Wideband Fading Channel Capacity with Training and Partial Feedback

Manish Agarwal, Michael L. Honig
EECS Department, Northwestern University
2145 Sheridan Road, Evanston, IL 60208 USA
{m-agarwal,mh}@northwestern.edu

Abstract

We consider the capacity of a wideband fading channel with partial feedback, subject to an average power constraint. A doubly block Rayleigh fading model is assumed with finite coherence time (L channel uses) and a large number of independent, finite bandwidth coherence bands (sub-channels). Without feedback, it is known that uniformly spreading the signal power beyond a critical number of coherence bands decreases the capacity. Here we assume that the transmitter *probes* a subset of sub-channels during each coherence time by transmitting pilot sequences for channel estimation. For each sub-channel probed, one bit of feedback indicates whether or not the transmitter should continue to transmit information symbols. Our problem is to optimize jointly the training (both length and power), number of sub-channels probed (probing bandwidth), and feedback threshold to maximize the achievable rate (lower bound on ergodic capacity) taking into account the sub-channel estimation error. Optimizing the probing bandwidth balances diversity against the quality of sub-channel estimate. We show that the achievable rate increases as $S \log L$, where S is the Signal-to-Noise Ratio, and exceeds the capacity with impulsive signaling (given by S) when L exceeds a (positive) threshold value. Moreover, the optimal probing bandwidth scales as $S\frac{L}{\log^2 L}$. In contrast, without feedback the optimal probing bandwidth for the probing scheme scales as $SL^{1/3}$ and the achievable rate converges to S, where the gap diminishes as $SL^{-1/3}$.

Index Terms

ultra-wideband, channel probing, pilot symbols, one-bit feedback

This paper was presented in part at the 2005 Allerton Conference on Communication, Control and Computing. The work was supported by ARO under grant DAAD19-99-1-0288 and NSF under grant CCR-0310809.

I. Introduction

As the bandwidth of a fading channel increases, accurate channel estimation becomes a major challenge. This is due to the increasing number of coherence bands, which must be estimated in a finite coherence time. The associated increase in channel estimation error leads to a noncoherence loss in capacity, which increases as the transmitted signal is spread across wider bandwidths [1]–[3]. To avoid this noncoherence loss, it is necessary to use impulsive, or "flash" signaling [4], [5], which concentrates the signal power in time or frequency.

In this paper we consider communications over a wideband channel assuming the receiver can relay limited feedback about the channel to the transmitter. We assume a doubly-selective block fading channel model in which the channel is partitioned in time and frequency into coherence blocks, each of which is associated with a constant channel gain. Different coherence bands (across frequency) are referred to as *sub-channels*. The sub-channel gains are independent and Rayleigh distributed across coherence blocks, and are unknown at the receiver and transmitter at the beginning of each coherence block.

If both the transmitter and receiver have full channel state information (CSI) (perfect estimates of all sub-channel gains), then the transmitter can optimize the power distribution in frequency during each coherence time. In that case, assuming an average power constraint, the capacity tends to infinity as $\log M$, where M is the number of sub-channels [6], [7].[1]

With limited feedback per coherence time, the capacity remains finite as $M \to \infty$, since the transmitter can obtain CSI for only a finite number of sub-channels.

Since we assume that neither the transmitter nor the receiver knows the channel at the beginning of each coherence block, any CSI, which is relayed to the transmitter, must be estimated at the receiver. For that purpose, we assume the transmitter *probes* a subset of sub-channels by transmitting a pilot or training sequence over each sub-channel. The number of probed, or active sub-channels N is called the *probing bandwidth*. The receiver estimates the active sub-channels and indicates to the transmitter which of those should be used for data transmission. That requires no more than one feedback bit per active sub-channel. The transmitted power for data transmission is then uniformly spread over that subset of active sub-channels ("on-off" power allocation).

[1]Throughout the paper we assume natural logarithms, so that capacity is measured in nats per channel use.

Given a large number of coherence bands (sub-channels) and limited training power, the quality of the channel estimate decreases as the training power is spread over a larger number of active sub-channels. However, increasing the set of active sub-channels also increases diversity. Balancing these two trends gives an optimal probing bandwidth, which depends on the training power and the coherence time.

We optimize the probing bandwidth with and without one-bit feedback by maximizing a lower bound on the capacity over the training length, training power, and feedback threshold for data transmission. The lower bound is the rate achieved with a Gaussian code book, assuming coherent linear detection with a linear Minimum Mean Square Error (MMSE) channel estimate obtained from the pilot.

We show that the capacity of the pilot-based scheme with partial feedback increases as $S \log L$, where S is the Signal-to-Noise Ratio (SNR) and L is the coherence time in number of channel uses. This is consistent with the observations made in [8], which considers a similar type of feedback scheme for a Rayleigh fading channel at low SNRs. Hence there is a critical coherence time beyond which the partial feedback scheme performs better than flash signaling, which achieves a capacity of S bits/channel use as $M \to \infty$. For the model considered, this critical value depends only on the Rayleigh fading assumption, and is independent of the system parameters (i.e., signal power, noise variance, and variance of channel gains). In addition, the probing bandwidth grows sublinearly, namely, as $S\frac{L}{\log^2 L}$.

Without feedback as the bandwidth $M \to \infty$, the pilot-based scheme considered cannot perform better than a flash signaling scheme, which achieves the same capacity as that achieved with complete channel knowledge at the receiver [4], [5]. The lower bound on capacity for the pilot-based scheme approaches the flash capacity as $L \to \infty$, although the gap diminishes slowly at the rate $\frac{S}{L^{1/3}}$. The optimal probing bandwidth without feedback increases as $SL^{1/3}$, which is much slower than the optimal growth with feedback. With feedback the transmitter therefore benefits from probing a significantly larger number of sub-channels, even with the increase in channel estimation error.

We also show that the preceding results apply with an additional peak power constraint on training. The main effect of this additional constraint is to increase the optimized training length.

Related work in [2] studies the tradeoff between diversity and noncoherence loss without feedback for a doubly-selective *i.i.d.* block fading model. There a range of optimal spreading bandwidths is determined under a peak power constraint. The tradeoff between diversity and channel estimation error without feedback is studied in [9] for a *sparse* wideband channel. Other related work, which considers different channel models and objectives, have been presented in [10]–[12]. The emphasis in [10], [11] is on designing a multi-channel probing protocol, which balances power for probing against power for data transmission. There the data transmission is limited to one sub-channel and the relation between number of probed channels and channel measurement error is not taken into account. Another multi-channel probing strategy based on a multi-armed bandit model is studied in [12]. There the objective is to maximize the received SNR, and data transmission is again limited to one *best* sub-channel. Here the emphasis is on characterizing the optimal probing bandwidth for a wideband fading channel taking into account the tradeoff between probing bandwidth and channel estimation error. Finally, optimizing probing bandwidth to maximize capacity is closely related to minimizing the energy-per-bit for each active sub-channel, as introduced in [5]. (See also [13] where the energy-per-bit objective is used to optimize a pilot signal without feedback.)

In the next section we present the wideband channel model, and in Section III we present the achievable rate performance objective. Our main results are presented in Section IV, along with numerical results, which compare the asymptotic results with those for a finite-size system.

II. System Model

The channel is assumed to be block fading in both time and frequency. That is, the bandwidth is divided into coherence bands (sub-channels), each of which experiences block fading across time. Each time-frequency coherence block is assumed to have the same duration of L channel uses and the same bandwidth. The channel gain within each time-frequency coherence block is constant, and is chosen from a Rayleigh distribution. The channel gains are *i.i.d.* across all coherence blocks.

To model a wideband system, we assume an infinite number of sub-channels. We will see, however, that the transmitter should always transmit on a finite subset of sub-channels. We refer to those sub-channels (for a given coherence time) as being “active”. The number of active sub-channels is N.

Focusing on a particular active coherence block corresponding to the i^{th} sub-channel, the $L \times 1$

vector of received samples across time is given by

$$\mathbf{Y}_i = h_i \, \mathbf{X}_i + \mathbf{Z}_i \tag{1}$$

where $\mathbf{X}_i$ is the vector of transmitted symbols, h_i is the sub-channel coefficient, assumed to be circularly symmetric, complex Gaussian (CSCG) with zero mean and variance σ_h^2, and $\mathbf{Z}_i$ is the vector of CSCG noise samples with covariance matrix ${\sigma_z}^2 \, \mathbf{I}$. We omit the dependence on the coherence time index for convenience.

Because the receiver does not know the channel, a sequence of T training symbols is transmitted for channel estimation at the beginning of each coherence block. We therefore have $L = T + D$, where D is the number of channel uses reserved for data transmission. The input vector is therefore $\mathbf{X}_i^\dagger = [\mathbf{X}_{Ti}^\dagger \; \mathbf{X}_{Di}^\dagger]$, where $\mathbf{X}_{Ti}$ and $\mathbf{X}_{Di}$ are $T \times 1$ and $D \times 1$ vectors of training and data symbols, respectively. Similarly, the received and noise vectors can be partitioned as $\mathbf{Y}_i^\dagger = [\mathbf{Y}_{Ti}^\dagger \; \mathbf{Y}_{Di}^\dagger]$ and $\mathbf{Z}_i^\dagger = [\mathbf{Z}_{Ti}^\dagger \; \mathbf{Z}_{Di}^\dagger]$. The average powers during the training and data phases are P_T and P_D, respectively, i.e., $\frac{1}{T} \sum_{i=1}^{N} E[\mathbf{X}_{Ti}^\dagger \mathbf{X}_{Ti}] = P_T$, $\frac{1}{D} \sum_{i=1}^{N} E[\mathbf{X}_{Di}^\dagger \mathbf{X}_{Di}] = P_D$, and we assume an average power constraint

$$\alpha \, P_T + (1 - \alpha) \, P_D = P \tag{2}$$

where $\alpha = \frac{T}{L}$ is the fraction of the coherence time spent on training.

