

Spectrum Sharing on a Wideband Fading Channel with Limited Feedback

(Invited Paper)

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

Abstract—We consider a multiple access, doubly-selective block Rayleigh fading channel in which the users coordinate spectrum sharing through a limited feedback scheme. Each user probes a random set of sub-channels, known to the receiver, by sending a pilot sequence at the beginning of each coherence block. Multiple users may probe the same sub-channel, causing interference. The receiver assigns each sub-channel to the user with the highest estimated sub-channel gain (via limited feedback), provided that this gain exceeds a predetermined threshold. Our problem is to optimize the number of channels to probe, or “probing bandwidth”, for each user. We maximize a lower bound on the ergodic capacity, and consider a large system limit in which the system bandwidth and number of users scale linearly with the coherence time. We show that the optimal probing bandwidth grows as $O\left(\frac{N}{(\log N)^2}\right)$, assuming a linear Minimum Mean Square Error channel estimator, and the achievable rate increases as $O(\log \log N)$ per user, where N is the number of available subchannels. In contrast, if the users are pre-assigned nonoverlapping subchannels on which they probe and transmit, then the capacity per user converges to a constant as N becomes large. Additionally, the optimal training length and training power are computed and the effect of system load (number of users per unit coherence time) on the achievable rate is studied.

I. INTRODUCTION

The current scarcity of spectrum for many types of services can be alleviated by dynamically sharing spectrum across a multitude of services. That possibility motivates the consideration of “wideband” systems in which each user can choose from among a large number of coherence bands. A primary challenge when the users are non-cooperative is the mitigation and control of interference.

In this work we assume that the available spectrum is shared by several independent devices, which communicate synchronously with a central transceiver. Although the devices do not coordinate transmissions directly, indirect coordination takes place through a feedback channel. We focus on the multiple access uplink, and assume that the transmitting devices use multi-carrier signaling. Given many available sub-channels, each device wishes to choose a subset of sub-channels, which

avoids interference from other devices and exploits available frequency diversity.

The achievable rate for a doubly-selective fading channel depends on what channel state information (CSI) is available at the receiver and transmitter. Namely, CSI at the receiver can increase the rate by allowing coherent detection, and CSI at the transmitter allows adaptive allocation of rate and power across sub-channels (e.g., see [1, Ch. 6]) in addition to opportunistic scheduling across the users. This information is all the more important given a wideband fading channel, which offers many degrees of freedom for diversity. Obtaining CSI at the receiver and/or transmitter typically requires overhead in the form of a pilot signal and feedback. Hence there is a fundamental trade-off in allocating available resources between learning CSI and data transmission.

The channel from each user to the central receiver is assumed to be *i.i.d.* doubly-selective block fading, and is divided into several independent flat Rayleigh fading sub-channels. That is, each sub-channel remains static for a certain number of channel uses (coherence block), and then changes to a new independent value in the next block. Furthermore, sub-channel gains are assumed to be independent across the users. To simplify the analysis, we assume that the users are both symbol- and frame-synchronous, where each frame corresponds to a coherence block.

Each user selects a random subset of available subchannels, known to the receiver¹, to *probe*, i.e., send pilot sequences at the beginning of each coherence block. Multiple users may choose to probe a particular sub-channel, creating interference. The receiver estimates the sub-channel gain using the (possibly interfering) pilots and assigns the sub-channel for rest of the coherence block to the user with highest estimated gain via a feedback channel, provided that it is above a predetermined “on-off” threshold. We assume that the assignment process takes a negligible amount of time.

The preceding assumptions lead to the following tradeoff: Probing a large number of channels gives inaccurate channel

This work was supported by the U.S. Army Research Office under grant DAAD19-99-1-0288 and NSF under grant CCR-0310809

¹We have assumed that the choice of subset, although random, is shared between each transmitter and the central receiver. For example, we can assume a synchronized pseudo random generator for sub-channel selection.

estimates due to the increased likelihood of interference and also because the probing power per sub-channel is small. On the other hand, probing a small number of channels is likely to give accurate channel estimates, but does not exploit available degrees of freedom. This work is extension of [5], which studies the preceding tradeoff with a single user. Related work on single-user wideband models, which accounts for the effect of channel probing and the associated channel estimation error on the achievable rate, is presented in [7], [8]. Other models for multi-channel probing by a single user are studied in [9], [10].

We characterize the optimal number of sub-channels to probe, or *probing bandwidth*, for a large system in which the total number of sub-channels (coherence bands) N and number of users K scale linearly with the coherence block, consisting of L channel uses. Namely, the optimal probing bandwidth increases as $O\left(\frac{N}{(\log N)^2}\right)$, assuming the receiver computes an optimal Linear Minimum Mean Square Error (LMMSE) estimate of the channel, and as $O\left(\frac{N}{(\log \log N)(\log N)^2}\right)$ with a sub-optimal channel estimate obtained from a matched filter.² The corresponding capacity per user grows as $O(\log \log N)$, which is the same as the order-growth with perfect channel knowledge [6]. For the model considered, we characterize the second-order loss in capacity due to channel estimation. In contrast, if the users are pre-assigned to non-overlapping subchannels, then the capacity per user converges to a constant as N , K , and L become large. We also characterize the associated optimal on-off threshold, training power, and training length. A comparison of the asymptotic bounds with results from numerical optimization of finite-size systems shows that the asymptotic analysis accurately predicts performance only when the number of users K is large (e.g., a few hundred), although the asymptotic trends are visible for small K as well.

