On the Capacity of Training-Based Transmissions with Input Peak Power Constraints

Mustafa Cenk Gursoy
Department of Electrical Engineering
University of Nebraska-Lincoln, Lincoln, NE 68588
Email: gursoy@engr.unl.edu

Abstract— [1] **In this paper, training-based transmissions over a priori unknown Rayleigh block fading channels are considered. The input signals are assumed to be subject to peak power constraints. Prior to data transmission, channel fading coefficients are estimated in the training phase with the aid of pilot symbols. In this setting, the capacity and capacity-achieving input distribution are studied. The magnitude distribution of the optimal input is shown to be discrete with a finite number of mass points. The capacity, bit energy requirements, and optimal resource allocation strategies are obtained through numerical analysis. The bit energy is shown to grow without bound as SNR decreases to zero due to the presence of peakedness constraints. Capacity and energy-per-bit are also analyzed under the assumptions that the transmitter interleaves the data symbols before transmission over the channel, and per-symbol peak power constraints are imposed. Comparisons of the performances of training-based and noncoherent transmission schemes are provided.**

I. Introduction

In wireless communications, channel conditions vary randomly over time due to mobility and changing environment, and the degree of channel side information (CSI) assumed to be available at the receiver and transmitter is a key assumption in the study of wireless fading channels. Unless there is very high mobility, practical wireless systems generally employ estimation techniques to learn the channel conditions, albeit with errors. Hence, it is of utmost interest to analyze fading channels with imperfect CSI. Médard [5] investigated the effect of imperfect channel knowledge upon the channel capacity, and obtained upper and lower bounds on the input-output mutual information. Lapidoth and Shamai [4] analyzed the effects of channel estimation errors on the performance if Gaussian codebooks are used and nearest neighbor decoding is employed. The capacity of imperfectly-known fading channels is characterized in the low-SNR regime in [6] and in the high-SNR regime in [2].

The aforementioned studies have not considered explicit training and estimation techniques, and resources allocated to them. Recently, Hassibi and Hochwald [8] studied training schemes to learn the multiple-antenna channels. In this work, power and time dedicated to training is optimized by maximizing a lower bound on the capacity. Similar training techniques are also discussed in [3]. Due to its practical significance, the information-theoretic analysis of training schemes has attracted much interest (see e.g., [9] and references therein). Since exact capacity expressions are difficult to find, these studies have optimized the training signal power, duration, and placement using capacity bounds. Since Gaussian noise is the worst uncorrelated additive noise in a Gaussian setting [8], a capacity lower bound is generally obtained by assuming the product of the estimate error and the transmitted signal as another source of Gaussian noise. In this paper, we depart from this approach and find the exact capacity under input peak power constraints. We study the energy efficiency of training-based schemes and optimize the training power allocation using the exact capacity values.

[1]This work was supported in part by the NSF CAREER Grant CCF-0546384.

II. Channel Model

We consider Rayleigh block-fading channels where the input-output relationship within a block of m symbols is given by

$$\mathbf{y} = h\mathbf{x} + \mathbf{n} \tag{1}$$

where $h \sim \mathcal{CN}(0, \gamma^2)$ [2] is a zero-mean circularly symmetric complex Gaussian random variable with variance $E\{|h|^2\} = \gamma^2$, and $\mathbf{n}$ is a zero-mean, m complex-dimensional Gaussian random vector with covariance matrix $E\{\mathbf{n}\mathbf{n}^\dagger\} = N_0\mathbf{I}$. $\mathbf{x}$ and $\mathbf{y}$ are the m complex-dimensional channel input and output vectors, respectively. The channel input is assumed to be subject to the following peak power constraint

$$\|\mathbf{x}\|^2 \overset{\text{a.s.}}{\leq} mP. \tag{2}$$

It is assumed that the fading coefficients stay constant for a block of m symbols and have independent realizations for each block. It is further assumed that neither the transmitter nor the receiver has prior knowledge of the realizations of the fading coefficients.