Based on the training segment of the coherence block, the receiver computes the Minimum Mean Squared Error (MMSE) channel estimate $\hat{h}_i$, and uses that estimate for both feedback and coherent detection. The feedback, to be described, occurs between the training and the data transmission, and its duration is assumed to be an insignificant part of the coherence time.

Denoting the channel estimation error as $e_i = h_i - \hat{h}_i$, we can therefore rewrite the data segment of (1) as

$$\mathbf{Y}_{Di} = \hat{h}_i \, \mathbf{X}_{Di} + e_i \, \mathbf{X}_{Di} + \mathbf{Z}_{Di} \tag{3}$$

where $\hat{h}_i$ and e_i are uncorrelated, zero-mean, CSCG random variables. The error variance is therefore ${\sigma_e}^2 = E(|h_i|^2) - E(|\hat{h}_i|^2)$. Since the sub-channel coefficients are independent, the training power P_T is divided equally among the N active sub-channels, and the channel estimation is performed separately for each active sub-channel. Hence the MMSE, or error variance, is given by

$${\sigma_e}^2 = {\sigma_h}^2 - {\sigma_h}^4 \left(\frac{T \, P_T}{{\sigma_h}^2 \, T \, P_T + N \, {\sigma_z}^2} \right). \tag{4}$$

For each active sub-channel i, the receiver feeds back one bit, which indicates whether or not the channel estimate exceeds a threshold t_0. If $|\hat{h}_i|^2 > t_0$, then the transmitter continues to transmit data on the sub-channel, otherwise, the sub-channel is idle. The data power P_D is divided evenly among the subset of active sub-channels on which data transmission occurs. We note that for a given set of active sub-channels, this uniform power allocation, corresponding to one-bit feedback per sub-channel, typically performs quite close to the optimal water pouring power allocation [6], [7].

The choice of active sub-channels N should balance the tradeoff between diversity and channel estimation error. Namely, as N increases, the likelihood of finding a set of good sub-channels increases, corresponding to increased diversity; however, the training power per channel decreases, which increases the channel estimation error. The optimal N, which balances this tradeoff, depends on the channel coherence time. As the channel coherence time L increases, more energy can be allocated to training without decreasing the data rate. We therefore expect the optimal N to increase with L.

III. Achievable Rate Objective

The transmitter is assumed to code over multiple coherence blocks in frequency and time, so that our performance objective is ergodic capacity. We wish to optimize this objective over the training power P_T, data power P_D, fraction of training symbols α, and number of active sub-channels N.

From (3), the ergodic capacity is given by

$$C = (1-\alpha)\frac{1}{D}\sum_{i=1}^{N} \max_{p(\mathbf{X}_{Di}|\hat{h}_i)} I(\mathbf{X}_{Di};\mathbf{Y}_{Di}|\hat{h}_i) \tag{5}$$

$$\text{Subject to:} \sum_{i=1}^{N} E_{\hat{h_i}}\left[tr(\mathbf{Q}_{\hat{h_i}})\right] = P_D\, D$$

where $p(\mathbf{X}_{Di}|\hat{h}_i)$ is the probability density of $\mathbf{X}_{Di}$ given $\hat{h}_i$, $\mathbf{Q}_{\hat{h_i}} = E[\mathbf{X}_{Di}\mathbf{X}_{Di}^{\dagger}|\hat{h}_i]$, and tr$(\mathbf{Q}_{\hat{h_i}})$ denotes the trace of the matrix $\mathbf{Q}_{\hat{h_i}}$. The input density, which maximizes the mutual information is unknown; however, a lower bound is obtained by assuming that $p(\mathbf{X}_{Di}|\hat{h}_i)$ is Gaussian [14], [15]. With this assumption the differential entropy $h(\mathbf{X}_{Di}|\hat{h}_i) = E_{\hat{h_i}}[\log(|\pi e \mathbf{Q}_{\hat{h_i}}|)]$. Also, $h(\mathbf{X}_{Di}|\mathbf{Y}_{Di}, \hat{h}_i)$ is upper bounded by the entropy of a Gaussian random variable with variance given by the Mean Squared Error (MSE) associated with the linear MMSE estimate of $\mathbf{X}_{Di}$ given $\mathbf{Y}_{Di}$ and $\hat{h}_i$ [14]. After some manipulation and application of the matrix inversion lemma we obtain the lower bound

$$I(\mathbf{X}_{Di};\mathbf{Y}_{Di}|\hat{h}_i) \geq \underline{I}(\mathbf{X}_{Di};\mathbf{Y}_{Di}|\hat{h}_i) = E_{\hat{h}_i}\left[\log\det\left(\mathbf{I} + |\hat{h}_i|^2({\sigma_e}^2\mathbf{Q}_{\hat{h}_i} + {\sigma_z}^2\mathbf{I})^{-1}\mathbf{Q}_{\hat{h}_i}\right)\right]$$

From Hadamard's inequality this lower bound is maximized when $\mathbf{Q}_{\hat{h}_i}$ is diagonal. Letting $\mathbf{Q}_{\hat{h}_i} = P(\hat{h}_i)\mathbf{I}_{D\times D}$, where $P(\hat{h}_i)$ is the power of a data symbol in the i^{th} subchannel as a function of the channel estimate $\hat{h}_i$, the lower bound on mutual information can be re-written as

$$\underline{I}(\mathbf{X}_{Di};\mathbf{Y}_{Di}|\hat{h}_i) = D\, E_{\hat{h}_i}\left[\log\left(1 + \frac{P(\hat{h}_i)|\hat{h}_i|^2}{P(\hat{h}_i)\,{\sigma_e}^2 + {\sigma_z}^2}\right)\right]. \tag{6}$$

This lower bound is the achievable rate with a Gaussian input distribution and a coherent linear receiver, which treats the channel estimation error as additive noise. In particular, the receiver does not attempt to improve upon the channel estimate during the data reception period.

Substituting this lower bound on mutual information into the capacity expression (5), we obtain the following lower bound on capacity given power allocation $P(\hat{h}_i)$,

$$\underline{C} = (1-\alpha)\,N\,E_{\hat{h}_i}\left[\log\left(1 + \frac{P(\hat{h}_i)|\hat{h}_i|^2}{P(\hat{h}_i)\,{\sigma_e}^2 + {\sigma_z}^2}\right)\right] \tag{7}$$

$$\text{with power constraint} \quad E_{\hat{h}_i}\left[P(\hat{h}_i)\right] \leq \frac{P_D}{N}.$$

Here we have used the assumption that the channel estimate $\hat{h}_i$ has the same distribution for all N active sub-channels.

IV. Optimal Parameters

The lower bound (7) depends on the transmitter power allocation. We consider two scenarios. In the first, the transmitter uniformly distributes the power across all data symbols (in time and frequency). That is, the transmitter does not make use of feedback. In the second scenario, the transmitter transmits with constant power only on sub-channels for which the channel estimate $|\hat{h}_i|^2$ is above a threshold. We refer to this as an "on-off" power allocation. Although the optimal power allocation strategy, which maximizes (7), resembles water pouring [16], the "on-off" power allocation is simpler to analyze and is known to achieve near-optimal performance with perfect channel estimates [6], [7].

A. No feedback

Since the power is spread evenly over all data symbols, $P(\hat{h}_i) = P_D/N$ for all $i = 1, 2, \ldots, N$. Also, $|\hat{h}_i|^2$ has an exponential distribution with expected value

$$\sigma_{\hat{h}}{}^2 = \sigma_h{}^2 - \sigma_e{}^2 = \sigma_h{}^4 \left(\frac{T\,P_T}{\sigma_h{}^2\,T\,P_T + N\,\sigma_z{}^2} \right), \tag{8}$$

so that the lower bound on capacity (7) can be evaluated as

$$\underline{C}_{nfb} = (1-\alpha)\,N\,e^x\,\gamma(x) \tag{9}$$

where $\gamma(x) = \int_x^\infty \frac{e^{-t}}{t}\,dt$ is the zeroth-order incomplete gamma function and

$$x = \frac{\sigma_h{}^2}{\sigma_{\hat{h}}{}^2} \left(1 + \frac{N\,\sigma_z{}^2}{P_D\,\sigma_h{}^2} \right) - 1. \tag{10}$$

The achievable rate in (9) depends on α, P_T and N. Note that for any permissible value of α and P_T, as $N \to \infty$, $\underline{C}_{nfb} \to 0$. This is because increasing the number of active sub-channels degrades the channel estimates. Similarily, for any fixed N, $\underline{C}_{nfb} \to 0$ as α or P_T approaches a boundary value of the constraint set (e.g., $\alpha \to 0$ or $\alpha \to 1$). Hence we can jointly optimize α, P_T, and N to maximize $\underline{C}_{nfb}$.

We begin by optimizing the training fraction α. For this case, we write (7) as an integral, and substitute $P(\hat{h}_i) = P_D/N$, where P_D is given by (2), to obtain

$$\underline{C}_{nfb} = (1-\alpha)\,N \int_0^\infty \log\left(1 + \frac{(P-\epsilon_T)\,t}{(P-\epsilon_T)\,\sigma_e{}^2 + N\,\sigma_z{}^2\,(1-\alpha)} \right) \frac{1}{\sigma_{\hat{h}}{}^2}\,e^{-t/\sigma_{\hat{h}}{}^2} dt \tag{11}$$

where $\epsilon_T = \alpha P_T$ is the average training power.

