Wideband Fading Channel Capacity with Training and Partial Feedback*

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

September 27, 2005

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 (M channel uses) and a large number of independent, finite coherence bands. 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 a pilot sequence is transmitted during each coherence time for channel estimation, and that feedback is used to designate a subset of coherence bands on which to transmit. Our problem is to optimize jointly the training length, average training power, and spreading bandwidth, taking into account the channel estimation error. We do this by maximizing a lower bound on the ergodic capacity. This lower bound becomes tight for large M, and we show that it increases as $O(\log M)$. The capacity of the partial feedback scheme therefore exceeds the capacity of "flash" signaling when M exceeds a (positive) threshold value.

1 Introduction

It has been shown for various channel models and input constraints that the capacity of a time-varying wideband fading channel goes to zero as the signal is spread across an increasing number of dimensions (e.g., coherence times and bands) [3–5]. This is due to the inability of the receiver to obtain a satisfactory estimate of the channel as the signal power per dimension decreases. To avoid this behavior, it is necessary to use "peaky", or "flash" signaling [1, 2], which constrains the amount of signal spreading in time and frequency.

In this paper we consider a doubly block Rayleigh fading channel model, and a training-based scheme for channel estimation. Namely, we assume that the channel is partitioned in frequency into multiple (flat fading) coherence bands, each of which experiences block Rayleigh fading over successive coherence times. The channel gains are

*This work was supported by ARO under grant DAAD19-99-1-0288, NSF under grant CCR-0310809, and the Motorola-Northwestern Center for Telecommunications.

assumed to be independent across coherence bands and coherence times. A sequence of pilot symbols are transmitted at the begining of each coherence time, which enables the receiver to obtain channel estimates over some subset of coherence bands.

The channel estimates provide an opportunity for adapting the transmitted signal, given a feedback channel. We consider a feedback scheme whereby the receiver indicates to the transmitter a subset of coherence bands over which to transmit (i.e., one feedback bit per estimated channel coefficient). The transmitted power is then uniformly spread over that subset ("on-off" power allocation). We optimize the spreading bandwidth with and without feedback by maximizing a lower bound on the capacity over the training length, training power, and threshold for activating a channel. 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.

Without feedback, in the infinite bandwidth limit, 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 [1,2]. The lower bound on capacity for the pilot-based scheme increases with coherence time, and approaches the capacity with flash signaling only as the coherence time tends to infinity.

Related work (which does not consider feedback) is presented in [4], in which a range of optimal spreading bandwidths is determined for a doubly-selective *i.i.d.* block fading model. Those results are derived under a peak power constraint, as opposed to the average power constraint considered here, and do not assume the availability of a pilot for channel estimation.

We show that the capacity of the pilot-based scheme with partial feedback increases as $O(\log M)$, where M is the coherence time in number of channel uses. This is consistent with the observations made in [10], which considers a similar type of feedback scheme for a Rayleigh fading channel at low SNRs. Hence there is a critical coherence time, measured in number of channel uses, beyond which the partial feedback scheme performs better than flash signaling. For the model considered, this 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). Without feedback we show that the difference between the capacity of the pilot-based scheme and the capacity with flash signaling tends to zero as $1/o(M)$.

2 System Model

We consider a time-varying wideband Rayleigh fading channel. To model finite coherence bandwidth, we assume that the channel is divided into equal frequency slices (subchannels) with bandwidth B, where each experiences *i.i.d.* flat Rayleigh fading. Since this is a wideband system, we assume an infinite number of available parallel subchannels. We will see, however, that the transmitter should always transmit on a finite subset of channels. Each subchannel symbol has duration T_s, which is approximately equal to $\frac{1}{B}$. We will refer to this duration as one channel use. To model finite coherence time, an *i.i.d.* block fading model in time is assumed with a block length of M channel uses (coherence time = $M\,T_s$). That is, subchannel fading coefficients remain constant for M channel uses and then change to a new independent value.

We impose an average energy constraint P per channel use, hence the (normalized) average power constraint is also P. The coherence block length is divided into training, which is placed at the begining, and data transmission. 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. The receiver obtains a Minimum Mean Squared Error (MMSE) estimate of the channel, based on the pilot symbols transmitted during training, and uses that estimate for feedback and data detection. The feedback determines a power allocation across subchannels for the data transmission. This procedure is repeated for each coherence block.