III. Training-Based Transmission and Reception

We assume that pilot symbols are employed in the system to facilitate channel estimation at the receiver. Hence, the system operates in two phases, namely training and data transmission. In the training phase, pilot symbols known at the receiver are sent from the transmitter and the received signal is

$$\mathbf{y}_t = h\mathbf{x}_t + \mathbf{n}_t \tag{3}$$

where $\mathbf{y}_t$, $\mathbf{x}_t$, and $\mathbf{n}_t$ are l-dimensional vectors signifying the fact that l out of m input symbols are devoted to training. It is assumed that the receiver employs minimum mean-square-error (MMSE) estimation to obtain the estimate

$$\hat{h} = E\{h|\mathbf{y}_t\} = \frac{\gamma^2}{\gamma^2\|\mathbf{x}_t\|^2 + N_0}\mathbf{x}_t^\dagger\mathbf{y}_t. \tag{4}$$

[2]$\mathbf{x} \sim \mathcal{CN}(\mathbf{d}, \mathbf{\Sigma})$ is used to denote that $\mathbf{x}$ is a complex Gaussian random vector with mean $E\{\mathbf{x}\} = \mathbf{d}$ and covariance $E\{(\mathbf{x}-\mathbf{d})(\mathbf{x}-\mathbf{d})^\dagger\} = \mathbf{\Sigma}$

978-1-4244-2075-9/08/$25.00 ©2008 IEEE

With this estimate, the fading coefficient can now be expressed as

$$h = \hat{h} + \tilde{h} \tag{5}$$

where

$$\hat{h} \sim \mathcal{CN}\left(0, \frac{\gamma^4 \|\mathbf{x}_t\|^2}{\gamma^2 \|\mathbf{x}_t\|^2 + N_0}\right) \text{ and } \tilde{h} \sim \mathcal{CN}\left(0, \frac{\gamma^2 N_0}{\gamma^2 \|\mathbf{x}_t\|^2 + N_0}\right). \tag{6}$$

Following the training phase, the transmitter sends the $(m-l)$-dimensional data vector $\mathbf{x}_d$ and the receiver equipped with the knowledge of the channel estimate operates on the received signal

$$\mathbf{y}_d = \hat{h}\mathbf{x}_d + \tilde{h}\mathbf{x}_d + \mathbf{n}_d \tag{7}$$

to recover the transmitted information. We note that since training-based schemes are studied in this paper, memoryless fading channels in which $m = 1$ are not considered, and it is assumed throughout the paper that $m \geq 2$.

IV. Capacity and Energy Efficiency

Since the peak power constraint is imposed on the input vector $\mathbf{x}$, transmission of a single pilot is optimal as the pilot power, if required, can be varied instead of increasing the number of pilot symbols. We assume that the pilot symbol power is

$$|x_t|^2 = \delta m P \tag{8}$$

where $\delta \in (0, 1)$. Therefore, the $(m-1)$-dimensional data vector $\mathbf{x}_d$ is subject to

$$\|\mathbf{x}_d\|^2 \overset{\text{a.s.}}{\leq} (1-\delta) m P. \tag{9}$$

Our goal is to solve the maximization problem

$$C = \sup_{\delta \in (0,1)} \sup_{\substack{\mathbf{x}_d \\ \|\mathbf{x}_d\|^2 \overset{\text{a.s.}}{\leq} (1-\delta) m P}} \frac{1}{m} I(\mathbf{x}_d; \mathbf{y}_d | \hat{h}) \tag{10}$$

and obtain the channel capacity, and identify the capacity-achieving input distribution and the optimal value of the power allocation coefficient δ. The input-output mutual information is

$$I(\mathbf{x}_d; \mathbf{y}_d | \hat{h}) = E_{\hat{h}} E_{\mathbf{x}_d} \int f_{\mathbf{y}|\mathbf{x}_d, \hat{h}}(\mathbf{y}|\mathbf{x}_d, \hat{h}) \log \frac{f_{\mathbf{y}|\mathbf{x}_d, \hat{h}}(\mathbf{y}|\mathbf{x}_d, \hat{h})}{f_{\mathbf{y}|\hat{h}}(\mathbf{y}|\hat{h})} \, d\mathbf{y} \tag{11}$$