Note from (4) that $\sigma_e{}^2$, and hence $\sigma_{\hat{h}}{}^2$, depend on α only through ϵ_T. This observation and the fact that $N\,(1-\alpha)\,\log(1 + \frac{a}{b+N\,(1-\alpha)})$ is a decreasing function of α for all positive a and b implies that the integrand, and hence the integral, is maximized by choosing the smallest possible α while keeping the average training power ϵ_T fixed. Since at least one channel use is needed for training, we must have $\alpha \geq 1/L$. In addition, we impose a peak power constraint P_{peak} on each training symbol over all sub-channels so that $\alpha \geq \epsilon_T/(NP_{peak})$, hence the optimal training length fraction is

$$\alpha = \max\left\{ \frac{\epsilon_T}{NP_{peak}}, \frac{1}{L} \right\}. \tag{12}$$

We substitute this value of α in (9) and proceed to optimize N and ϵ_T. Define the average SNR $S = \frac{P\sigma_h^2}{\sigma_z^2}$, the peak SNR $S_{peak} = \frac{P_{peak}\sigma_h^2}{\sigma_z^2}$, normalized training power $\bar{\epsilon}_T = \frac{\epsilon_T}{P}$, and normalized bandwidth $\bar{N} = \frac{N}{S}$, so that we can rewrite (10) as

$$x = \left(1 + \frac{\bar{N}}{\bar{\epsilon}_T L}\right)\left(1 + \frac{\bar{N}(1-\alpha)}{1-\bar{\epsilon}_T}\right) - 1. \tag{13}$$

Note that the flash capacity is $C_{flash} = S$, which is also the capacity with perfect channel knowledge at the receiver [5]. With α given by (12), maximizing (9) with respect to N and ϵ_T is equivalent to maximizing with respect to the normalized values $\bar{N}$ and $\bar{\epsilon}_T$. The optimal values $\bar{N}^\star$ and $\bar{\epsilon}_T^\star$ depend only on L and S_{peak}, hence we can write

$$\epsilon_T{}^\star = P\bar{\epsilon}_T^*(L, S_{peak}) \qquad N^\star = S\bar{N}^*(L, S_{peak}). \tag{14}$$

The lower bound on capacity is then

$$\underline{C}_{nfb}^\star = C_{flash}\bar{C}_{nfb}^\star(L, S_{peak}) \tag{15}$$

where $\bar{C}_{nfb}^\star(L, S_{peak})$ is obtained by substituting $N^\star$, $\epsilon_T^\star$, and the α in (12) into (9). Clearly, C_{flash} is an upper bound on the performance of the training-based scheme without feedback. We present numerical examples in Section IV-A.2, which illustrate the behavior of the optimized parameters, but first examine their asymptotic behavior as L becomes large.

Consider the special case in which the peak training power is unbounded. If $S_{peak} \to \infty$, then $\alpha = 1/L$. In that case, the optimized parameters $\bar{\epsilon}_T^\star$, $\bar{N}^\star$, and the achievable rate $\bar{C}_{nfb}^\star$ depend only on L. To determine these functions we note that $e^x\,\gamma(x)$ in (9) is a decreasing function of x, and depends on $\bar{\epsilon}_T$ only through x. Therefore the optimal value of $\bar{\epsilon}_T$ minimizes x. The solution is given in terms of the function

$$g(u) = -z + (z^2 + z)^{1/2} \tag{16}$$

where $z(u) = \frac{(1-\frac{1}{L})u+1}{L-2}$. Namely,

$$\bar{\epsilon}_T^*(L) = g(\bar{N}^*) \tag{17}$$

and

$$\bar{N}^*(L) = \arg\max_y \left[y\, e^{x'}\,\gamma(x')\right]$$

where

$$x' = \left(1 + \frac{y(1-\frac{1}{L})}{1-g(y)}\right)\left(1 + \frac{y}{L\,g(y)}\right) - 1.$$

To obtain more insight we analyze the behavior of the optimal parameters asymptotically as L becomes large.

1) Asymptotic Analysis: For the more general case of finite S_{peak}, the behavior of the optimal parameters as L increases, and the rate at which $\underline{C}^{\star}_{nfb}$ approaches C_{flash} is given by the next theorem. We use the following notation. Suppose that $\lim_{L\to\infty} \frac{f_1(L)}{f_2(L)} = c$. If $c = 0$, then we write $f_1 \prec f_2$; if $c \in [0, 1]$, then we write $f_1 \precsim f_2$, and if $c = 1$, then $f_1 \asymp f_2$. Furthermore, the values $\bar{\epsilon}_T = \bar{\epsilon}^a_T(L)$ and $\bar{N} = \bar{N}^a(L)$ are called *asymptotically optimal* if they satisfy the following condition,

$$C_{flash} - \underline{C}_{nfb}(\bar{\epsilon}^a_T, \bar{N}^a) \asymp C_{flash} - \underline{C}^{\star}_{nfb}. \tag{18}$$

That is, the capacity with the values $\bar{\epsilon}^a_T$ and $\bar{N}^a$ converges to the flash capacity at the optimal (first-order) rate.

Theorem 1: As $L \to \infty$, if S_{peak} satisfies

$$L \precsim S_{peak}, \tag{19}$$

then the optimal parameters for the training scheme without feedback satisfy

$$\bar{\epsilon}^{\star}_T \asymp \frac{1}{(2L)^{1/3}} \tag{20}$$

$$\bar{N}^{\star} \asymp \left(\frac{L}{4}\right)^{1/3} \tag{21}$$

and the corresponding achievable rate satisfies

$$C_{flash} - \underline{C}^{\star}_{nfb} \asymp \frac{3C_{flash}}{(2L)^{1/3}}. \tag{22}$$

In addition, if $\frac{1}{L^{1/3}} \prec S_{peak} \precsim L$ then the parameter values given by (20) and (21) are asymptotically optimal.

The proof is given in Appendix I. There it is also shown that for $\frac{1}{L^{1/3}} \prec S_{peak} \precsim L$, the asymptotically optimal values $\bar{\epsilon}^a_T = \frac{1}{(2L)^{1/3}}$ and $\bar{N}^a = \left(\frac{L}{4}\right)^{1/3}$ satisfy the stationarity conditions as L becomes large, that is,

$$\lim_{L\to\infty} \frac{\partial \underline{C}_{nfb}}{\partial \bar{\epsilon}_T}\Big|_{\bar{\epsilon}_T = \bar{\epsilon}^a_T, \bar{N} = \bar{N}^a} = 0 \tag{23}$$

$$\lim_{L\to\infty} \frac{\partial \underline{C}_{nfb}}{\partial \bar{N}}\Big|_{\bar{\epsilon}_T = \bar{\epsilon}^a_T, \bar{N} = \bar{N}^a} = 0. \tag{24}$$

The theorem specifies the first-order growth rate of the optimized bandwidth, training power, and achievable rate with L provided that S_{peak} does not tend to zero faster than $1/L^{1/3}$. (Of course, this includes the important cases in which S_{peak} is constant, or increases with L.). From (22), the gap between the achievable rate for the training scheme and the flash capacity without feedback approaches zero slowly as $\frac{1}{L^{1/3}}$. From (20) the optimized training power ϵ_T decreases to zero as $L \to \infty$. The training *energy* per subchannel is $\frac{\epsilon_T L}{N}$, and combining this with the asymptotic growth for N in (21) shows that that the training energy *increases* as $L^{1/3}$, independent of the peak training power. Furthermore, the channel estimation error $\sigma_e^2 = \sigma_h^2 - \sigma_{\hat{h}}^2$ decreases to zero as $\frac{1}{L^{1/3}}$. Interestingly, the theorem also implies that this remains true if the training symbol power S_{peak} goes to zero more slowly than the optimized data power per subchannel, given by $\frac{P_D}{N}$, which decreases as $\frac{1}{L^{1/3}}$.

Although the choice of S_{peak} does not effect the first-order behavior of the optimal training power, bandwidth or capacity, it does influence the optimal training length. Recall that the training length fraction is given by (12). Hence with parameter values chosen as (20) and (21), if $L^{1/3} \precsim S_{peak}$ then $\alpha = 1/L$, corresponding to a training length of one symbol. For smaller peak powers in the range $L^{-1/3} \prec S_{peak} \precsim L^{1/3}$, the training length is larger than one. From (12), with fixed peak power S_{peak} the optimized training length $\alpha^\star L$ increases as $L^{1/3}$, and increases faster if S_{peak} decreases with L. Namely, $\alpha^\star L$ increases as $L^{2/3}$ if S_{peak} decreases at a rate close to $\frac{1}{L^{1/3}}$.

Given a large S_{peak}, the first-order values given by the theorem provide accurate estimates of the corresponding optimal parameters, even for relatively small coherence times L. However, with smaller values of S_{peak}, the second-order terms can be non-negligible, so that the coherence time must be much larger for the first-order values to be accurate. These asymptotic trends are illustrated by the numerical results presented next.

2) Numerical Results: Here we compare the preceding asymptotic results with numerical results obtained by optimizing α, ϵ_T, and N directly in (9) with finite values for L and S_{peak}. Results for both scenarios with and without feedback are shown; however, discussion of the results with feedback is postponed to the next section.