The transmitter sends pilot symbols over a finite number of subchannels K. Those channels are referred to as being "active". (Note that a subset of active channels will carry transmitted data.) Since subchannels are assumed to be *i.i.d.*, the channel estimation can be performed separately for each active subchannel by equally dividing the training power among them. Let T denote the training length in number of channel uses, P_T denote the training power during the training period, and D and P_D denote the data length and average data power, respectively. Then we have that

$$\begin{aligned} T + D &= M \qquad (1) \\ \alpha\, P_T + (1-\alpha)\, P_D &= P \qquad (2) \end{aligned}$$

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

For the i^{th} active subchannel, the transmitted vector symbol during the data part of coherence block is represented by $\mathbf{X}_i = (x_{i1}\, x_{i2} \ldots x_{iD})^T$, and the corresponding received vector symbol, $\mathbf{Y}_i$ has a similar form. We omit the dependence on the coherence time index for convenience. We have that

$$\mathbf{Y}_i = h_i\, \mathbf{X}_i + \mathbf{N}_i \qquad (3)$$

where $\mathbf{N}_i$ is a $D \times 1$ zero-mean, circularly symmetric, complex Gaussian (CSCG) noise vector with covariance $E(\mathbf{N}_i\, \mathbf{N}_i^H) = {\sigma_n}^2\, \mathbf{I}$ and h_i is the i^{th} subchannel fading coefficient. We will assume that the channel gains during a coherence time, $\{h_i\}$, are also *i.i.d.* CSCG random variables with mean zero and variance ${\sigma_h}^2$. Each channel gain changes independently across the coherence blocks. Also, the additive noise random variables across subchannels are independent.

Based on the training segment of the coherence block, the receiver obtains a channel estimate $\hat{h}_i$ with error e_i. Since $h_i = \hat{h}_i + e_i$, we rewrite (3) as

$$\mathbf{Y}_i = \hat{h}_i\, \mathbf{X}_i + e_i\, \mathbf{X}_i + \mathbf{N}_i \qquad (4)$$

where $\hat{h}_i$ and e_i are uncorrelated, zero-mean, complex Gaussian. The error variance is therefore ${\sigma_e}^2 = E(|h_i|^2) - E(|\hat{h}_i|^2)$. Since each pilot symbol has power P_T/K and the training length is T symbols, we can write the MMSE, or error variance, as

$${\sigma_e}^2 = {\sigma_h}^2 - {\sigma_h}^4 \left(\frac{T\, P_T}{{\sigma_h}^2\, T\, P_T + K\, {\sigma_n}^2} \right). \qquad (5)$$

3 Capacity Performance Objective

In what follows we adopt a capacity performance metric to optimize the training power P_T, data power P_D, fraction of training symbols α, and number of active channels K. Since the capacity achieving input distribution with imperfect channel knowledge is not known [7], we derive a lower bound on capacity. This lower bound is the rate achieved with a Gaussian input distribution and a coherent linear receiver, which uses the MMSE channel estimate based on the pilot symbols. In particular, the receiver does not attempt to improve upon the channel estimate during the data reception period.

¿From (4), the ergodic capacity is given by[1]

$$C = (1-\alpha)\frac{1}{D}\sum_{i=1}^{K} max_{p(\mathbf{X}_i|\hat{h}_i)} I(\mathbf{X}_i;\mathbf{Y}_i|\hat{h}_i) \tag{6}$$

$$Subject\, to: \sum_{i=1}^{K} E_{\hat{h}_i}\left[tr(\mathbf{Q}_{\hat{h}_i})\right] \leq P_D\, D$$

where $p(\mathbf{X}_i|\hat{h}_i)$ is the probability distribution of $\mathbf{X}_i$ given $\hat{h}_i$, and $\mathbf{Q}_{\hat{h}_i} = E[\mathbf{X}_i\mathbf{X}_i^\dagger|\hat{h}_i]$. We compute a lower bound on the mutual information $I(\mathbf{X}_i;\mathbf{Y}_i|\hat{h}_i)$, and subsequently on capacity. The input distribution that maximizes the mutual information is unknown; however, a lower bound is obtained by assuming that $p(\mathbf{X}_i|\hat{h}_i)$ is Gaussian [6, 7]. With this assumption the differential entropy $h(\mathbf{X}_i|\hat{h}_i) = E_{\hat{h}_i}[\log(|\pi e\mathbf{Q}_{\hat{h}_i}|)]$. Also, $h(\mathbf{X}_i|\mathbf{Y}_i,\hat{h}_i)$ 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 $\mathbf{X}_i$ given $\mathbf{Y}_i$ and $\hat{h}_i$ [7]. After some manipulation and application of the matrix inversion lemma we obtain the lower bound