where

$$f_{\mathbf{y}|\mathbf{x}_d, \hat{h}}(\mathbf{y}|\mathbf{x}_d, \hat{h}) = \frac{\exp\left(-(\mathbf{y} - \hat{h}\mathbf{x}_d)^\dagger (\tilde{\gamma}^2 \mathbf{x}_d \mathbf{x}_d^\dagger + N_0 \mathbf{I})^{-1} (\mathbf{y} - \hat{h}\mathbf{x}_d)\right)}{\pi^{m-1} N_0^{m-2} (\tilde{\gamma}^2 \|\mathbf{x}_d\|^2 + N_0)}$$

and

$$\tilde{\gamma}^2 = E\{|\tilde{h}|^2\} = \frac{\gamma^2 N_0}{\gamma^2 \delta m P + N_0}. \tag{12}$$

First, we have the following preliminary result on the structure of the capacity-achieving input distribution. We note that due to space limitations, the results are presented without proofs throughout the paper. However, the proofs can be found in [11].

Theorem 1: For the block fading channel (7) where the input is subject to a peak power limitation (9), the capacity-achieving input vector can be written as $\mathbf{x}_d = \|\mathbf{x}_d\| \mathbf{v}$ where $\|\mathbf{x}_d\|$ is a nonnegative real random variable and $\mathbf{v}$ is an independent isotropically distributed unit random vector.

With this characterization, the problem has been reduced to the optimization of the input magnitude distribution, $F_{\|\mathbf{x}_d\|}$. We first obtain an equivalent expression for the mutual information when the the input vector has the structure described in Theorem 1.

Theorem 2: When the input is $\mathbf{x}_d = \|\mathbf{x}_d\| \mathbf{v}$ where $\mathbf{v}$ is an isotropically distributed unit vector that is independent of the magnitude $\|\mathbf{x}_d\|$, the input-output mutual information of the channel (7) can be expressed as

$$I(F_r | \hat{h}) = -E_{\mathsf{K}, r}\left\{\int_0^\infty f_{R|r,\mathsf{K}}(R|r, \mathsf{K}) \log g(R, F_r, \mathsf{K}) \, dR\right\} - E_r\{\log(1 + r^2)\} - (m-1) \tag{13}$$

where

$$f_{R|r,\mathsf{K}}(R|r, \mathsf{K}) = \begin{cases} \frac{R^{m-2}}{(m-3)!} \frac{e^{-R - \frac{\mathsf{K} r^2}{1+r^2}}}{1+r^2} \times \int_0^1 (1-a)^{m-3} e^{\frac{a r^2 R}{1+r^2}} I_0\left(\frac{2\sqrt{\mathsf{K} R}\, r \sqrt{a}}{1+r^2}\right) da & m \geq 3 \\ \frac{e^{-\frac{R + \mathsf{K} r^2}{1+r^2}}}{1+r^2} I_0\left(\frac{2\sqrt{\mathsf{K} R}\, r}{1+r^2}\right) & m = 2 \end{cases}$$

and

$$g(R, F_r, \mathsf{K}) = \frac{(m-2)!}{R^{m-2}} \int_0^\infty f_{R|r,\mathsf{K}}(R|r, \mathsf{K}) \, dF_r. \tag{14}$$

In the above formulations, $R = \frac{\|\mathbf{y}\|^2}{N_0}$, $r = \frac{\tilde{\gamma}\|\mathbf{x}_d\|}{\sqrt{N_0}}$, and $\mathsf{K} = \frac{|\hat{h}|^2}{\tilde{\gamma}^2}$. Furthermore, F_r denotes the distribution function of r. K is an exponential random variable with mean $E\{\mathsf{K}\} = \frac{E\{|\hat{h}^2|\}}{\tilde{\gamma}^2} = \frac{\gamma^2 \delta m P}{N_0}$.