Figs. 1 and 2 show plots of the optimal normalized number of sub-channels $\bar{N}^\star(L, S_{peak})$ and normalized average training power $\bar{\epsilon}_T^\star(L, S_{peak})$ versus L, respectively. Fig. 3 shows the normalized capacity $\bar{C}_{nfb}^\star(L, S_{peak})$. All figures show results with $S_{peak} =$ 10, 0, and -5 dB.

As the coherence time L increases, we can obtain more accurate channel estimates with less training power per subchannel. Hence the optimal training power decreases with coherence time, and the optimal number of active sub-channels increases. This allows for an increase in data power and number of subchannels used for data transmission, and hence the capacity increases with L.

The figures show that the curves for $\bar{N}^{\star}(L, S_{peak})$, $\bar{\epsilon}_T^{\star}(L, S_{peak})$, and $\bar{C}_{nfb}^{\star}(L, S_{peak})$ are insensitive to S_{peak}. (In fact, they almost overlap for the different values of S_{peak} shown.) This is consistent with Theorem 1, since S_{peak} does not affect the first-order growth rates. Also shown are the first-order asymptotic values of these function given by (20)-(22). These values are quite close to the actual optimized values.

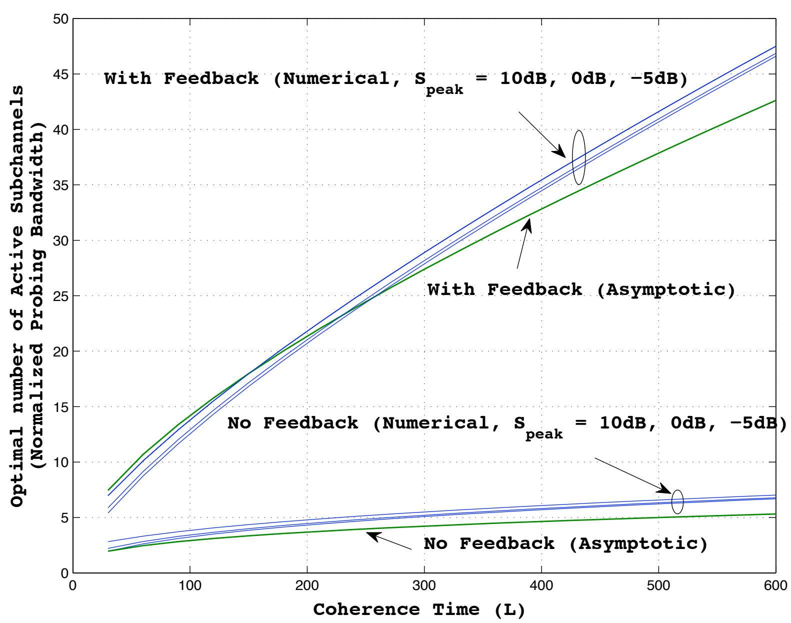

Fig. 1. Optimal normalized probing bandwidth $\bar{N}^{\star}(L, S_{peak})$ with and without feedback versus the coherence time. The peak power values are taken to be $S_{peak} = 10dB, 0dB, -5dB$.

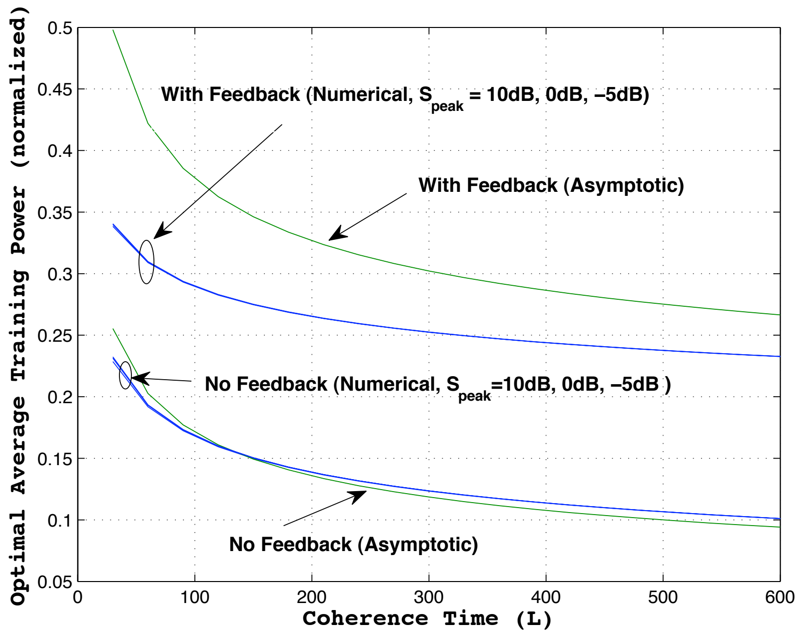

Fig. 2. Optimal normalized training power $\bar{\epsilon}_T^{\star}(L, S_{peak})$ with and without feedback versus the coherence time. The peak power values are taken to be $S_{peak} = 10dB, 0dB, -5dB$.

B. Partial Feedback

The feedback is used to specify the on-off power allocation

$$P(\hat{h}_i) = \begin{cases} P_o & \text{if } |\hat{h}_i|^2 \geq t_0 \\ 0 & \text{otherwise} \end{cases}$$

which requires one feedback bit per subchannel per coherence block. From (7), the lower bound on ergodic capacity is

$$\underline{C}_{fb} = (1-\alpha)\, N \int_{t_0}^{\infty} \log\left(1 + \frac{P_o\, t}{P_o\, {\sigma_e}^2 + {\sigma_z}^2}\right) f(t)dt \tag{25}$$

with the power constraint

$$\int_{t_0}^{\infty} P_o f(t)dt = \frac{P_D}{N}$$

where $f(t) = \frac{1}{{\sigma_{\hat{h}}}^2}\, e^{-t/{\sigma_{\hat{h}}}^2}$. This can be rewritten as

$$\underline{C}_{fb} = (1-\alpha)\, N\, e^{-\frac{t_0}{{\sigma_{\hat{h}}}^2}} \left[\log\left(\frac{y}{y - \frac{t_0}{{\sigma_{\hat{h}}}^2}}\right) + e^y \gamma(y)\right] \tag{26}$$

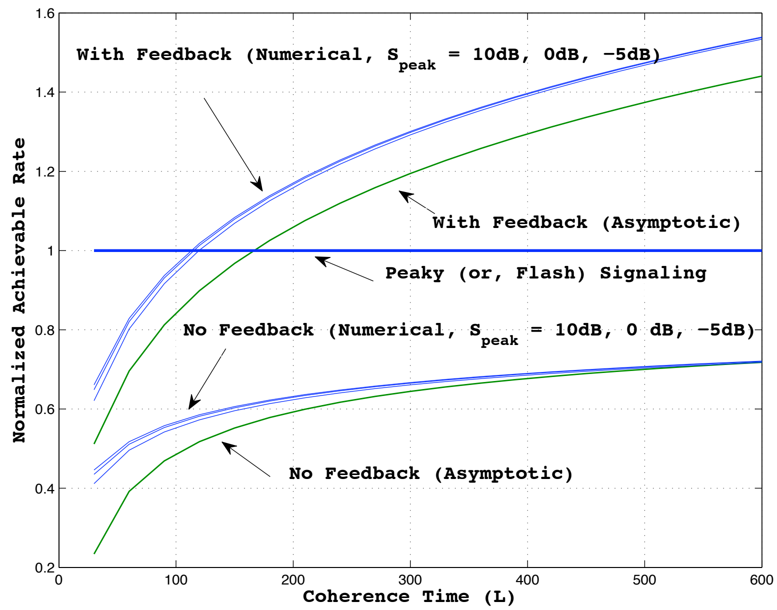

Fig. 3. Optimal normalized achievable rate with feedback $\overline{C}^{\star}_{fb}(L, S_{peak})$ and without feedback $\overline{C}^{\star}_{nfb}(L, S_{peak})$ versus the coherence time. The peak power values are taken to be $S_{peak} = 10dB, 0dB, -5dB$.

where

$$y = \frac{\sigma_h{}^2}{\sigma_{\hat{h}}{}^2}\left(1 + \frac{N\,\sigma_z{}^2\, e^{-\frac{t_0}{\sigma_{\hat{h}}{}^2}}}{P_D\,\sigma_h{}^2}\right) - 1 + \frac{t_0}{\sigma_{\hat{h}}{}^2}.$$