II. MODEL AND CAPACITY OBJECTIVE

We consider a synchronous multiple access channel model in which K users transmit through a wideband Rayleigh fading channel. The channel for each user is divided into N identical independent flat Rayleigh fading sub-channels. Furthermore, the wideband channel is assumed to be independent across the users. Block fading is assumed, so that the channel gains are constant within a coherence block of L symbols (for all users), and are independent across coherence blocks. At the beginning of a coherence block, each user chooses a random subset of sub-channels on which to probe, i.e., transmit a training sequence. In what follows, without any loss of generality, we do not indicate the dependence on coherence block index l explicitly and focus on sub-channel 1. For the l^{th} coherence block, let \mathcal{K} denote the set of indices of users who select sub-channel 1 to probe, and the corresponding number of users $|\mathcal{K}| = K'$. The received (scalar) signal on the sub-channel is given by

$$y = \mathbf{h}^\dagger \mathbf{s} + n \quad (1)$$

²The result with LMMSE channel estimation also applies to the single-user model considered in [5].

where \mathbf{s} is the $K' \times 1$ vector of transmitted symbols across users $k \in \mathcal{K}$, \mathbf{h} is the corresponding $K' \times 1$ vector of independent zero mean circularly symmetric complex Gaussian (CSCG) channel gains, each with variance σ_h^2 , and the noise n is also CSCG, zero-mean, and white with variance σ_n^2 .

Each coherence block consists of T training symbols followed by D data symbols, assuming the sub-channel is used for data transmission. We therefore partition the $1 \times L$ vector of received symbols on the sub-channel within coherence block l as

$$\mathbf{y} = [\mathbf{y}_T \ \mathbf{y}_D] = \mathbf{h}^\dagger [\mathbf{S} \ \mathbf{X}] + \mathbf{n} \quad (2)$$

where \mathbf{y}_T and \mathbf{y}_D are the $1 \times T$ and $1 \times D$ received vectors during training and data transmission, respectively, \mathbf{S} is the $K' \times T$ matrix of training symbols, \mathbf{X} is the $K' \times D$ matrix of transmitted data symbols, and \mathbf{n} is the $1 \times L$ vector of noise samples. Note that the rows of \mathbf{S} and \mathbf{X} contain the training and data symbols transmitted by users in \mathcal{K} . The training symbols have zero mean and variance $\frac{P_T}{N'}$ where P_T is the power during the training period for a particular user, which is split uniformly among $N' (\leq N)$ randomly probed sub-channels. The row of \mathbf{X} corresponding to user $k \in \mathcal{K}$, say \mathbf{x}_k , has data symbols with power (variance) P_D if user k is “assigned” the sub-channel (explained below), otherwise it has zero power. Given an average power constraint P_{av} , for each user, we have

$$\epsilon_T + q N (1 - \alpha) P_D = P_{av} \quad (3)$$

where $\epsilon_T = \alpha P_T$ is the average training power, $\alpha = T/L$ denotes the fraction of the coherence block devoted to training and, q is the probability that user k is assigned the sub-channel.

For each user $i \in \mathcal{K}$ with channel gain h_i , the receiver computes the corresponding channel estimate \hat{h}_i with estimation error $e_i = h_i - \hat{h}_i$. The error variance $\sigma_{e|K'}^2(i) = E[|e_i|^2 \mid K']$ in general depends upon the particular training sequences transmitted by the users probing the sub-channel. Since the training sequences are assigned randomly, the variance $\sigma_{e|K'}^2(i)$ is random in general and can be different for different users in \mathcal{K} . The receiver uses the channel estimates for both coherent detection, and to “assign” sub-channels. That is, sub-channel 1 is assigned to user $k \in \mathcal{K}$ if its estimated channel gain is largest among users probing the sub-channel, and it exceeds the threshold $t(K')$, i.e., letting $\hat{\mu}_i = |\hat{h}_i|^2$,

$$\hat{\mu}_k \geq \hat{\mu}_i, \quad \text{for all } i \in \mathcal{K}, \quad \text{and} \quad \hat{\mu}_k \geq t(K'). \quad (4)$$

Here we explicitly denote the dependence of the threshold on K' . The assigned user then transmits data on the sub-channel during the rest of the coherence block. If $\hat{\mu}_i < t(K')$ for all $i \in \mathcal{K}$, then the sub-channel is not used. The sub-channel assignments are made after training and before data transmission. Here we assume that these assignments take up a negligible fraction of the coherence block.

The capacity for user k (and hence for any user by symmetry) summed over all N sub-channels is given by³

$$C_k = N(1 - \alpha) \frac{1}{D} I(\mathbf{x}_k; (\mathbf{y}_D, \mathcal{K}, \{\sigma_{e|K'}^2(i)\}, \{\hat{h}_i\})) \quad (5)$$

where the mutual information for a particular sub-channel is

$$I(\mathbf{x}_k; (\mathbf{y}_D, \mathcal{K}, \{\sigma_{e|K'}^2(i)\}, \{\hat{h}_i\})) \geq qDE \left[\log \left(1 + \frac{P_D \hat{\mu}_k}{P_D \sigma_{e|K'}^2(k) + \sigma_n^2} \right) \middle| \begin{array}{l} \text{user } k \text{ is assigned} \\ \text{the subchannel} \end{array} \right] \quad (6)$$

The first line (5) is the mutual information for a fading channel with channel estimation error and the lower bound in (6) is taken from [4], [5]. Our performance objective is ergodic capacity per user and hence the expectation in (6) is over the joint distributions of K' , $\{\hat{\mu}_i\}$, and $\{\sigma_{e|K'}^2(i)\}$ given that user k is assigned the sub-channel. In general, the channel estimates across the users, $\{\hat{\mu}_i\}$, can be correlated due to non-orthogonal training sequences. In what follows, we will take a large system limit in which the channel estimates are independent and the channel estimation error variance converges to a constant, which depends only on K' . That is, in this limit the variance does not depend on the particular realizations of random training sequences. We therefore simplify the analysis by assuming that this is true for a large but finite-size system. Hence, dropping the dependence on the user index i , the error variance for all users probing the same sub-channel is denoted as $\sigma_{e|K'}^2$.