Note that the integral in the mutual information expression in (11) is in general an $2(m-1)$-fold integral. In (13), this has been reduced to a double integral providing a significant simplification especially for numerical analysis. With this result, the channel capacity in nats per symbol can now be reformulated as

$$C = \sup_{\delta \in (0,1)} C_\delta = \sup_{\delta \in (0,1)} \sup_{\substack{F_r \\ r \overset{\text{a.s.}}{\leq} \sqrt{L}}} \frac{1}{m} I(F_r | \hat{h}) \tag{15}$$

where $L = \frac{\gamma^2 (1-\delta) m P}{\gamma^2 \delta m P + N_0}$. Hence, the capacity is obtained through the optimal choices of the power allocation coefficient δ and normalized input magnitude distribution F_r. Since the inner maximization is over a continuous alphabet, the existence of the capacity-achieving distribution F_r is not guaranteed. Next, we prove the existence of a capacity-achieving input distribution and provide a sufficient and necessary condition for an input to be optimal.

Theorem 3: Fix the value of $\delta \in (0, 1)$ and consider the inner maximization in (15). There exists an input distribution F_r that maximizes the mutual information $I(F_r | \hat{h})$. Moreover, an input distribution F_r is capacity-achieving if and only if the following Kuhn-Tucker condition is satisfied:

$$\Phi(r) = E_{\mathsf{K}}\left\{\int_0^\infty f_{R|r,\mathsf{K}}(R|r, \mathsf{K}) \log g(R, F_r, \mathsf{K}) \, dR\right\} + \log(1 + r^2) + m C_\delta + (m-1) \geq 0 \quad \forall r \in [0, \sqrt{L}] \tag{16}$$

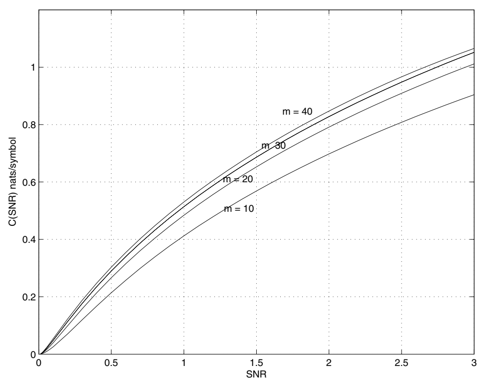

Fig. 1. Capacity (nats/symbol) vs. SNR for block lengths of $m = 10, 20, 30$ and 40 when the input is subject to peak power limitations.

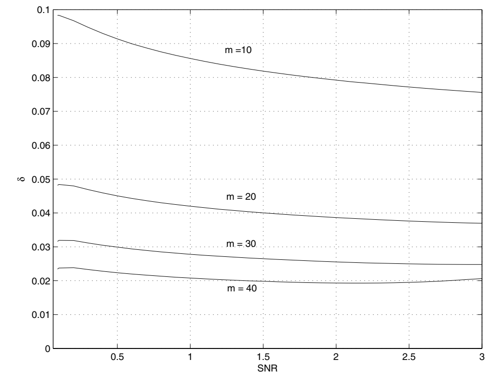

Fig. 2. Optimal fraction of power δ allocated to the pilot symbol vs. SNR for block lengths of $m = 10, 20, 30$ and 40.

with equality at the points of increase of F_r[3]. In the above condition, C_δ denotes the result of the inner maximization in (15) .

Having shown the existence of the capacity-achieving input distribution and a sufficient and necessary condition for an input distribution to be optimal, we turn our attention to the characterization of the optimal input.

Theorem 4: Fix the value of $\delta \in (0, 1)$. The input distribution that maximizes the mutual information $I(F_r|\hat{h})$ is discrete with a finite number of mass points

After the characterization of the discrete nature of the optimal input, the optimization problem in (15) can be solved using vector optimization techniques. Numerical results indicate that the optimal magnitude distribution F_r has a single mass at the peak level $r = \sqrt{L}$ for low-to-medium received peak $\text{SNR} = \frac{\gamma^2 P}{N_0}$ levels. Hence, all the information is carried by the isotropically distributed directional unit vector. Therefore, information transmission is achieved by sending points on the surface of an $(m-1)$-dimensional complex sphere with radius $\frac{\sqrt{LN_0}}{\tilde{\gamma}}$. Note that the mutual information (in nats per m symbols) achieved by having a single-mass at $r = \sqrt{L}$ is

$$I_{cm} = -E_{\mathsf{K}}\left\{\int_0^\infty f_{R|r,\mathsf{K}}(R|r=\sqrt{L},\mathsf{K}) \log g(R, F_r, \mathsf{K})\, dR\right\} - \log(1+L) - (m-1). \tag{17}$$