As before, we wish to maximize $\underline{C}_{fb}$ with respect to the training fraction α, average training power ϵ_T, on-off threshold t_0, and number of subchannels N. The normalized values are again defined as $\bar{N} = \frac{N}{S}$, $\bar{\epsilon}_T = \frac{\epsilon_T}{P}$, and $\bar{C}_{fb} = \frac{\underline{C}_{fb}}{S}$. We also define the normalized threshold $\bar{t}_o = \frac{t_o}{\sigma_h^2}$. Following an argument analogous to that given in the preceding subsection, it can be shown that the optimal values satisfy

$$\alpha^\star = \max\left\{\frac{\bar{\epsilon}_T^\star}{\bar{N}^\star S_{peak}}, \frac{1}{L}\right\} \tag{27}$$

$$\epsilon_T{}^\star = P\,\bar{\epsilon}_T^\star(L, S_{peak}) \qquad N^\star = S\bar{N}^\star(L, S_{peak}) \qquad t_0{}^\star = \sigma_h{}^2\,\bar{t_0}{}^\star(L, S_{peak}) \tag{28}$$

and the lower bound on capacity

$$\underline{C}^\star_{fb} = S\,\bar{C}^\star_{fb}(L, S_{peak}), \tag{29}$$

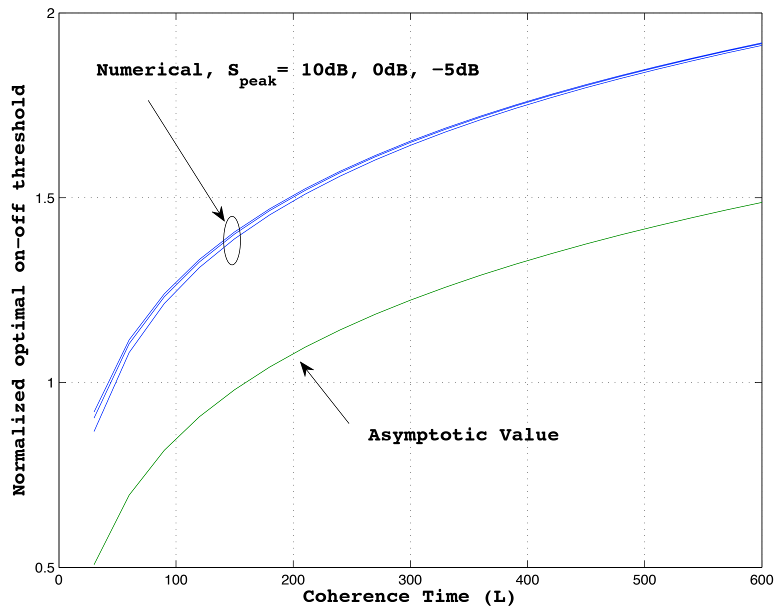

Fig. 4. Optimal normalized on-off threshold $\bar{t}_o^\star(L, S_{peak})$ versus the coherence time. The peak power values are taken to be $S_{peak} = 10dB, 0dB, -5dB$.

where $\bar{\epsilon}_T^\star, \bar{N}^\star, \bar{t_0}^\star$ and $\bar{C}_{fb}^\star$ represent the optimal values of the normalized parameters and capacity, respectively, which are functions of L and S_{peak} only. Here $N^\star$ is the total number of subchannels probed with pilot symbols. Data is transmitted only on the subset of $N^\star$ channels with estimated channels gains $|\hat{h}_i|^2 \geq t_0^\star$. In contrast, with no feedback data is transmitted on all $N^\star$ subchannels. In either case, $N^\star$ represents the optimal probing bandwidth. Although it appears to be difficult to evaluate the preceding optimal system values in closed-form, it is straightforward to evaluate them numerically. Examples will be discussed in Section IV-B.2.

1) Asymptotic Analysis: Here we give the growth rate of the optimal parameters and the corresponding achievable rate versus coherence time when the coherence time is large. We then compare these results to the analogous results without feedback.

Similar to the case without feedback, we say that $\bar{N} = \bar{N}^a(L)$, $\bar{\epsilon}_T = \bar{\epsilon}_T^a(L)$, and $\bar{t}_o = \bar{t}_o^a(L)$ are *asymptotically optimal* if they satisfy the following condition,

$$\underline{C}_{fb}(\bar{\epsilon}_T^a, \bar{N}^a, \bar{t}_o^a) \asymp \underline{C}_{fb}^\star. \tag{30}$$

That is, the capacity with the values $\bar{\epsilon}_T^a, \bar{N}^a$ and $\bar{t}_o^a$ has the optimal (first-order) growth rate with L. In contrast to the case without feedback, here the capacity grows without bound as $L \to \infty$. Hence we are concerned with the first-order growth of the capacity, as opposed to the convergence rate to a constant.

Theorem 2: As $L \to \infty$, if S_{peak} satisfies

$$L \precsim S_{peak}, \tag{31}$$

then the optimal parameters for the training scheme with feedback satisfy

$$\bar{N}^\star \asymp \frac{L}{\log^2 L} \tag{32}$$

$$\bar{\epsilon}_T^\star \asymp \frac{1}{\log \bar{N}^\star} \tag{33}$$

and the corresponding achievable rate satisfies

$$\underline{C}_{fb}^\star \asymp S \log \bar{N}^\star. \tag{34}$$

Moreover the optimal threshold is

$$\bar{t}_o^\star = \log \bar{N}^\star - (1+\delta) \log\log \bar{N}^\star + o(\log\log \bar{N}^\star) \tag{35}$$

where $\delta \in (0,1)$ is given by $\delta = \frac{\log v}{\log\log \bar{N}^\star} - 1$ and v satisfies

$$\log\left(\frac{v}{v + \log(\bar{N}^\star/v)}\right) = v \exp\left(v + \log(\bar{N}^\star/v)\right) \gamma\left(v + \log(\bar{N}^\star/v)\right). \tag{36}$$

In addition, for $\frac{\log^2 L}{L} \prec S_{peak} \precsim L$ the parameter values given by (32), (33) and (35) are asymptotically optimal.

The proof is given in Appendix II. There it is also shown that for $\frac{\log^2 L}{L} \prec S_{peak} \precsim L$, the asymptotically optimal values $\bar{\epsilon}_T^a$, $\bar{N}^a$ and $\bar{t}_o^a$ given by (33), (32) and (35), respectively, satisfy the stationarity conditions as L becomes large, that is,

$$\lim_{L\to\infty} \frac{\partial \underline{C}_{fb}}{\partial \bar{\epsilon}_T}|_{\bar{\epsilon}_T=\bar{\epsilon}_T^a, \bar{N}=\bar{N}^a, \bar{t}_o=\bar{t}_o^a} = 0 \tag{37}$$

$$\lim_{L\to\infty} \frac{\partial \underline{C}_{fb}}{\partial \bar{N}}|_{\bar{\epsilon}_T=\bar{\epsilon}_T^a, \bar{N}=\bar{N}^a, \bar{t}_o=\bar{t}_o^a} = 0 \tag{38}$$

$$\lim_{L\to\infty} \frac{\partial \underline{C}_{fb}}{\partial \bar{t}_o}|_{\bar{\epsilon}_T=\bar{\epsilon}_T^a, \bar{N}=\bar{N}^a, \bar{t}_o=\bar{t}_o^a} = 0. \tag{39}$$

The theorem states that the lower bound on capacity grows as $\log L$, in contrast to the corresponding lower bound without feedback, which is upper bounded by the flash capacity (a finite constant). Hence for large coherence times, partial feedback provides a substantial gain in achievable rate. The asymptotic growth with feedback is enabled by the optimal on-off threshold $t_o^\star$, which allows the average number of subchannels above the threshold, i.e., $Ne^{-t_o/\sigma_{\hat{h}}^2}$, to grow approximately as $(\log N)^{1+\delta}$, where the constant δ in (35) is small and positive.

As in the case without feedback, the training *energy* per subchannel $\frac{\epsilon_T L}{N}$ *increases* as $\log L$ independent of S_{peak}. The average training *power* $\epsilon_T^\star \asymp \frac{P}{\log L}$ is much larger with partial feedback than without; however, the bandwidth $N^\star \asymp S\frac{L}{\log^2 L}$ is also substantially larger so that the training energy per active sub-channel $\frac{\epsilon_T L}{N}$ grows as $\log L$, as opposed to $L^{1/3}$, without feedback. Correspondingly, the channel estimation error with feedback goes to zero at a slower rate of $1/\log L$. However, the loss in rate due to the larger estimation error is offset by the diversity gain from the on-off power adaptation and larger bandwidth. For these results to hold the optimized training power can be almost as small as the optimized data power per subchannel, given by $\frac{P_D}{N}$, which decreases as $\frac{\log^2 L}{L}$.

With the exception of training length the first-order behavior of the optimal parameters does not depend on S_{peak} provided that $\frac{\log^2 L}{L} \prec S_{peak}$. From (27) if $\log L \precsim S_{peak}$, then $\alpha = 1/L$, corresonding to a training length of one symbol. For smaller peak powers in the range $\frac{\log^2 L}{L} \prec S_{peak} \precsim \log L$, the optimal training length is larger than one. The training length grows as $\log L$ with fixed S_{peak}, and can be as large as $\frac{L}{\log L}$ if the peak power is close to $\frac{\log^2 L}{L}$. Given a fixed S_{peak}, the optimized training length grows much more slowly with partial feedback, i.e., as $\log L$, than without feedback (namely, as $L^{1/3}$).

The larger S_{peak} is, the more accurately the asymptotic estimates predict the corresponding values for finite L. The second-order terms can become significant for small values of S_{peak}. From (32) we can replace the probing bandwidth $\bar{N}^\star$ by L in the remaining asymptotic expressions. However, the asymptotic expressions more accurately predict the corresponding values for a finite size system when stated in terms of $\bar{N}^\star$. The asymptotic values are compared with the corresponding optimal values for a finite size system in the next subsection.

2) Numerical Results: Figs. 1-4 show numerically optimized parameters, i.e., $\bar{N}^\star(L, S_{peak})$, $\bar{\epsilon}_T^\star(L, S_{peak})$, $\bar{t}_0^\star(L, S_{peak})$, and the normalized achievable rate $\bar{C}_{fb}^\star(L, S_{peak})$ with $S_{peak} = 10$ dB, 0 dB, and -5 dB. As for the results without feedback, the curves for different S_{peak} almost overlap. Note that according to (29), $\bar{C}_{fb}^\star(L, S_{peak}) = 1$ corresponds to the capacity with impulsive, or flash signaling.