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

¿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

$$\underline{I}(\mathbf{X}_i;\mathbf{Y}_i|\hat{h}_i) = D\, E_{\hat{h}_i}\left[\log\left(1 + \frac{P(\hat{h}_i)|\hat{h}_i|^2}{P(\hat{h}_i)\,{\sigma_e}^2 + {\sigma_n}^2}\right)\right]. \tag{7}$$

Substituting this lower bound on mutual information into the capacity expression (6), we obtain the following lower bound on capacity given power allocation strategy $P(\hat{h}_i)$,

$$\underline{C} = (1-\alpha)\, K\, E_{\hat{h}_i}\left[\log\left(1 + \frac{P(\hat{h}_i)|\hat{h}_i|^2}{P(\hat{h}_i)\,{\sigma_e}^2 + {\sigma_n}^2}\right)\right] \tag{8}$$

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

Here we have used the assumption that the channel estimate $\hat{h}_i$ has the same distribution for all K active subchannels.

4 Optimal Parameters

The lower bound (8) depends on the transmitter power allocation strategy. We consider two scenarios. In the first, the transmitter uniformly distributes the power across all

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

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 channels for which the channel estimate $|\hat{h}_i|^2$ is above a threshold. (That is, the power across active channels is a constant.) We refer to this as an "on-off" power allocation. Although the optimal power allocation strategy, which maximizes (8), resembles water pouring [9], the "on-off" power allocation is simpler to analyze and is known to achieve near-optimal performance with perfect channel estimates [8].

4.1 No feedback

Since the power is spread evenly over all data symbols, $P(\hat{h}_i) = P_D/K$ for all $i = 1, 2, \ldots, K$. Also, $|\hat{h}_i|^2$ has an exponential distribution with expected value ${\sigma_{\hat{h}}}^2 = ({\sigma_h}^2 - {\sigma_e}^2)$, so that the lower bound on capacity (8) can be evaluated as

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

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

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

The capacity expression (9) depends on the design parameters α, P_T and K. Note that for any permissible value of α and P_T, as $K \to \infty$, $\underline{C}_{nfb} \to 0$. This is because as the number of active subchannels increases, the channel estimates degrade. Similarily, for fixed K, $\underline{C}_{nfb} \to 0$ at the constraint values on training fraction and training power. Hence we can jointly optimize α, P_T, and K to maximize $\underline{C}_{nfb}$.

We begin by optimizing the training fraction α. For this case, we write (8) as an integral, substitute for $P(\hat{h}_i)$, and subsequently for P_D from equation (2), to obtain

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

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

Note from (5) that ${\sigma_e}^2$ and hence ${\sigma_{\hat{h}}}^2$ depend on ϵ_T but not on α explicitly. This observation and the fact that $K\,(1-\alpha)\, \log(1+\frac{a}{b+K\,(1-\alpha)})$ is a decreasing function of α for all positive a and b implies that the integrand, and hence the integral, is maximized by letting $\alpha \to 0$ while keeping the average training power ϵ_T fixed. (Our results in Section 5 show that this dependence on α vanishes as K and M become large.)

We substitute the optimal value $\alpha \to 0$ in (9) and proceed to optimize K and ϵ_T. Let ${\epsilon_T}' = \frac{\epsilon_T}{P}$ and $K' = \frac{K {\sigma_n}^2}{P {\sigma_h}^2}$, and note that x in (10) is a function of ${\epsilon_T}'$, K' and M only. The capacity (9) can now be rewritten as

$$\underline{C}_{nfb} = \frac{P\, {\sigma_h}^2}{{\sigma_n}^2}\, K'\, e^x\, \gamma(x) \tag{12}$$