Figure 1 plots the capacity values as a function of SNR for block lengths of $m = 10, 20, 30$ and 40. These capacity values are achieved with optimal power allocation. The optimal fractions of power allocated to the pilot symbol are plotted in Fig. 2. Note that for the range of SNR values considered in the figure, the optimal value of δ is slightly smaller than $1/m$ and approaches $1/m$ as SNR tends to 0.

In the low-SNR regime, the tradeoff between spectral efficiency and energy per bit obtained from $\frac{E_b}{N_0} = \frac{\text{SNR} \log 2}{C(\text{SNR})}$ is the key performance measure [6]. If we assume, without loss of generality, that one symbol occupies a $1\text{s} \times 1\text{Hz}$ time-frequency slot, then the maximum spectral efficiency is $\mathsf{C}(E_b/N_0) = C(\text{SNR}) \log_2 e$ bits/s/Hz where we have assumed that $C(\text{SNR})$ is in nats/symbol. Fig. 3 plots the bit energy values as a function of the spectral efficiency. It is observed that the minimum bit energy is achieved at a nonzero spectral efficiency and the required bit energy values grow without bound as SNR and hence the spectral efficiency is further decreased. Indeed, we can show the following result.

Theorem 5: Assume that the normalized input magnitude distribution has a single mass and hence the magnitude is fixed at $r = \sqrt{L}$. For any value of $\delta \in (0, 1)$, the normalized bit energy required by this input grows without bound as the signal-to-noise ratio decreases to zero, i.e.,

$$\left.\frac{E_{b,cm}}{N_0}\right|_{I_{cm}=0} = \lim_{\text{SNR}\to 0} \frac{m\text{SNR}}{I_{cm}(\text{SNR})} \log 2 = \frac{m \log 2}{\dot{I}_{cm}(0)} = \infty. \tag{18}$$

In the very low SNR regime, the channel estimate deteriorates and the performance approaches that of noncoherent Rayleigh block fading channels. As shown in [7], bit energy values required in these channels under input peak power constraints grow without bound as $\text{SNR} \to 0$ and the same phenomenon is observed here as well. Fig. 4 provides a comparison of the bit energy values required in the worst case scenario and the scenario where peak power constraints are imposed and optimal signaling and decoding is employed. In the worst-case scenario, the channel estimate is assumed to be perfect and transmission and reception is designed for a known channel. This is obviously a poor assumption in the low-SNR regime and in Fig. 4 we observe bit energy gains of approximately 1.5 dB when optimal techniques are employed in the case of $m = 10$.

In training-based systems, certain fractions of time and power which otherwise will be used for data transmission are allocated to the pilot symbols to facilitate channel estimation. Hence, there is a potential for performance loss in terms of data rates. However, at the same time, the availability of channel estimates at the receiver tends to improve the performance. On the other hand, in noncoherent communications, there is no attempt for channel estimation, and communication is performed over unknown channels. Figures 5 and 6 compare the performances of training-based and noncoherent communication systems. In Fig. 5, the bit energy values are plotted for both schemes when the block length is $m = 20$. It is observed that for this relatively small value of the block length, both schemes achieve almost

[3] The set of points of increase of a distribution function F is $\{r : F(r-\epsilon) < F(r+\epsilon)\ \forall \epsilon > 0\}$.