Averaging the lower bound in (6) over the distribution of K' and $\{\hat{\mu}_i\}$ gives the achievable rate per user,

$$C = (1 - \alpha) \frac{\beta}{\rho} \sum_{K'=1}^K \left[\int_{t(K')}^{\infty} \log \left(1 + \frac{P_D t}{P_D \sigma_{e|K'}^2 + \sigma_n^2} \right) f_{max;K'}(t) dt \right] p(K')$$

where $\beta = \frac{N}{L}$ is the normalized available bandwidth and $\rho = \frac{K}{L}$ is the normalized load (number of users). Also, $f_{max;K'}(\cdot)$ is the pdf of $\hat{\mu}_{max;K'} = \max_{i \in \mathcal{K}} \hat{\mu}_i$ given $|\mathcal{K}| = K'$, and $p(K')$ is the probability that $|\mathcal{K}| = K'$. With Rayleigh fading and the assumption (justified in next section) that $\{\hat{h}_i\}$ are independent CSCG random variables with variance,

$$E[|\hat{h}_i|^2] = \sigma_{\hat{h}|K'}^2 = \sigma_h^2 - \sigma_{e|K'}^2, \quad \forall i \in \mathcal{K} \quad (8)$$

we have,

$$f_{max;K'}(t) = \frac{K'}{\sigma_{\hat{h}|K'}^2} \left(1 - e^{-t/\sigma_{\hat{h}|K'}^2} \right)^{(K'-1)} e^{-t/\sigma_{\hat{h}|K'}^2}, \quad t \geq 0. \quad (9)$$

³The notation $\{\sigma_{e|K'}^2(i)\} = \{\sigma_{e|K'}^2(i) : i \in \mathcal{K}\}$ and similar definitions holds for $\{\hat{h}_i\}$ and $\{\hat{\mu}_i\}$.

Assuming that each user probes a randomly chosen subset of N' channels out of the available N channels, we also have that

$$p(K') = \binom{K}{K'} r^{K'} (1-r)^{(K-K')} \quad K' = 0, 1, \dots, K \quad (10)$$

where $r = \frac{N'}{N}$ is the probability that sub-channel 1 is probed by a particular user. Here N' is assumed to be the same for all users and is defined as "probing bandwidth". Hence the average number of users, which probe a particular sub-channel is rK .

Letting

$$u(K') = \Pr\{\hat{\mu}_{max;K'} \geq t(K') | K'\} = \int_{t(K')}^{\infty} f_{max;K'}(t) dt, \quad (11)$$

the probability that user k is assigned to sub-channel 1 is

$$q = \frac{1}{K} \sum_{K'=1}^K u(K') p(K'). \quad (12)$$

Given the system parameters K , L , and N (users, coherence time, and bandwidth), our objective is to determine the training length T , probing bandwidth N' and average training power ϵ_T , which maximize the achievable rate (7). This can be done numerically; however, to gain further insight, we will consider a large system limit in which all the system parameters $\{K, L, N\}$ tend to infinity with fixed ratios $\{\beta, \rho\}$. In that case, we wish to characterize the growth in optimal α , ϵ_T and r (training length, training power and probing bandwidth) with total number of users K .

To achieve the rate (7), the receiver must feed back the index of the probing user with the highest estimated channel gain, which requires $\log_2(K)$ bits per coherence block per sub-channel. We also remark that in sub-channel assignment rule (4) the users are assigned sub-channels based only on their estimated channel gains, which does not account for estimation error. A more general scheme could also take into account the error variance $\sigma_{e|K'}^2(i)$.

III. CHANNEL ESTIMATION ERROR VARIANCE

In this section we compute the channel estimation error variance, $\sigma_{e|K'}^2$, which appears in the capacity expression (7). We consider two linear channel estimators, which give different asymptotic growth rates for the optimal parameters. In both cases, since we compute a Linear Minimum Mean Squared Error (LMMSE) estimate, the relation (8) holds. Furthermore, because the sub-channels are *i.i.d.*, without loss of generality, we again focus on subchannel 1.

A. Matched Filter Estimator

Given the vector of received samples \mathbf{y}_T , corresponding to T training symbols in (2), for user $k \in \mathcal{K}$ the matched filter estimator first computes

$$z_k = \mathbf{s}_k \mathbf{y}_T^\dagger = (\mathbf{s}_k \mathbf{s}_k^\dagger) h_k + n_I \quad (13)$$

where \mathbf{s}_k is the $1 \times T$ row vector containing the training symbols of user k and n_I contains the interference from other users and noise. The channel estimate is then $\hat{h}_k = cz_k$, where c is selected to minimize $E[|h_k - \hat{h}_k|^2]$. Here we assume that the training sequences consist of binary symbols $\pm \sqrt{\frac{P_T}{N'}}$. With K' users probing the sub-channel, we have

$$\sigma_{\hat{h}|K'}^2 = E[\hat{\mu}_k] = \frac{\sigma_h^4}{\left[\sigma_h^2 + \left(\frac{K'-1}{T}\right)\sigma_h^2 + \left(\frac{N'}{TP_T}\right)\sigma_n^2\right]} \quad (14)$$

B. LMMSE Channel Estimator

For the model (2) with known training sequence, the LMMSE estimate of the vector of channel gains across users probing the sub-channel is

$$\hat{\mathbf{h}} = \sigma_h^2 \mathbf{S} [\sigma_h^2 (\mathbf{S}^\dagger \mathbf{S}) + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{y}_T^\dagger \quad (15)$$

and the covariance matrix is

$$\Phi_{\hat{\mathbf{h}}} = E[\hat{\mathbf{h}}\hat{\mathbf{h}}^\dagger] = \sigma_h^4 \mathbf{S} [\sigma_h^2 (\mathbf{S}^\dagger \mathbf{S}) + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{S}^\dagger. \quad (16)$$

Here the training symbols, i.e., the entries of \mathbf{S} , are complex *i.i.d.* random variables with mean zero and variance $\frac{P_T}{N'}$. To guarantee that the estimation error variance converges to a large system limit, we also assume that the symbols have finite fourth moment [2].