Since P and M are system parameters, maximizing (12) with respect to K and ϵ_T is equivalent to maximizing with respect to K' and ${\epsilon_T}'$. The optimal values depend only on M, i.e., $K'^\star = g_K(M)$ and ${\epsilon_T}'^\star = g_{\epsilon_T}(M)$, hence the optimal values of ϵ_T and K have the forms

$${\epsilon_T}^\star = P\, g_{\epsilon_T}(M) \qquad K^\star = P\, \frac{{\sigma_h}^2}{{\sigma_n}^2} g_K(M). \tag{13}$$

To evaluate the preceding functions of M, we note that $e^x \gamma(x)$ in (12) is a decreasing function of x, and depends on $\epsilon_T{}'$ only through x. Therefore the optimal value of $\epsilon_T{}'$ minimizes x, which gives

$$g_{\epsilon_T}(M) = -z + (z^2 + z)^{1/2} \text{ where } z = \frac{g_K(M) + 1}{M - 1}.$$

Also,

$$g_K(M) = \arg\max_{x_k} \left[x_k \, e^{x'} \gamma(x') \right]$$

where

$$x' = \left(1 + \frac{x_k}{1 + v - (v^2 + v)^{1/2}}\right)\left(1 + \frac{x_K}{((v^2 + v)^{1/2} - v)\, M}\right) - 1$$

and $v = (x_k + 1)/(M - 1)$. The lower bound on capacity is then

$$\underline{C}^{\star}_{nfb} = \frac{P\,{\sigma_h}^2}{{\sigma_n}^2}\, g_C(M) = C_{flash}\; g_C(M) \tag{14}$$

where $C_{flash} = \frac{P{\sigma_h}^2}{{\sigma_n}^2}$ is the capacity with peaky, or flash signaling, and is also the capacity with perfect channel knowledge at the receiver [2], and $g_C(M)$ is obtained by substituting $K^\star$ and $\epsilon_T^\star$ in (12). Plots of the functions $g_K(M)$, $g_{\epsilon_T}(M)$, and $g_C(M)$ are given in Section 6. Clearly, C_{flash} is an upper bound on the performance of the training-based scheme without feedback. The rate at which $\underline{C}^{\star}_{nfb}$ approaches this bound with M is discussed in Section 5.

4.2 Partial Feedback

We assume the 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 (8), the lower bound on ergodic capacity is

$$\underline{C}_{fb} = (1-\alpha)\, K \int_{t_0}^{\infty} \log(1 + \frac{P_o\, t}{P_o\,{\sigma_e}^2 + {\sigma_n}^2}) f(t) dt \tag{15}$$

with the power constraint

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

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

$$\underline{C}_{fb} = (1-\alpha)\, K\, e^{-\frac{t_0}{{\sigma_{\hat{h}}}^2}} \left[\log\left(\frac{y}{y - \frac{t_0}{{\sigma_{\hat{h}}}^2}} \right) + e^y \gamma(y) \right] \tag{16}$$

where

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

As before, we can maximize $\underline{C}_{fb}$ with respect to the training fraction (α), training power (P_T), on-off threshold (t_0) and number of subchannels (K). Following an argument analogous to that given in the preceding subsection, it can be shown that the optimal values satisfy

$$\alpha^\star \longrightarrow 0 \qquad {\epsilon_T}^\star = P\, f_{\epsilon_T}(M) \qquad K^\star = P\, \frac{{\sigma_h}^2}{{\sigma_n}^2} f_K(M) \qquad {t_0}^\star = {\sigma_h}^2\, f_t(M) \tag{17}$$

and the lower bound on capacity

$$\underline{C}^\star_{fb} = \frac{P\, {\sigma_h}^2}{{\sigma_n}^2}\, f_C(M). \tag{18}$$

where the "f" functions depend only on M. Here $K^\star$ is the total number of subchannels monitored with pilot symbols. Data is transmitted only on the subset of $K^\star$ channels with estimated channels gains $|\hat{h}_i|^2 \geq t_0^\star$. In contrast, with no feedback data is transmitted on all $K^\star$ subchannels. In either case, $K^\star$ represents the optimal spreading bandwidth. Although it appears to be difficult to evaluate the preceding optimal system values in closed-form, it is straightforward to evaluate them numerically. Plots of the preceding functions of coherence time M are given in Section 6.