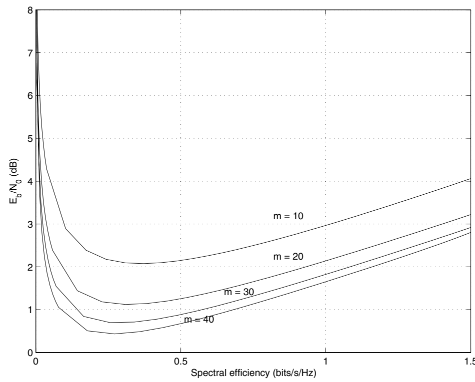

Fig. 3. Bit energy $\frac{E_b}{N_0}$ vs. Spectral efficiency $\mathsf{C}\left(\frac{E_b}{N_0}\right)$ in pilot-assisted systems with block lengths $m = 10, 20, 30$ and 40.

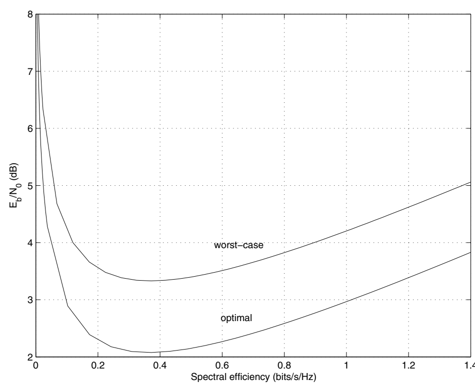

Fig. 4. Bit energy $\frac{E_b}{N_0}$ vs. Spectral efficiency $\mathsf{C}\left(\frac{E_b}{N_0}\right)$ in the worst-case scenario and the scenario of optimal coding-decoding under input peak-power constraints. The block length is $m = 10$.

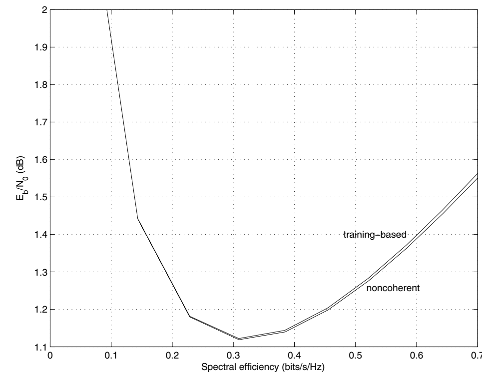

Fig. 5. Bit energy $\frac{E_b}{N_0}$ vs. Spectral efficiency $\mathsf{C}\left(\frac{E_b}{N_0}\right)$ for training-based and noncoherent communication systems when $m = 20$.

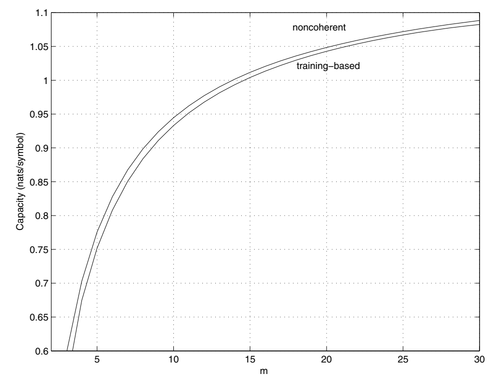

Fig. 6. Capacity (nats/symbol) vs. block length m for training-based and noncoherent communication systems. SNR $= 5$ dB.

the same minimum bit energy value, and therefore, the training-based performance is surprisingly rather close to that of the noncoherent scheme even in the low-SNR regime. Fig. 6 plots the capacity values as a function of the block length at SNR $= 5$ dB. Here, we also observe that the performance of training-based schemes comes very close to that of the noncoherent scheme. Therefore, if having the channel estimate reduces the complexity of the receiver and/or pilot signals are additionally used for timing and frequency-offset synchronization or channel equalization, training-based schemes can be preferred over noncoherent communications with small loss in data rates.

A. Capacity with Ideal Interleaving and Per-symbol Peak Power Constraints

Since most of the well-known codes are designed to correct errors that occur independently from the location of other errors [10], practical communication systems employ interleavers at the transmitters to gain protection against error bursts. Deinterleavers are used at the receiver to reverse the interleaving operation. In this section, we consider such systems and assume that ideal interleaving is used so that each data symbol experiences independent channel conditions. Pilot symbols are inserted periodically after the interleaver. Since interleaving breaks the channel correlation seen by the data symbols, channel memory can no longer be taken advantage of in the transmission. Hence, interleaving in general decreases the capacity. Therefore, the capacity results in this section can also be regarded as lower bounds on the capacity of a non-interleaved system.