For a finite size system, the covariance matrix depends on the particular realization of signatures. However, in the large system limit considered here, under certain conditions on parameters $\{\alpha, r, \epsilon_T\}$ to be described subsequently, the diagonal elements of $\Phi_{\hat{\mathbf{h}}}$ converge to a *deterministic* value given by [2], [3]

$$\sigma_{\hat{h}|K'}^2 = \frac{\sigma_h^4}{\sigma_h^2 + \frac{1}{a\xi}} \quad (17)$$

where

$$\xi = \frac{\left(1 - \frac{K'}{T}\right)}{2\sigma_n^2} - \frac{1}{2a\sigma_h^2} + \left[\frac{\left(1 - \frac{K'}{T}\right)^2}{4\sigma_n^4} + \frac{\left(1 + \frac{K'}{T}\right)}{2\sigma_n^2(a\sigma_h^2)} + \frac{1}{4(a\sigma_h^2)^2} \right]^{1/2} \quad (18)$$

and $a = \frac{\epsilon_T}{r\beta}$. In what follows, we will take this to be the channel estimate variance, even though for finite K and T , $\sigma_{\hat{h}|K'}^2$ depends on the realization of training sequences. This substitution still leads to the correct large system limit provided that the variance of $\sigma_{\hat{h}|K'}^2$ tends to zero sufficiently fast with K and T . More precisely, the results in [3] can be used to show that this variance is bounded by $\kappa \frac{a^2 K}{T^2}$, where κ is a constant. According to the results in the next section, the optimal parameters scale in such a way that this variance and the error in achievable rate incurred by using (17) are expected to go to zero in the large system limit.

For both channel estimators considered, the set of channel estimates are zero-mean CSCG random variables, and are assumed to be independent for the following reasons. For the LMMSE estimator with parameter values of interest, the off-diagonal terms of $\Phi_{\hat{\mathbf{h}}}$ converge to zero in the large system limit, so that the channel estimates become pair-wise independent. However, the set of channel estimates across probing users are still dependent since the eigenvalue distribution of $\Phi_{\hat{\mathbf{h}}}$ is non-degenerate in the large system limit [2]. This dependence increases with the ratio $\frac{K'}{T}$. Since we expect to operate at small values of $\frac{K'}{T}$ in the large system limit (i.e., to avoid interference), the error in calculating the achievable rate due to this independence approximation is expected to be quite small. A similar argument justifies the independence assumption in the case of the matched filter estimator. Moreover, assuming that the estimates are independent in a finite system implies more diversity than is actually available, and should therefore give an optimistic estimate of the achievable rate.

IV. SIMPLIFIED RATE OBJECTIVE

In this section, we characterize the parameters α (training duration), ϵ_T (training power), and r (probing bandwidth), which maximize the asymptotic growth rate of the capacity (7) for both channel estimators presented in the last section. It is difficult to work with the capacity expression (7) directly, so that instead we optimize the following simpler expression, which has the same asymptotic behavior:

$$R = (1 - \alpha) \frac{\beta}{\rho} \log \left(1 + \frac{(P_{av} - \epsilon_T) \sigma_{\hat{h}|rK}^2 \log(rK)}{(1 - \alpha) \frac{\beta}{\rho} \sigma_n^2} \right) \quad (19)$$

where $0 < \alpha, r \leq 1$, $0 \leq \epsilon_T \leq P_{av}$ and $\sigma_{\hat{h}|rK}^2$ is the variance of the sub-channel estimate given that rK users probe the sub-channel. We wish to maximize this expression over α, ϵ_T , and r . The following theorem states that the optimal values also maximize the asymptotic growth rate of the capacity C , given by (7).

Theorem 1: Let $(\alpha^*, \epsilon_T^*, r^*) = \arg \max_{\alpha, \epsilon_T, r} R$, R^* be the corresponding maximum, and C^* be the value of C evaluated at $(\alpha^*, \epsilon_T^*, r^*)$ with thresholds chosen as,

$$t(K') = \sigma_{\hat{h}|r^*K}^2 \log(r^*K) - \sigma_{\hat{h}|r^*K}^2 \log \log(r^*K), \quad \forall K' \quad (20)$$

If C has a unique maximum, denoted by $C_{max} = \max_{\alpha, \epsilon_T, r, \{t(K')\}} C$, then for fixed β and ρ ,

$$\lim_{(K, N, L) \rightarrow \infty} (|R^* - C^*| + |C_{max} - C^*|) = 0 \quad (21)$$

with either the Matched filter or LMMSE channel estimator.

The proof is omitted to save space. Hence the parameters, which maximize R , also maximize C asymptotically, and the optimized R and C_{max} exhibit the same asymptotic behavior. Further, the theorem implies that to maximize the asymptotic capacity, the threshold can be chosen as in (20). That is, the threshold can be set according to the *average* number of probing users, r^*K , as opposed to varying it with the instantaneous number of probing users, K' .

A. Numerical Comparison

Fig. 1 compares the parameter values obtained by maximizing R with the corresponding values obtained by maximizing the actual capacity C . The system load $\rho = 0.1$. Here and in the numerical results that follow, $P_{av} = \sigma_h^2 = 1$, the SNR $P_{av}\sigma_h^2/\sigma_n^2 = 0$ dB, $\beta = 1$, and we only show results for the LMMSE channel estimator. (Similar curves were obtained for the matched filter channel estimator.) Also, to reduce the computational complexity, we assume that the threshold $\{t(K')\} = t_0$, independent of K' , and numerically maximize C in (7) with respect to (α, ϵ_T, r) and t_0 for each K .

