor with the Department of Electrical and Computer Engineering, 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 200 journal papers, 350 conference papers, and two edited books. His current research interests include transmitter and receiver diversity techniques for single- and multiuser fading communication channels, complex-field and space-time coding, multicarrier, ultrawide-band wireless communication systems, cross-layer designs, and distributed sensor networks.

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