We continue considering the block fading channel model. Hence, the channel stays constant for a block of m symbols. However, after deinterleaving, the channel input-output relationship is

$$y_{d,i} = \hat{h}_i x_{d,i} + \tilde{h}_i x_{d,i} + n_i \quad i = 1, 2, 3 \ldots \tag{19}$$

Note that due to interleaving, each data symbol $x_{d,i}$ is affected by independent and identically distributed fading coefficients $h_i = \hat{h}_i + \tilde{h}_i$. In this section, we consider per-symbol peak power constraints, $|x_i|^2 \overset{a.s.}{\leq} P \ \forall i$. Therefore, the pilot symbol power is $|x_t|^2 = P$. Note that the use of more than one pilot may be optimal. The channel capacity in this setting is formulated as follows:

$$C = \sup_{1 \leq l \leq m} \sup_{\substack{x_d \\ |x_d|^2 \overset{\text{a.s.}}{\leq} P}} \frac{m-l}{m} I(x_d; y_d | \hat{h}) \tag{20}$$

where l denotes the number of pilot symbols per m symbols, and

$$\hat{h} \sim \mathcal{CN}\left(0, \frac{\gamma^4 l P}{\gamma^2 l P + N_0}\right) \text{ and } \tilde{h} \sim \mathcal{CN}\left(0, \frac{\gamma^2 N_0}{\gamma^2 l P + N_0}\right).$$

The inner maximization in (20) becomes a special case of the inner maximization in (10) when we reduce the dimensionality of the optimization problem in (10) by choosing $m = 2$. Therefore, the results on the structure of the capacity-achieving input immediately apply to the setting we consider in this section. The optimal input has a uniformly distributed phase. With this characterization, the capacity is

$$C = \sup_{1 \le l \le m} \sup_{\substack{F_r \\ r \overset{\text{a.s.}}{\le} \sqrt{L}}} \frac{m-l}{m} I(F_r|\hat{h}) \tag{21}$$

where

$$I(F_r|\hat{h}) = -E_{\mathsf{K},r}\left\{\int_0^\infty f_{R|r,\mathsf{K}}(R|r,\mathsf{K}) \log g(R, F_r, \mathsf{K})\, dR\right\} - E_r\{\log(1+r^2)\} - 1 \tag{22}$$

and

$$f_{R|r,\mathsf{K}}(R|r,\mathsf{K}) = \frac{e^{-\frac{R+\mathsf{K}r^2}{1+r^2}}}{1+r^2} I_0\left(\frac{2\sqrt{\mathsf{K}R}\,r}{1+r^2}\right), \tag{23}$$

$$g(R, F_r, \mathsf{K}) = \int_0^\infty f_{R|r,\mathsf{K}}(R|r,\mathsf{K})\, dF_r, \tag{24}$$

and, $R = \frac{|y_d|^2}{N_0}$, $r = \frac{\tilde{\gamma}|x_d|}{\sqrt{N_0}}$, $\mathsf{K} = \frac{|\hat{h}|^2}{\tilde{\gamma}^2}$, $\tilde{\gamma}^2 = \frac{\gamma^2 N_0}{\gamma^2 l P + N_0}$, and $L = \frac{\gamma^2 P/N_0}{l\gamma^2 P/N_0 + 1} = \frac{\text{SNR}}{l\text{SNR}+1}$. Note that K is an exponential random variable with mean $E\{\mathsf{K}\} = \frac{E\{|\hat{h}^2|\}}{\tilde{\gamma}^2} = \frac{l\gamma^2 P}{N_0} = l\text{SNR}$. Since the inner maximization in (21) is a special case of that in (15), we immediately have the following result.

Theorem 6: Fix the value of $1 \le l \le m$. The input distribution that maximizes the mutual information $I(F_r|\hat{h})$ in (21) is discrete with a finite number of mass points.