Figure 2 shows the corresponding plots of R^* and C_{max} . The figures show that the optimized system parameters and the corresponding rates exhibit the same asymptotic trends. The gaps between the optimized parameters and between R^* and C_{max} close, albeit very slowly, as the system size becomes large.

Fig. 3 compares the threshold t_0 , which maximizes C with the asymptotically optimal threshold given by (20). The figure indicates that the growth rate is the same in each case, although there is a significant gap between the curves.

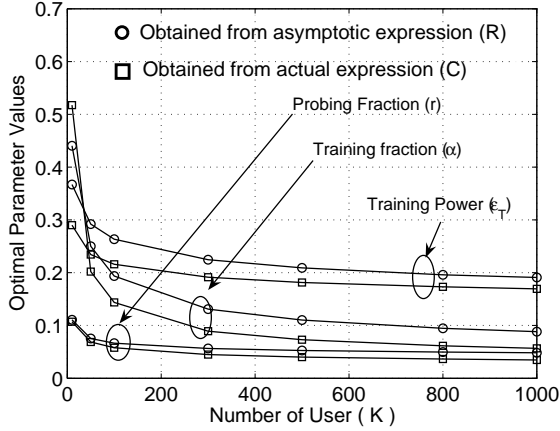


Fig. 1. Comparison of parameters obtained by maximizing R in (19) and C in (7).

V. OPTIMAL PARAMETERS

Setting the derivatives of R with respect to ϵ_T, r, α to zero gives the necessary conditions

$$-\sigma_{\hat{h}|rK}^2 + (P_{av} - \epsilon_T) \frac{\partial \sigma_{\hat{h}|rK}^2}{\partial \epsilon_T} = 0 \quad (22)$$

$$\frac{\sigma_{\hat{h}|rK}^2}{r} + \log(rK) \frac{\partial \sigma_{\hat{h}|rK}^2}{\partial r} = 0 \quad (23)$$

$$-\frac{R}{(1-\alpha)} + (1-\alpha) \frac{\beta}{\rho} \left(\frac{e^{R\rho/(1-\alpha)\beta} - 1}{e^{R\rho/(1-\alpha)\beta}} \right) \cdot \frac{1}{\sigma_{\hat{h}|rK}^2} \left[\frac{\partial \sigma_{\hat{h}|rK}^2}{\partial \alpha} + \frac{\sigma_{\hat{h}|rK}^2}{(1-\alpha)} \right] = 0 \quad (24)$$

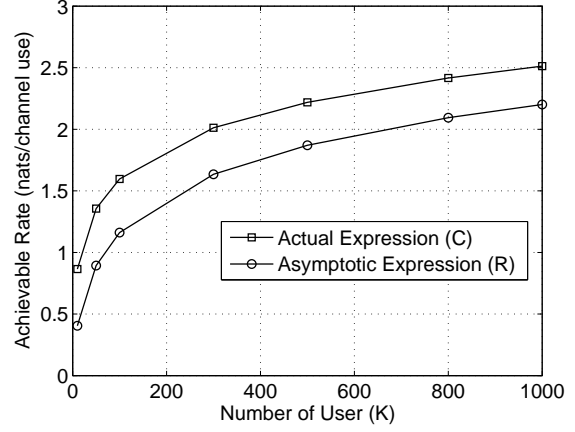


Fig. 2. Optimized rates R^* and C_{max} .

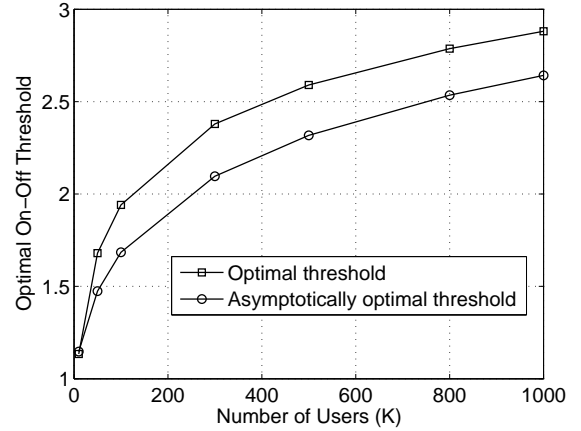


Fig. 3. Comparison of optimal on-off threshold t_0 with the asymptotically optimal threshold (20).

These conditions appear to be difficult to solve directly, so that we consider asymptotic properties as $K \rightarrow \infty$ with fixed β and ρ . We assume certain properties of the asymptotic solution, which simplifies these conditions. We can then determine the asymptotic behavior of r , α , and ϵ_T , and verify that the corresponding solution indeed satisfies the original assumptions.

A. Matched Filter Channel Estimator

We first simplify the necessary conditions (22)-(24) by drawing analogies with the single-user analysis in [5]. Namely, there it is shown that as the system size scales, both the optimal training length α and training power ϵ_T tend to zero. However, the optimal probing bandwidth $r \rightarrow 0$ fast enough so that the channel estimation error tends to zero. With this in mind, we assume that $(\alpha, \epsilon_T, r) \rightarrow 0$, $\epsilon_T \log(rK) \rightarrow 0$ and $\left(\frac{\sigma_{\hat{h}}^2 \rho}{\alpha} + \frac{\beta \sigma_n^2}{\epsilon_T} \right) r \rightarrow 0$ as $K \rightarrow \infty$. With these assumptions,

(22), (23) and (24) imply⁴

$$\epsilon_T^2 \asymp \frac{P_{av}\sigma_n^2}{\sigma_h^2} (\beta r) \quad (25)$$

$$r \log(rK) \asymp \frac{\sigma_h^2}{\left(\frac{\sigma_h^2 \rho}{\alpha} + \frac{\beta \sigma_n^2}{\epsilon_T}\right)} \quad (26)$$

$$\alpha^2 \asymp \frac{\rho r}{\log \left[\frac{P_{av} \sigma_h^2 \log(rK)}{e \sigma_n^2 \frac{\beta}{\rho}} \right]} \quad (27)$$