As stated in Theorem 2, the achievable rate with feedback asymptotically increases as $\log L$, whereas the corresponding rate without feedback approaches C_{flash}. Hence there exists a critical coherence time L_{crit} for which $\underline{C}^{\star}_{fb} > C_{flash}$ when $L > L_{crit}$. From (29) L_{crit} satisfies $\bar{C}^{\star}_{fb}(L_{crit}, S_{peak}) = 1$. Solving this numerically gives $L_{crit} \approx 120$, which is independent of the system parameters, and is insensitive to variations in S_{peak}. Hence for the block *i.i.d.* Rayleigh fading model, if the coherence time of the channel exceeds 120 channel uses, then the partial feedback scheme considered achieves a higher rate than the optimal impulsive signaling scheme without feedback, irrespective of the channel variance, noise variance, and average power constraint. If the channel coherence time is less than 120 channel uses, then this feedback scheme (with Gaussian codewords) does not achieve as high a capacity as impulsive signaling without feedback.

Also shown in the figures are the values obtained through asymptotic analysis. However, instead of using the asymptotic values for normalized bandwidth $\bar{N}^{\star}$ given by (32), we use the more accurate value (obtained in Appendix B) given by the solution of $L(\log \bar{N}^{\star})^2 = \bar{N}^{\star}$. The asymptotic values of $\bar{\epsilon}^{\star}_{T}$ and $\bar{t}^{\star}_{o}$ in the figures are given by (33) and (35), respectively. The asymptotic capacity in Fig. 3 is obtained by computing (26) at those parameter values. Note that the effect of the training fraction α (and hence S_{peak}) on this asymptotic value is negligible since α is small for the cases considered. These plots show that the asymptotic values are accurate estimates of the corresponding optimal parameters for finite coherence times. An exception is the optimal on-off threshold, although the asymptotic trend is evident for small L.

V. CONCLUSIONS

We have considered a time- and frequency-selective wideband channel with pilot-assisted training and feedback. The performance of this scheme, relative to the performance of impulsive, or flash signaling without feedback, depends critically on the coherence time L. Namely, for the block *i.i.d.* Rayleigh fading channel model, the capacity with the one-bit feedback scheme considered grows as $\log L$ when L is large, and surpasses the capacity without feedback when L exceeds 120 channel uses. As L becomes large, our results also show that the overhead (fraction of available power) devoted to training for channel estimation decreases to zero. However, the rate of decrease is sufficiently slow so that the training energy per sub-channel increases as $\log L$, and the channel estimation error tends to zero as $1/\log L$. Of course, other fading distributions can be analyzed within the framework presented, and may lead to different trends.

The model and results presented here can be extended in a few different directions. For example, although the on-off feedback scheme considered is known to have optimal properties [6], [7], an open question is whether or not other finite-rate feedback schemes can achieve higher capacities (e.g., see [17], [18]). Also, our model has assumed that the channel gains are *i.i.d.* across both frequency and time. A natural extension of this work is to consider a dynamic scheme for allocating training and data power with correlated fading (e.g., see [19], which optimizes training power for a narrow-band channel with correlated fading). An extension of this model to a multiple access channel, motivated by dynamic spectrum sharing applications, is studied in [20]. Finally, we have assumed that the transmitter codes across several coherence blocks so that our objective is the ergodic rate. A similar trade-off between channel estimation error and diversity could also be studied with outage capacity as the objective (e.g., see [21]).

APPENDIX I
PROOF OF THEOREM 1

Rewriting (9), the capacity without feedback is given by $\underline{C}_{nfb} = (1-\alpha)SC$, where

$$C = \bar{N}e^{x}\,\gamma(x), \qquad \alpha = \max\left\{\frac{\bar{\epsilon}_T}{\bar{N}S_{peak}}, \frac{1}{L}\right\} \tag{40}$$

and x is given by (13). Setting the partial derivatives of $\underline{C}_{nfb}$ with respect to $\bar{\epsilon}_T$ and $\bar{N}$ to zero gives the necessary conditions

$$\frac{\partial \underline{C}_{nfb}}{\partial \bar{\epsilon}_T} = S\left[-\frac{\partial \alpha}{\partial \bar{\epsilon}_T}C - \alpha\frac{\partial C}{\partial \bar{\epsilon}_T} + \frac{\partial C}{\partial \bar{\epsilon}_T}\right] = 0 \tag{41}$$

$$\frac{\partial \underline{C}_{nfb}}{\partial \bar{N}} = S\left[-\frac{\partial \alpha}{\partial \bar{N}}C - \alpha\frac{\partial C}{\partial \bar{N}} + \frac{\partial C}{\partial \bar{N}}\right] = 0 \tag{42}$$

where

$$\frac{\partial C}{\partial \bar{\epsilon}_T} = \bar{N}\left[e^{x}\gamma(x) - \frac{1}{x}\right]\left\{-\frac{\bar{N}}{\bar{\epsilon}_T^2 L}\left(1+\frac{\bar{N}(1-\alpha)}{1-\bar{\epsilon}_T}\right) + \left(1+\frac{\bar{N}}{\bar{\epsilon}_T L}\right)\left(\frac{\bar{N}(1-\alpha)}{(1-\bar{\epsilon}_T)^2} - \frac{1}{S_{peak}(1-\bar{\epsilon}_T)}\right)\right\} \tag{43}$$

and

$$\frac{\partial C}{\partial \bar{N}} = e^{x}\gamma(x) + \bar{N}\left[e^{x}\gamma(x) - \frac{1}{x}\right]\left\{\frac{1}{\bar{\epsilon}_T L}\left(1+\frac{\bar{N}(1-\alpha)}{1-\bar{\epsilon}_T}\right) + \left(1+\frac{\bar{N}}{\bar{\epsilon}_T L}\right)\left(\frac{1}{(1-\bar{\epsilon}_T)}\right)\right\} \tag{44}$$

Next we compute the optimal parameters for large L. We first consider the case where $L \precsim S_{peak}$ with the following assumptions as $L \to \infty$,

(A1) $\bar{\epsilon}_T \to 0,\ \bar{N} \to \infty,$ and, $\frac{\bar{N}}{\bar{\epsilon}_T L} \to 0.$

We proceed to obtain the first-order properties of the optimal parameters and subsequently verify that (A1) is in fact a necessary condition for optimality. Since $L \precsim S_{peak}$, we have $\frac{\bar{\epsilon}_T}{\bar{N} S_{peak}} < \frac{1}{L}$ and hence $\alpha = \frac{1}{L}$, which implies $\frac{\partial \alpha}{\partial \bar{\epsilon}_T} = 0$, $\frac{\partial \alpha}{\partial \bar{N}} = 0$, $\alpha \frac{\partial C}{\partial \bar{N}} \prec \frac{\partial C}{\partial \bar{N}}$ and $\alpha \frac{\partial C}{\partial \bar{\epsilon}_T} \prec \frac{\partial C}{\partial \bar{\epsilon}_T}$. The condition (41) therefore implies

$$\frac{\bar{N}}{\bar{\epsilon}_T^2 L}\left(1 + \frac{\bar{N}(1-\alpha)}{1-\bar{\epsilon}_T}\right) \asymp \frac{\bar{N}(1-\alpha)}{(1-\bar{\epsilon}_T)^2}, \tag{45}$$

or equivalently,

$$\bar{N} \asymp \bar{\epsilon}_T^2 L. \tag{46}$$

Using the expansion

$$\gamma(u) = e^{-u}\left[\frac{1}{u} - \frac{1}{u^2} + \frac{2}{u^3} - O\left(\frac{1}{u^4}\right)\right], \tag{47}$$

the condition (42) becomes

$$\frac{1}{x} - \frac{1}{x^2} + O\left(\frac{1}{x^3}\right) \asymp \bar{N}\left(\frac{1}{x^2} - \frac{2}{x^3} + O\left(\frac{1}{x^4}\right)\right)\left[\frac{1}{(1-\bar{\epsilon}_T)} + \frac{(2-\alpha)\bar{N}}{\bar{\epsilon}_T L(1-\bar{\epsilon}_T)} + \frac{1}{\bar{\epsilon}_T L}\right], \tag{48}$$

which from (13) simplifies to

$$\frac{1}{x} \asymp \frac{2\bar{N}}{\bar{\epsilon}_T L} \qquad \Rightarrow \qquad \bar{\epsilon}_T L \asymp 2\bar{N}^2. \tag{49}$$

Combining (46) and (49) gives

$$\bar{\epsilon}_T^\star \asymp \frac{1}{(2L)^{1/3}}, \qquad \bar{N}^\star \asymp \frac{L^{1/3}}{2^{2/3}} \tag{50}$$

Substituting these first-order growth terms into the expression for the achievable rate gives

$$\begin{aligned}
\underline{C}_{nfb} &= S(1-\alpha)\bar{N}e^x\gamma(x) & (51)\\
&= S(1-\alpha)\frac{\bar{N}}{x}\left[1 - \frac{1}{x} + \frac{2}{x^2} - O\left(\frac{1}{x^3}\right)\right] & (52)\\
&= S\bar{N}\left(\frac{1}{x} - \frac{1}{x^2}\right) + o\left(\frac{1}{L^{1/3}}\right) & (53)
\end{aligned}$$

and substituting for x from (13), using the first-order values in (50), gives

$$\underline{C}_{nfb} = S(1-\bar{\epsilon}_T - \frac{1}{\bar{N}}) + o\left(\frac{1}{L^{1/3}}\right) \quad (54)$$

$$= S - S\left(\frac{3}{2^{1/3}}\right)\frac{1}{L^{1/3}} + o\left(\frac{1}{L^{1/3}}\right). \quad (55)$$

Now we show that the conditions in (A1) are in fact necessary for optimality. Suppose that the optimal training power is bounded away from zero as $L \to \infty$, that is, $\bar{\epsilon}_T \geq \bar{\epsilon}_o > 0$. We then have $P_D \leq P(1-\bar{\epsilon}_o)$, and applying Jensen's inequality to (11) gives $\underline{C}_{nfb} \leq S(1-\bar{\epsilon}_o)$. This contradicts the fact that, by choosing parameter values as in (50), we can achieve rates arbitrarily close to S as L becomes large. Hence for optimality we must have $\bar{\epsilon}_T \to 0$ as $L \to \infty$.