5 Asymptotic Analysis

We now examine the behavior of the capacity as both the number of channels K and the coherence time M becomes large. Specifically, we can evaluate the lower bounds in (9) and (16) as $K \to \infty$ and $M \to \infty$ with fixed ratio $\frac{K}{M} = \beta$, which represents the spreading bandwidth. Using the fact that $u\, e^u\, \gamma(u) \to 1$ as $u \to \infty$, it is easily shown that (9) and (16) converge respectively to

$$\underline{C}^\infty_{nfb} = \frac{(P - \alpha P_T)\, {\sigma_{\hat{h}}}^2}{{\sigma_n}^2} \tag{19}$$

and

$$\underline{C}^\infty_{fb\infty} = \frac{(P - \alpha P_T)\, {\sigma_{\hat{h}}}^2}{{\sigma_n}^2} + \frac{(P - \alpha P_T)\, t_0}{{\sigma_n}^2} \tag{20}$$

where the variance of the channel estimate

$${\sigma_{\hat{h}}}^2 = {\sigma_h}^4 \left(\frac{\alpha\, P_T}{{\sigma_h}^2\, \alpha\, P_T + \beta\, {\sigma_n}^2} \right). \tag{21}$$

Comparing (19) and (20), the second term in (20) represents the gain due to feedback. Although this term can be arbitarily large, depending on the choice of t_0, the choice of t_0 influences how large M must be in order to achieve the corresponding rate in (20). This relation is clarified in the Theorem (and proof), which follows.

We observe that these asymptotic capacity expressions depend only on the training power $\alpha\, P_T$, and not on the particular value of α. Note that $\alpha\, P_T$ determines the quality of the channel estimate, reflected in the variance ${\sigma_{\hat{h}}}^2$. If we increase α, keeping the training power fixed, the loss in data transmission time is therefore exactly compensated by the increase in data symbol power. We also note that because $|1 - u\, e^u\, \gamma(u)|$ decreases as $O(1/u)$ with $u \to \infty$, we can show that the difference $\underline{C}^\infty_{nfb} - \underline{C}_{nfb}$ and $\underline{C}^\infty_{fb} - \underline{C}_{fb}$ with fixed t_0 both go to zero as $O(1/K)$.

We now consider maximizing the asymptotic capacity expressions over the average training power αP_T and β, which represents the spreading bandwidth. It turns out that both expressions are maximized by letting $\alpha P_T \rightarrow 0$ and $\beta \rightarrow 0$ while maintaining $\alpha P_T \sigma_h{}^2 \gg \beta \sigma_n{}^2$, so that the channel variance in (21) asymptotically converges to $\sigma_{\hat{h}}{}^2 = \sigma_h{}^2$, and the estimation error converges to zero. That is, both the average training power and number of channels per coherence time tend to zero in such a way that perfect channel estimates are obtained (asymptotically) for all active subchannels. With this choice of αP_T and β, the capacity without feedback approaches $C_{flash} = \frac{P \sigma_h{}^2}{\sigma_n{}^2}$. However to achieve this, since $\beta \rightarrow 0$, $K \rightarrow \infty$ sublinearly with M, i.e., $K = o(M)$. Hence, we conclude that with optimized training power and spreading bandwidth, $C_{flash} - \underline{C}_{nfb} \rightarrow 0$ as $1/o(M)$. This is illustrated numerically in the next section.

Returning to the scenario with partial feedback, we observe that although (20) is derived from the lower bound on capacity for finite K and M, the lower bound approaches the true capacity as the channel estimation error tends to zero (i.e., the receiver performs coherent detection given the channel). Hence letting $\alpha P_T \rightarrow 0$ in (20), according to the preceding discussion, implies that the asymptotic capacity with partial feedback becomes $\underline{C}_{fb}^{\infty} = C_{flash} + P t_0 / \sigma_n^2$ for any fixed t_0. To achieve this capacity, we must let $\beta \rightarrow 0$, and furthermore, the number of channels above the threshold t_0, given by $K\, e^{-t_0/\sigma_{\hat{h}}{}^2}$, must tend to infinity. (Otherwise, the capacity must converge to zero.) Combining these facts leads to the following Theorem.