From these relations we obtain the following asymptotic behavior for the optimal r , ϵ_T , and α ,

$$r^* \asymp \frac{1}{\rho (\log \log(K)) (\log(K))^2} \quad (28)$$

$$\epsilon_T^* \asymp \sqrt{\frac{P_{av}\sigma_n^2\beta}{\sigma_h^2\rho} \frac{1}{(\log \log(K))^{1/2} (\log(K))}} \quad (29)$$

$$\alpha^* \asymp \frac{1}{(\log \log(K)) (\log(K))} \quad (30)$$

It is easy to verify that these relations satisfy the initial assumptions about the solution. The asymptotic expressions (25)-(27) accurately estimate the optimal values provided that the system size K is large enough so that $\left(\frac{\sigma_h^2\rho}{\alpha} + \frac{\beta\sigma_n^2}{\epsilon_T}\right)r \ll \sigma_h^2$ and $\epsilon_T \ll P_{av}$. Numerical results in the next section show that this is satisfied only when K is in the range of several hundred or greater. Nevertheless, the asymptotic trends are present for finite-size systems of interest.

B. LMMSE Channel Estimator

Substituting $K' = rK$ in (17), we have

$$a\xi = \frac{\left(\frac{\epsilon_T}{r\beta} - \frac{\epsilon_T\rho}{\beta\alpha}\right)}{2\sigma_n^2} - \frac{1}{2\sigma_h^2} + \left[\frac{\left(\frac{\epsilon_T}{r\beta} - \frac{\epsilon_T\rho}{\beta\alpha}\right)^2}{4\sigma_n^4} + \frac{\left(\frac{\epsilon_T}{r\beta} + \frac{\epsilon_T\rho}{\beta\alpha}\right)}{2\sigma_n^2\sigma_h^2} + \frac{1}{4\sigma_h^4} \right]^{1/2} \quad (31)$$

Now if we assume that as $K \rightarrow \infty$, the optimal values of the parameters $(\alpha, \epsilon_T, r) \rightarrow 0$, and $\frac{\epsilon_T}{r}, \frac{\alpha}{r}$, and rK all tend to infinity, then (22), (23) and (24) simplify to

$$\epsilon_T^2 \asymp \frac{P_{av}\sigma_n^2}{\sigma_h^2} (\beta r) \quad (32)$$

$$\epsilon_T \log(rK) \asymp P_{av} \quad (33)$$

$$P_{av}\alpha^2 \asymp \frac{\epsilon_T\rho r}{\log \left[\frac{P_{av} \sigma_h^2 \log(rK)}{e \sigma_n^2 \frac{\beta}{\rho}} \right]} \quad (34)$$

which imply

⁴The notation $F_1(K) \asymp F_2(K)$ denotes $\lim_{K \rightarrow \infty} \frac{F_1(K)}{F_2(K)} = 1$.

$$r^* \asymp \frac{P_{av}\sigma_h^2}{\sigma_n^2\beta} \frac{1}{(\log^2(K))} \quad (35)$$

$$\epsilon_T^* \asymp \frac{P_{av}}{\log(K)} \quad (36)$$

$$\alpha^* \asymp \sqrt{\left(\frac{\rho}{\beta}\right) \left(\frac{P_{av}\sigma_h^2}{\sigma_n^2}\right)} \frac{1}{(\log \log(K))^{1/2} (\log(K))^{3/2}} \quad (37)$$

These relations are consistent with our initial assumptions about asymptotic behavior. In this case, for (32)-(34) to give an accurate estimate of the optimal values, K should be large enough so that $\frac{\alpha}{r} \gg \rho$, $\frac{\epsilon_T}{r} \gg \frac{\beta\sigma_n^2}{\sigma_h^2}$ and $\epsilon_T \ll P_{av}$. As for the matched filter, this implies that K must be at least several hundred.

VI. ASYMPTOTIC TRENDS AND NUMERICAL RESULTS

We now highlight some asymptotic properties, and show numerical results for finite-size systems. We also compare the preceding results for the multiple-access channel (MAC) with the analogous results for a single user presented in [5]. The model in [5] assumes the same probing and on-off feedback scheme as that considered here. There the asymptotic results are presented as the coherence time $L \rightarrow \infty$ with an infinite number of independently fading subchannels.

A. Capacity

The asymptotic rate in (19) has the form

$$R = \frac{c_1}{\rho} \log(1 + c_2\rho \log(r\rho L)), \quad (38)$$

where c_1 and c_2 are constants. For large L , R therefore grows as $O(\log \log L)$ and decreases with system load ρ . In contrast, the capacity grows as $O(\log L)$ for single user model considered in [5].

The $O(\log \log L)$ growth for the MAC model is optimal, i.e., matches the order growth with perfect CSI at the transmitter [6]. Here there is a second-order term, which denotes the (additive) loss in capacity due to channel estimation and one-bit feedback. Specifically, substituting the optimal asymptotic parameters into (19) suggests that this term decreases as $O\left(\frac{1}{\log L}\right)$ for the matched filter estimator, and as $O\left(\frac{(\log \log L)^{1/2}}{(\log L)^{3/2}}\right)$ for the LMMSE estimator. As expected, the loss in capacity is greater for the matched filter estimator. Because (19) approximates the capacity for a finite-size system, these error terms are only estimates. A more accurate analysis must take into account the rate at which the gap between (19) and (7) closes.

Rather than allowing the users to choose sub-channels to probe at random, suppose that the users probe *non-overlapping* sets of sub-channels. In that scenario the number of subchannels assigned to each user is $\frac{N}{K} = \frac{\beta}{\rho}$. (A non-integer value of $\frac{\beta}{\rho}$ implies time-sharing of some of the sub-channels.) Even if the transmitter has perfect CSI for all users, as $(K, L, N) \rightarrow \infty$, the capacity per user converges to a constant, which is upper

bounded by $\frac{\beta}{\rho} C^{w,f}$, where $C^{w,f}$ is the ergodic capacity per subchannel achieved by water-pouring over the channel states across time. Hence probing overlapping sets of subchannels exploits multi-user diversity, in spite of the interference present during channel estimation.