A similar argument shows that if $\frac{\bar{\epsilon}_T L}{\bar{N}}$ is bounded from above, which from (4) implies that the estimation error σ_e^2 is bounded away from zero, or if N is bounded from above, then $\underline{C}_{nfb}$ will be bounded away from S. Hence, for optimality $\bar{N} \to \infty$ and $\frac{\bar{\epsilon}_T L}{\bar{N}} \to \infty$ as $L \to \infty$, and (50) and (55) give the first-order asymptotic growth rate of the optimal parameters and capacity, respectively, when the peak power lies in the range $L \precsim S_{peak}$. In addition, since we obtain a unique solution to the optimality conditions (41)-(42), the results must correspond to the global maximum.

Next we show that the parameter values (50) are asymptotically optimal when the peak training power lies in the range $\frac{1}{L^{1/3}} \prec S_{peak} \precsim L$. From (50) the optimal training fraction α, given by (12), is $\frac{1}{L}$ if $L^{1/3} \prec S_{peak} \precsim L$, and is $\frac{2^{1/3}}{L^{2/3} S_{peak}}$ if $\frac{1}{L^{1/3}} \prec S_{peak} \precsim L^{1/3}$. Substituting these values of α and (50) into (51) gives the first-order behavior for the capacity $\underline{C}_{nfb}$ shown in (55). Since $\underline{C}_{nfb}$ is an increasing function of S_{peak}, the convergence rate to the limit S as $L \to \infty$ in (55), which corresponds to an infinite S_{peak}, is the best achievable. Hence we conclude that choosing the parameters in (50) achieves the asymptotic optimality condition (18) when $\frac{1}{L^{1/3}} \prec S_{peak} \precsim L$. This proves the theorem.

Furthermore, for $\frac{1}{L^{1/3}} \prec S_{peak} \precsim L$, it is easy to check that with the parameter values in (50), we have $-\frac{\partial \alpha}{\partial \bar{\epsilon}_T} C - \alpha \frac{\partial C}{\partial \bar{\epsilon}_T} \prec \frac{\partial C}{\partial \bar{\epsilon}_T}$ and $-\frac{\partial \alpha}{\partial \bar{N}} C - \alpha \frac{\partial C}{\partial \bar{N}} \prec \frac{\partial C}{\partial \bar{N}}$. It is then easy to verify from (43)-(44) that $\frac{\partial C}{\partial \bar{\epsilon}_T} \to 0$ and $\frac{\partial C}{\partial \bar{N}} \to 0$ as $L \to \infty$.

Appendix II
Proof of Theorem 2

The achievable rate can be written as $\underline{C}_{fb} = S(1-\alpha)C$ where

$$C = \bar{N}e^{-\bar{t}_o\theta}\left[\log\left(\frac{y}{y-\bar{t}_o\theta}\right) + e^y\gamma(y)\right] \tag{56}$$

and

$$y = \theta\left[1 + \frac{\bar{N}(1-\alpha)}{1-\bar{\epsilon}_T}e^{-\bar{t}_o\theta}\right] - 1 + \bar{t}_o\theta, \qquad \theta = 1 + \frac{\bar{N}}{\bar{\epsilon}_T L}, \qquad \alpha = \max\left\{\frac{\bar{\epsilon}_T}{\bar{N}S_{peak}}, \frac{1}{L}\right\}. \tag{57}$$

We consider the asymptotic limit $L \to \infty$ where $L \precsim S_{peak}$. In addition, we assume that the optimal parameters satisfy (A1) in Appendix I with the additional assumption

(A2) $\bar{t}_o \to \infty$, $\bar{N}e^{-\theta\bar{t}_o} \to \infty$ and, $\bar{t}_o \prec \bar{N}e^{-\theta\bar{t}_o}$.

We will later verify that these assumptions are in fact necessary conditions for optimality. As in Appendix I, for this range of peak powers we have $\alpha = \frac{1}{L}$. Setting the partial derivatives of $\underline{C}_{fb}$ with respect to $\bar{N}$, $\bar{\epsilon}_T$, and $\bar{t}_o$ to zero, and using assumptions (A1) and (A2) gives the necessary conditions

$$\frac{1}{1-\bar{\epsilon}_T} \asymp \frac{(1-\bar{\epsilon}_T)}{\bar{\epsilon}_T^2 L\theta e^{-\bar{t}_o\theta}}\left[1 + \frac{\bar{N}e^{-\bar{t}_o\theta}}{1-\bar{\epsilon}_T}\right] \tag{58}$$

$$-\log\left(\frac{y}{y-\bar{t}_o\theta}\right) \asymp \frac{\theta\bar{N}e^{-\bar{t}_o\theta}}{1-\bar{\epsilon}_T}\left[e^y\gamma(y) - \frac{1}{(y-\bar{t}_o\theta)}\right] \tag{59}$$

$$-\log\left(\frac{y}{y-\bar{t}_o\theta}\right) + \left(\frac{\theta\bar{N}e^{-\bar{t}_o\theta}}{1-\bar{\epsilon}_T}\right)\frac{1}{(y-\bar{t}_o\theta)} \asymp \frac{\bar{N}}{\bar{\epsilon}_T L}\left[1 + \frac{\bar{N}e^{-\bar{t}_o\theta}}{1-\bar{\epsilon}_T}\right]\log\left(\frac{y}{y-\bar{t}_o\theta}\right). \tag{60}$$

Combining (59) and (60) with (A1) and (A2) gives

$$e^y\gamma(y) \asymp \frac{\bar{N}}{\bar{\epsilon}_T L}\log\left(\frac{y}{y-\bar{t}_o\theta}\right). \tag{61}$$

Furthermore, (58) reduces to

$$\bar{\epsilon}_T^2 L \asymp \bar{N}. \tag{62}$$

Define the threshold in terms of the variable v as

$$\bar{t}_o = \frac{1}{\theta}\log\left(\frac{\bar{N}}{v}\right) \tag{63}$$

so that

$$y = \theta\left[1 + \frac{v(1-\alpha)}{(1-\bar{\epsilon}_T)}\right] - 1 + \log\left(\frac{\bar{N}}{v}\right) \asymp v + \log\left(\frac{\bar{N}}{v}\right). \tag{64}$$

To satisfy (59), we can choose v to satisfy

$$\log\left(\frac{v}{v+\log(\bar{N}/v)}\right) = v\exp(v+\log(\bar{N}/v))\gamma(v+\log(\bar{N}/v)). \tag{65}$$

It can be shown that the solution to (65) has the form $v=(\log\bar{N})^{1+\delta}$ where $\delta\in(0,1)$. Also, from (A1) we have $\frac{\bar{N}}{\bar{\epsilon}_T L}\to 0$ which implies $\theta\to 1$. Combining these, (63) implies that the optimal threshold satisfies $\bar{t}_o \asymp \log\bar{N}$.

Using the fact that $\bar{t}_o/y\to 0$ (from (A2)) and $\gamma(y)$ can be expanded as (47), (61) simplifies to

$$\begin{aligned}
\frac{1}{y}-\frac{1}{y^2}+O\left(\frac{1}{y^3}\right) &\asymp -\frac{\bar{N}}{\bar{\epsilon}_T L}\log\left(1-\frac{\bar{t}_o\theta}{y}\right) && (66)\\
\Rightarrow \qquad \frac{1}{y} &\asymp -\frac{\bar{N}}{\bar{\epsilon}_T L}\left(\frac{\bar{t}_o\theta}{y}\right) && (67)\\
\Rightarrow \qquad \frac{\bar{\epsilon}_T L}{\bar{N}} &\asymp \bar{t}_o. && (68)
\end{aligned}$$

Finally combining (62) and (68) we have

$$\bar{\epsilon}_T \asymp \frac{1}{\bar{t}_o} \tag{69}$$

$$\frac{L}{\log^2\bar{N}} \asymp \bar{N}, \tag{70}$$

which gives the asymptotic optimal parameters

$$\bar{\epsilon}_T^\star \asymp \frac{1}{\log\bar{N}} \tag{71}$$

$$\bar{N}^\star \asymp \frac{L}{\log^2 L}. \tag{72}$$

Substituting (71) and (72) into (63) gives the optimal threshold (35).