Theorem 1 *In the asymptotic limit of large coherence time M, the capacity of the pilot-based scheme with partial feedback grows as $O(\log M)$.*

Proof. Since we must have $K\, e^{-\frac{t_0}{\sigma_{\hat{h}}{}^2}} \rightarrow \infty$ with K, t_0 cannot grow faster than $\sigma_{\hat{h}}{}^2 \log K$, and since $\beta \rightarrow 0$ implies $K = o(M)$, the capacity in (20) cannot grow faster than $O(\log M)$. Taking $t_0 = \sigma_{\hat{h}}{}^2 \log[K^{1-\epsilon_1}]$ with $\epsilon_1 > 0$ and $K = M^{1-\epsilon_2}$ with $\epsilon_2 > 0$ shows that the $O(\log M)$ growth is achievable. ■

6 Numerical Results and Capacity Comparison

Here we present numerical results obtained by optimizing α, P_T, and K directly in (9) and (16) for finite values of M. Fig. 1 shows plots of the optimal number of channels $K^{\star}$ and the optimal average training power $(\alpha P_T)^{\star}$ normalized by the appropriate factors. Results for both scenarios with and without feedback are shown. (That is, the plots show $g_{\epsilon_T}(M)$, $g_K(M)$ in (13) and $f_{\epsilon_T}(M)$, $f_K(M)$ in (17) versus M.) As the coherence time M increases, we can obtain more accurate channel estimates while spending less training power per subchannel. Hence, the optimal training power decreases with coherence time, and the optimal number of active subchannels increases. This allows for an increase in data power and number of subchannels used for data transmission, and hence the capacity increases with M.

Fig. 2 shows plots of capacity versus M for the schemes considered in this paper. We have plotted the normalized values $g_C(M)$ and $f_C(M)$, so that unity corresponds to the capacity with peaky, or flash signaling. As stated in the Theorem, the lower bound on capacity with feedback increases as $\log M$, asymptotically, whereas the corresponding capacity without feedback approaches C_{flash} (a constant independent of M).

Beyond a critical value of M, say M_{crit}, the feedback scheme gives a higher capacity than flash signaling. Since $\underline{C}_{fb}^{\star} = \underline{C}_{flash}\, f_C(M)$, it follows that M_{crit} satisfies

$f_C(M_{crit}) = 1$. Solving this numerically gives $M_{crit} \approx 120$, which, perhaps surprisingly, is independent of the system parameters. That is, 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 capacity than the optimal peaky 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 peaky signaling without feedback.

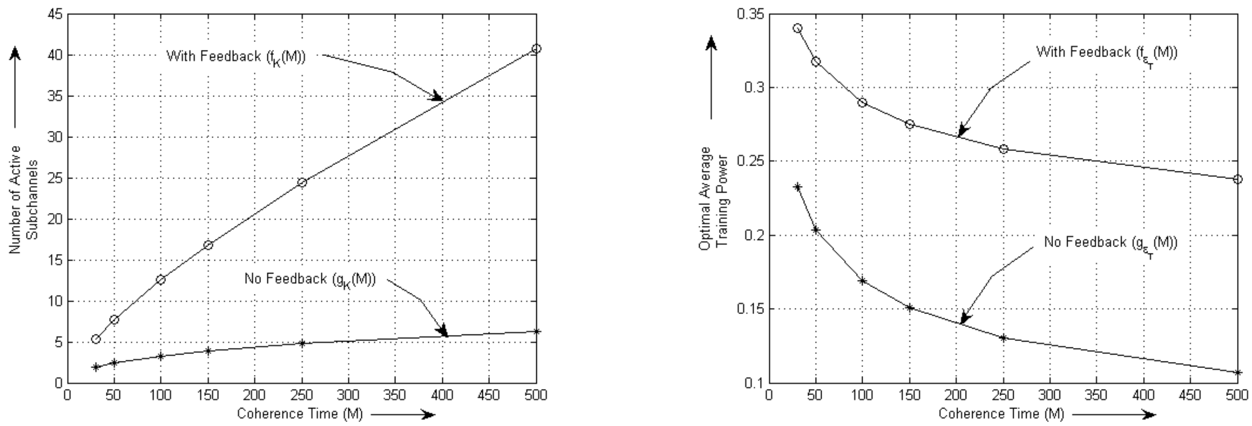

Figure 1: Optimal number of active subchannels (normalized) and optimal average training power (normalized) versus coherence time M.