Fig. 4 shows plots of the capacity for the MAC channel model C versus coherence time L . As the load ρ increases, the per user capacity decreases due to interference, as shown in the figure. The gap between the single-user curve and the MAC curves should increase as L increases. (The range of coherence times is not large enough to make this apparent.)

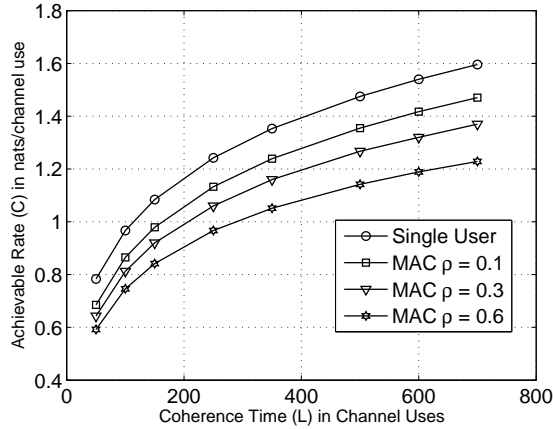


Fig. 4. Achievable rate versus L with different user loads ρ .

B. Probing Bandwidth

The asymptotic relations (28) and (35) imply that the optimal probing bandwidth $N' = r^* N$ scales sublinearly with L . More specifically, the probing bandwidth grows as $O\left(\frac{L}{\log \log(\rho L)(\log(\rho L))^2}\right)$ and $O\left(\frac{L}{(\log(\rho L))^2}\right)$ for the matched filter and LMMSE estimators, respectively. For the single-user model with an LMMSE channel estimator, it can be shown that the optimal probing bandwidth grows as $O\left(\frac{L}{(\log L)^2}\right)$. However, the asymptotic results also imply that for the MAC model, the optimal probing bandwidth decreases with ρ . This is because of the additional interference associated with larger ρ , which degrades the channel estimates. Hence as ρ increases, users should probe fewer sub-channels. Figure 5 shows optimized probing bandwidth versus L . The figure shows that the single-user curve has the same shape as the curves for the MAC channel. Although not shown here, the optimized probing bandwidth with the matched filter estimator is less than that with the LMMSE estimator.

C. Training Power

From (29) and (36), the average training power per user $\epsilon_T^* \rightarrow 0$ at the rate $O\left(\frac{1}{(\log \log(\rho L))^{1/2}(\log(\rho L))}\right)$ and $O\left(\frac{1}{\log(\rho L)}\right)$ for the matched filter and LMMSE estimators, respectively. For the single-user model with the LMMSE

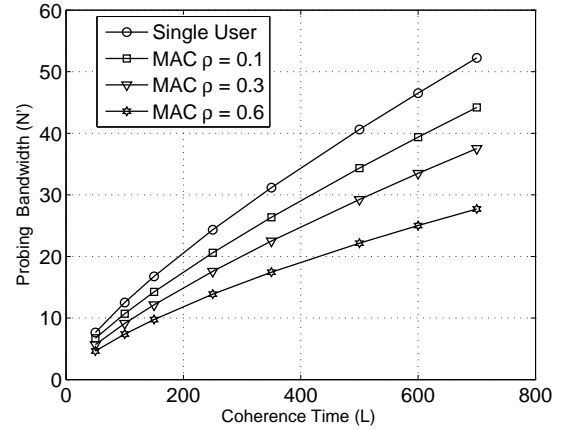


Fig. 5. Optimal probing bandwidth versus L for different system loads ρ .

channel estimator, it can be shown that the optimal training power decreases as $O\left(\frac{1}{(\log L)}\right)$.

Although the training power tends to zero asymptotically, the rate of decrease is sufficiently slow to guarantee increasingly accurate channel estimates. To see this, note that the training energy per fading coefficient, given by $\frac{\epsilon_T^* L}{r^* N}$, tends to infinity as $O((\log \log(\rho L))^{1/2}(\log(\rho L)))$ and $O(\log(\rho L))$ for matched filter and LMMSE estimators, respectively. Similarly, for the single-user model it increases as $O(\log L)$. The optimized training energy for the matched filter estimator is therefore larger than that for the LMMSE estimator, whereas the optimized training power is smaller. Similarly, as ρ increases, the optimized training power decreases, whereas the optimized training energy increases. This reversal of trends is due to the asymptotic behavior of the probing bandwidth, which decreases as ρ increases. Consequently, the optimized training power is spread over fewer sub-channels.

As an example, Figure 6 shows optimal training power versus L . The average training power decreases at a similar asymptotic rate for both the MAC and single-user channel models, and also decreases with load.

D. Training Duration

Figure 7 shows plots of optimal training length versus L with different values of ρ . This figure shows that for the single-user model the training length should be minimized, i.e., $T = 1$, so that $\alpha \rightarrow 0$ as L increases. That is, for the single-user model, the training energy is concentrated in a single symbol.⁵ However, the training length increases almost linearly with L for $\rho > 0$, and the slope increases with ρ . The increase in the length of the training sequence allows for interference suppression during the channel estimation phase. That is, the LMMSE channel estimator can suppress the interferers provided that the length of the training sequence exceeds the number of users probing the particular sub-channel.

⁵It is shown in [5] that as $L \rightarrow \infty$, the performance depends on α only through ϵ_T .