Next, we present numerical results. Fig. 7 plots, for different values of the block length, the capacity curves as a function of SNR for training-based schemes. We observe that the capacity values increase with the block length even though the channel in (19) is memoryless. This performance gain should be attributed to the fact that the channel estimate improves with increasing block length. Fig. 7 also plots the capacity of the interleaved noncoherent communications in which no attempt is made to learn the channel. From the comparison of the capacity curves, we observe that training significantly enhances the data rates when data symbols are interleaved at the transmitter. In Fig. 8, bit energy curves as a function of the spectral efficiency are plotted. Again, we see that training-based schemes perform much better in terms of energy efficiency than the noncoherent scheme. In all cases, the minimum bit energy is achieved at a nonzero spectral efficiency level below which one should not operate. The bit energy requirement increases without bound as spectral efficiency decreases to zero. Finally, when we compare Figs. 3 and 8, we note that while simplifying the system design, interleaving also incurs a penalty in energy efficiency.

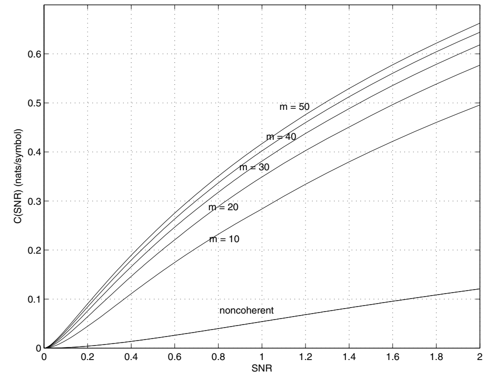

Fig. 7. Capacity (nats/symbol) vs. SNR for interleaved, training-based transmissions when block lengths are $m = 10, 20, 30, 40$ and 50, and for interleaved noncoherent transmission over the unknown Rayleigh fading channel.

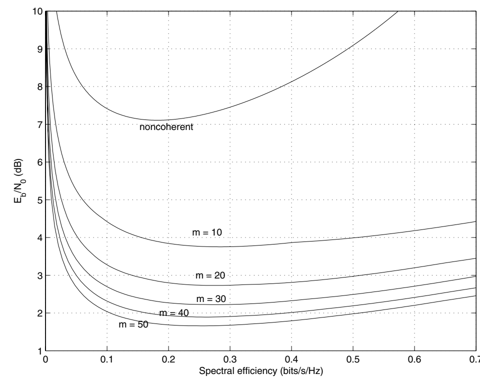

Fig. 8. Bit energy $\frac{E_b}{N_0}$ vs. Spectral efficiency $\mathsf{C}\left(\frac{E_b}{N_0}\right)$ for interleaved, training-based transmissions when block lengths are $m = 10, 20, 30, 40$ and 50, and for interleaved noncoherent transmission over the unknown Rayleigh fading channel.

REFERENCES

[1] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139-157, Jan. 1999.

[2] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2426-2467, Oct. 2003.

[3] L. Zheng and D. N. C. Tse, "Communication on the Grassman manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359-383, Feb. 2002.

[4] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need 'perfect side information' be?," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1118-1134, May 2002.

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

[6] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1319-1343, June 2002.

[7] M. C. Gursoy, H. V. Poor and S. Verdú, "Spectral Efficiency of Peak Power Limited Rician Block-Fading Channels," *Proc. 2004 IEEE Int'l. Symp. Inform. Theory,* Chicago, IL, June 27 - July 2, 2004a.

[8] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. Inform. Theory*, vol. 49, pp. 951-963, Apr. 2003.

[9] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmission," *IEEE Signal Processing Mag.*, pp. 12-25, Nov. 2004.

[10] J. G. Proakis, *Digital Communications 3rd. Ed.* McGraw-Hill: New York, 1995.

[11] M. C. Gursoy, "On the capacity and energy efficiency of training-based transmissions over fading channels" available at http://www.ee.unl.edu/faculty-staff/gursoy.shtml and also at http://arxiv.org.