To compute the growth rate for the capacity, we rewrite (26) as

$$\underline{C}_{fb} = S(1-\alpha)\bar{N}e^{-\bar{t}_o\theta}\left[-\log\left(1-\frac{\bar{t}_o\theta}{y}\right)+\frac{1}{y}-\frac{1}{y^2}+O\left(\frac{1}{y^3}\right)\right]. \tag{73}$$

Substituting the asymptotic values for the optimal parameters, we have

$$\underline{C}_{fb} = -S(1-\alpha)\bar{N}e^{-\bar{t}_o\theta}\log\left(1-\frac{\bar{t}_o\theta}{y}\right)+S-o(S) \tag{74}$$

$$= -S\bar{N}e^{-\bar{t}_o\theta}\log\left(1-\frac{\bar{t}_o\theta}{y}\right)+S-o(S) \tag{75}$$

$$= -Sv\log\left(1-\frac{\bar{t}_o\theta}{y}\right)+S-o(S) \tag{76}$$

$$= Sv\left[\frac{\bar{t}_o\theta}{y}+O\left[\left(\frac{\bar{t}_o\theta}{y}\right)^2\right]\right]+S-o(S) \tag{77}$$

$$= S\log\bar{N}+o(\log\bar{N}). \tag{78}$$

Now we show that (A1) and (A2) are necessary for optimality. Using (25), we can bound the achievable rate as

$$S\,\bar{N}e^{-\bar{t}_o\theta}\ \log\left(1+\frac{(1-\bar{\epsilon}_T)\bar{t}_o}{1+\bar{N}e^{-\bar{t}_o\theta}}\right)\leq\ \underline{C}_{fb}\ \leq S\ \bar{N}e^{-\bar{t}_o\theta}\ \log\left(1+\frac{(1-\bar{\epsilon}_T)(\bar{t}_o+1)}{\bar{N}e^{-\bar{t}_o\theta}}\right), \tag{79}$$

where the upper bound follows from Jensen's inequality and the fact that $\int_{t_o}^{\infty} tf(t)dt \leq t_o+\sigma_h^2$. Recall that we have $\bar{\epsilon}_T \in [0,1]$ and $\theta \geq 1$. If the optimal $\bar{N}$ is bounded from above as $L \to \infty$, then irrespective of the choice of $\bar{\epsilon}_T$ and $\bar{t}_o$, the upper bound in (79) is always less than a constant. This contradicts the fact that the capacity is unbounded as $L \to \infty$ (as in (78)) when $\bar{N}$ satisfies (72). We therefore conclude that the optimal $\bar{N} \to \infty$ as $L \to \infty$. Furthermore, given that $\bar{N} \to \infty$, choosing the threshold such that $\bar{t}_o \to \infty$, $\bar{N}e^{-\bar{t}_o\theta} \to \infty$ and $\frac{\bar{t}_o}{\bar{N}e^{-\bar{t}_o\theta}} \to 0$ simultaneously maximizes the lower and upper bounds in (79) and hence maximizes the achievable rate.

If $\bar{\epsilon}_T \geq \bar{\epsilon}_{To} > 0$, then by computing the value of $\bar{t}_o$, which maximizes the upper bound in (79), and substituting this value into the corresponding capacity lower bound, we can show that the maximum achievable growth rate for capacity is $S(1-\bar{\epsilon}_{To})\log\bar{N}$ as $\bar{N} \to \infty$. This is less than the asymptotic achievable rate in (78), which shows that a pre-$\log\bar{N}$ factor of S can be achieved. To maximize the asymptotic achievable rate we must therefore have $\bar{\epsilon}_T \to 0$ as $L \to \infty$.

Similarly, using the bounds in (79) we can show that if $\frac{\bar{N}}{\bar{\epsilon}_T L} \geq \kappa > 0$, then the maximum achievable growth rate for capacity is $\frac{S}{(1+\kappa)}\log\bar{N}$. Again the pre-log factor is less than the achievable factor of S, so that we must have $\frac{\bar{N}}{\bar{\epsilon}_T L} \to 0$.

We have therefore established that (A1) and (A2) are necessary conditions for optimality. Hence (63), (71), (72) and (78) give the first-order asymptotic growth rates of the optimal parameters and the capacity when the peak power satisfies $L \precsim S_{peak}$. In addition, since we obtain a unique solution to the optimality conditions, the results correspond to the global maximum.

Next we show that the parameter values (32), (33) and (35) are asymptotically optimal when the peak training power lies in the range $\frac{\log^2 L}{L} \prec S_{peak} \precsim L$. From (32), (33) and (35), the optimal training fraction α given by (57) is $\frac{1}{L}$ if $\log L \prec S_{peak} \precsim L$ and is $\frac{\log L}{LS_{peak}}$ if $\frac{\log^2 L}{L} \prec S_{peak} \precsim \log L$. Substituting these values of α and the optimal parameters into (73) shows that the capacity $\underline{C}_{fb}$ satisfies the first-order behavior in (78). Since $\underline{C}_{fb}$ is an increasing function of S_{peak}, the growth rate as $L \to \infty$ in (78), which corresponds to an infinite S_{peak}, is the best achievable. Hence we conclude that selecting the parameters in (32), (33), and (35) satisfies the asymptotic optimality condition (30) when $\frac{\log^2 L}{L} \prec S_{peak} \precsim L$. This proves the theorem.

Furthermore, if S_{peak} satisfies $\frac{\log^2 L}{L} \prec S_{peak} \precsim L$, then it is easy to show that (32), (33) and (35) imply $-\frac{\partial \alpha}{\partial \bar{\epsilon}_T} C - \alpha \frac{\partial C}{\partial \bar{\epsilon}_T} \prec \frac{\partial C}{\partial \bar{\epsilon}_T}$, $-\frac{\partial \alpha}{\partial \bar{N}} C - \alpha \frac{\partial C}{\partial \bar{N}} \prec \frac{\partial C}{\partial \bar{N}}$, and $-\frac{\partial \alpha}{\partial \bar{t}_o} C - \alpha \frac{\partial C}{\partial \bar{t}_o} \prec \frac{\partial C}{\partial \bar{t}_o}$. It is then easy to verify that the conditions (58)-(60) are satisfied, which establishes (37)-(39).

REFERENCES

[1] M. Medard and R. Gallager, "Bandwidth scaling for fading multipath channels," *Information Theory, IEEE Transactions on*, vol. 48, no. 4, pp. 840–852, Apr 2002.

[2] M. Medard and D. N. C. Tse, "Spreading in block-fading channels," in *Asilomar Conference on Signals, Systems and Computers*, vol. 2, 2000, pp. 1598–1602.

[3] D. Schafhuber, H. Boleskei, and G. Matz, "System capacity of wideband ofdm communications over fading channels without channel knowledge," *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, pp. 389–389, 2004.

[4] I. Telatar and D. Tse, "Capacity and mutual information of wideband multipath fading channels," *Information Theory, IEEE Transactions on*, vol. 46, no. 4, pp. 1384–1400, Jul 2000.

[5] S. Verdu, "Spectral efficiency in the wideband regime," *Information Theory, IEEE Transactions on*, vol. 48, no. 6, pp. 1319–1343, Jun 2002.

[6] Y. Sun and M. L. Honig, "Asymptotic capacity of multi-carrier transmission over a fading channel with feedback," *IEEE Trans. Inform. Theory.*, vol. 54, no. 7, pp. 2879–2902, July 2008.

[7] S. Sanayei and A. Nosratinia, "Opportunistic downlink transmission with limited feedback," *IEEE Trans. Inform. Theory.*, vol. 53, no. 11, pp. 4363–4372, Nov. 2007.

[8] S. Borade and L. Zheng, "Wideband fading channels with feedback," in *Allerton Conference on Communication, Control and Computing, Urbana, IL, USA,*, September, 2004.

[9] G. Hariharan and A. M. Sayeed, "Minimum probability of error in sparse wideband channels," in *Allerton Conference on Communication, Control and Computing, Urbana, IL, USA,*, September, 2006.

[10] A. Sabharwal, A. Khoshnevis, and E. Knightly, "Opportunistic spectral usage: Bounds and a multi-band csma/ca protocol," *Networking, IEEE/ACM Transactions on*, vol. 15, no. 3, pp. 533–545, June 2007.

[11] S. Guha, K. Munagala, and S. Sarkar, "Jointly optimal transmission and probing strategies for multi-channel wireless systems," in *Conference on Information Sciences ad Systems(CISS)*, Princeton, NJ, Mar. 2006.

[12] J. Singh and C. Rose, "Channel probing under power budget," in *Conference on Information Sciences ad Systems(CISS)*, Princeton, NJ, Mar. 2006.

[13] M. C. Gursoy, "An energy efficiency perspective on training for fading channels," in *International Symposium on Information Theory*. Nice, France, Jun. 2007.

[14] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *Information Theory, IEEE Transactions on*, vol. 46, no. 3, pp. 933–946, May 2000.

[15] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory.*, vol. 49, no. 4, pp. 951–963, April 2003.

[16] T. Klein and R. Gallager, "Power control for the additive white gaussian noise channel under channel estimation errors," *Information Theory, 2001. Proceedings. 2001 IEEE International Symposium on*, pp. 304–, 2001.

[17] D. Love and J. Heath, R.W., "Ofdm power loading using limited feedback," *Vehicular Technology, IEEE Transactions on*, vol. 54, no. 5, pp. 1773–1780, Sept. 2005.

[18] M. Agarwal, D. Guo, and M. L. Honig, "Multi-carrier transmission with limited feedback: Power loading over sub-channel groups." in *IEEE International Conference on Communication*. Beijing, China, May 2008.

[19] M. Agarwal, M. Honig, and B. Ata, "Adaptive allocation of pilot and data power for time-selective fading channels with feedback," in *Information Theory, 2006 IEEE International Symposium on*, July 2006, pp. 168–172.

[20] M. Agarwal and M. L. Honig, "Spectrum sharing on a wideband fading channel with limited feedback," in *Proc. CrownCom Conf.*, Orlando, Florida, Aug. 2007.

[21] W. Zhang and U. Mitra, "On outage behavior of wideband slow-fading channels," *IEEE Trans. Inform. Theory.*, 2007, submitted.