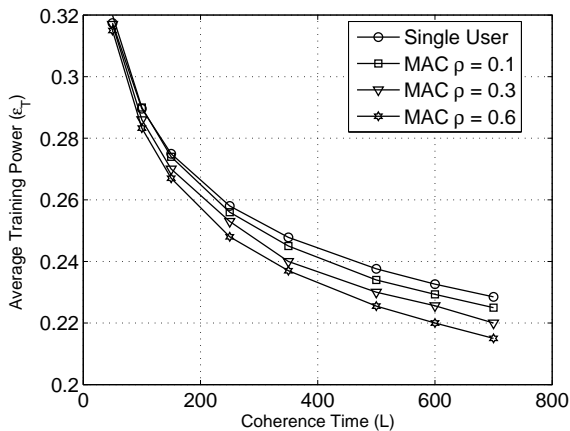


Fig. 6. Optimized training power versus L for different loads ρ .

The average number of users probing a particular sub-channel is r^*K , which increases sublinearly with K . Although the optimal $\alpha \rightarrow 0$, it is also true that $r^*/\alpha^* \rightarrow 0$. Hence the average number of users probing a particular sub-channel should be small relative to the duration of the training sequence. From (30) and (37) the training length grows as $O\left(\frac{L}{(\log \log L)(\log L)}\right)$ for the matched filter estimator, and as $O\left(\frac{L}{(\log \log L)^{1/2}(\log L)^{3/2}}\right)$ for the LMMSE estimator. As expected, the optimal training length for the matched filter estimator is larger than that for the LMMSE estimator.

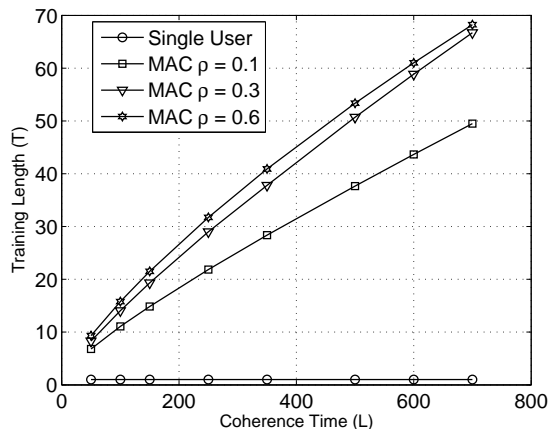


Fig. 7. Optimal training length versus L for different loads ρ .

VII. CONCLUSIONS

We have studied spectrum sharing over a doubly-selective block Rayleigh fading multiple access channel. A random access protocol is assumed, in which the users coordinate transmissions based on limited feedback about channel conditions. Our main results show how the optimal probing bandwidth, training power, and training duration behave asymptotically as the number of users K , bandwidth N , and coherence time L

scale linearly. The scheme is order-optimal in the sense that the capacity exhibits the same growth with coherence time as the capacity with complete channel knowledge at the transmitters (i.e., $O(\log \log L)$). The loss due to training overhead and channel estimation is a second-order term, which tends to zero as $L \rightarrow \infty$. In contrast, if the users avoid interference by probing and transmitting on non-overlapping sub-channels, then the ergodic capacity per user is a constant, i.e., does not increase with L .

The asymptotic analysis shows that the number of users, which probe a particular sub-channel, increases as $O(K/\log^2 K)$. This provides increasingly accurate channel estimates, even though the training power per user tends to zero. The optimal probing bandwidth and training power decrease with the load ρ . An extension of the single-user results in [5] shows that each parameter exhibits the same order-growth with L as for the MAC model. In contrast, the training length must be at least as long as the expected number of users per sub-channel in order to effectively suppress interference from other users probing the same sub-channel. Although the asymptotic results accurately predict the performance only for very large K , the asymptotic trends presented are visible for system sizes of interest.

The model and results presented here might be extended in different ways. For example, we have assumed synchronous training, whereas uncoordinated users are likely to be asynchronous. Also, the sub-channels across frequency and time are likely to be correlated. This may reduce the amount of overhead needed to obtain accurate sub-channel estimates, although it may not change the order of capacity growth from the *i.i.d.* model. Finally, extensions to other network configurations (e.g., peer-to-peer) may also relate to practical spectrum sharing scenarios.

REFERENCES

- [1] D. Tse and P. Viswanath, "Fundamentals of Wireless Communication," Cambridge University Press, 2005.
- [2] J. Evans and D.N.C. Tse, "Large System Performance of Linear Multi-user Receivers in Multipath Fading Channels," *IEEE Trans. on Info. Th.*, vol. 46, pp. 2059-2078, Sep. 2000.
- [3] D.N.C. Tse and O. Zeitouni, "Linear Multiuser Receivers in Random Environments," *IEEE Trans. on Info. Th.*, vol. 46, pp. 171-188, Jan. 2000.
- [4] 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.
- [5] M. Agarwal and M. Honig, "Wideband Fading Channel Capacity with Training and Partial Feedback," *Proc. Allerton Conference*, Sept. 2005.
- [6] Y.K. Sun, "Transmitter and Receiver Techniques for Wireless Fading Channels," *PhD thesis, Northwestern University, December 2004*.
- [7] V. Raghavan, G. Hariharan, and A.M. Sayeed, "Exploiting Time-Frequency Coherence to Achieve Coherent Capacity in Wideband Wireless Channels," *43rd Allerton Conference*, September 2005.
- [8] S. Borade and L. Zheng, "Wideband Fading Channels with Feedback," *Allerton Conference on Communication, Control, and Computing*, Urbana, IL, USA, September 2004.
- [9] S. Guha, K. Munagala, S. Sarkar, "Jointly optimal transmission and probing strategies for multichannel wireless systems", *Conference on Information Sciences and Systems (CISS)*, March 22-24, 2006, Princeton, NJ.
- [10] A. Sabharwal, A. Khoshnevis and E. Knightly, "Opportunistic Spectral Usage: Bounds and a Multi-band CSMA/CA Protocol," *ACM Transactions on Networking*, 2006, accepted for publication.