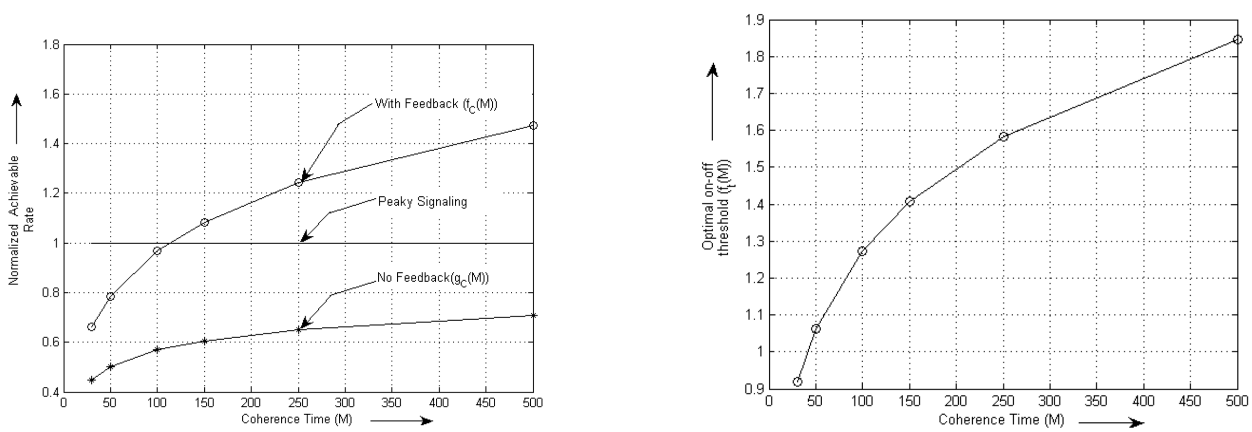

Figure 2: Achievable rates (normalized) for various signaling schemes versus coherence time M, and optimal on-off threshold value (normalized) with partial feedback versus M.

Finally, Fig. 2 also shows the normalized optimal threshold $f_t(M)$ for the feedback scheme versus M. As discussed in the previous section, for large M the optimal threshold grows as $\log M$. This relatively slow increase, combined with sublinear growth in K as a function of M, guarantees that that for large M, there are a large number of channels with gains that exceed the threshold.

7 Conclusions

We have considered a time-varying wideband system with pilot-assisted training and feedback. The performance of this scheme, relative to the performance of peaky, or flash signaling without feedback, depends critically on the coherence time. Namely, for the block *i.i.d.* Rayleigh fading channel model, the capacity with the feedback scheme considered grows as $\log M$ when M is large, and surpasses the capacity without feedback when M exceeds 120 channel uses. Other fading distributions can be analyzed within the framework presented, and, of course, may lead to quite different relative performance.

Although the on-off feedback scheme considered is known to have optimal properties [8], an open question is whether or not other finite-rate feedback schemes can achieve higher capacities. 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.

References

[1] I.E. Telatar and D.N.C. Tse, "Capacity and Mutual Information of Wideband Multipath Fading Channels ," *IEEE Trans. Inform. Theory*, vol. 46, No. 4, July 2000.

[2] S. Verdu, "Spectral Efficiency in the Wideband Regime ," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1319-1343 , June 2002.

[3] M. Medard and R. G. Gallager, "Bandwidth Scaling for Fading Multipath Channels," *IEEE Trans. Inform. Theory*, vol. 36, No. 4, April 2002.

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

[5] H. Boleski, D. Schafhuber and G. Matz, "System capacity of wideband ofdm communication over fading channels without channel knowledge,"*IEEE ISIT, Chicago*, July 2004.

[6] B. Hassibi and B. M. Hochwald, "How Much Training is Needed in Multiple- Antenna Wireless Links?," *IEEE Trans. Inform. Theory*, vol. 49, No. 4, pp. 951-963, April 2003.

[7] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, pp. 933-946, May 2000.

[8] Y.K. Sun, "Transmitter and Receiver Techniques for Wireless Fading Channels," *PhD thesis, Northwestern University, December 2004.*

[9] T. E. Klein and R. G. Gallager, "Power Control for the Additive White Gaussian Noise Channel under Channel Estimation Errors,"*IEEE ISIT, Washington, D.C.*, June 2001.

[10] S. Borade and L. Zheng, "Wideband Fading Channels with Feedback,"*Allerton Conference on Communication, Control, and Computing*, Urbana, IL, USA, September 2004.