An Energy Efficiency Perspective on Training for Fading Channels

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

***Abstract*—[1] In this paper, the bit energy requirements of training-based transmission over block Rayleigh fading channels are studied. Pilot signals are employed to obtain the minimum mean-square-error (MMSE) estimate of the channel fading coefficients. Energy efficiency is analyzed in the worst case scenario where the channel estimate is assumed to be perfect and the error in the estimate is considered as another source of additive Gaussian noise. It is shown that bit energy requirement grows without bound as the SNR goes to zero, and the minimum bit energy is achieved at a nonzero SNR value below which one should not operate. The effect of the block length on both the minimum bit energy and the SNR value at which the minimum is achieved is investigated. Flash training schemes are analyzed and shown to improve the energy efficiency in the low-SNR regime. Energy efficiency analysis is also carried out when peak power constraints are imposed on pilot signals.**

I. Introduction

One of the challenges of wireless systems is to establish communication under time-varying channel conditions experienced due to mobility and changing environment. One technique employed in practical systems to cope with this challenge is to periodically send training signals to estimate the channel. Often, the channel estimate is considered as perfect, and transmission and reception is designed for a known channel. Due to its practical significance, training schemes has been studied extensively. Tong *et al.* in [1] present an overview of pilot-assisted wireless transmissions and discuss design issues from both information-theoretic and signal processing perspectives. The information-theoretic approach considers the optimization of training parameters to maximize the achievable data rates. For instance, Hassibi and Hochwald [2] optimized the training data, power, and duration in multiple-antenna wireless links by maximizing a lower bound on the channel capacity. The general theme in information-theoretic approaches (see e.g., [1] and references therein) is that training-based transmission schemes are close to being optimal at high signal-to-noise (SNR) values but highly suboptimal in the low-SNR regime due to poor channel estimates.

Another important concern in wireless communications is the efficient use of limited energy resources. In systems where energy is at a premium, minimizing the energy cost per unit transmitted information will improve the efficiency. Hence, the energy required to reliably send one bit is a metric that can be adopted to measure the performance. Generally, energy-per-bit requirement is minimized, and hence the energy efficiency is maximized, if the system operates in the low-SNR regime [5], [6]. Since training-based schemes perform poorly at low SNRs especially if the channel estimate is presumed to be perfect, this immediately calls into question the energy-efficiency of pilot-assisted systems. With this motivation, we present in this paper an energy-efficiency perspective on pilot-assisted wireless transmission schemes and seek to answer the question of how low should the SNR be so that the energy efficiency is compromised.

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[3] 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 input is subject to an average power constraint

$$E\{\|\mathbf{x}\|^2\} \le mP. \tag{2}$$

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

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

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

[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 $E\{(\mathbf{x}-\mathbf{d})(\mathbf{x}-\mathbf{d})^\dagger\} = \mathbf{\Sigma}$

[3] Note that in the channel model (1), $\mathbf{y}$, $\mathbf{x}$, and $\mathbf{n}$ are column vectors.

1-4244-1429-6/07/$25.00 ©2007 IEEE

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

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

where

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

and

$$\tilde{h} \sim \mathcal{CN}\left(0, \frac{\gamma^2 N_0}{\gamma^2 \|\mathbf{x}_t\|^2 + N_0}\right). \tag{7}$$

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

$$\mathbf{y}_d = \hat{h}\mathbf{x}_d + \tilde{h}\mathbf{x}_d + \mathbf{n}_d \tag{8}$$

to recover the transmitted information.

IV. Energy Efficiency in the Worst Case Scenario

A. Average Power Limited Case

Our overall goal is to identify the bit energy values that can be attained with optimized training parameters such as the power and duration of pilot symbols. The least amount of energy required to send one information bit reliably is given by[4]

$$\frac{E_b}{N_0} = \frac{\text{SNR}}{C(\text{SNR})} \tag{9}$$

where $C(\text{SNR})$ is the channel capacity in bits/symbol. In general, it is difficult to obtain a closed-form expression for the capacity of the channel (8). Therefore, we consider a lower bound on the channel capacity by assuming that

$$\mathbf{z} = \tilde{h}\mathbf{x}_d + \mathbf{n}_d \tag{10}$$

is a Gaussian noise vector that has a covariance of

$$E\{\mathbf{z}\mathbf{z}^\dagger\} = \sigma_{\tilde{h}}^2 E\{\mathbf{x}_d\mathbf{x}_d^\dagger\} + N_0\mathbf{I}, \tag{11}$$

and is uncorrelated with the input signal $\mathbf{x}_d$. With this assumption, the channel model becomes

$$\mathbf{y}_d = \hat{h}\mathbf{x}_d + \mathbf{z}. \tag{12}$$

This model is called the worst-case scenario since the channel estimate is assumed to be perfect, and the noise is modeled as Gaussian, which presents the worst case [2]. The capacity of the channel in (12), which acts as a lower bound on the capacity of the channel in (8), is achieved by a Gaussian input with

$$E\{\mathbf{x}_d\mathbf{x}_d^\dagger\} = \frac{(1-\delta^*)mP}{m-1}\mathbf{I}$$

where δ^* is the optimal fraction of the power allocated to the pilot symbol, i.e., $|x_t|^2 = \delta^* mP$. The optimal value is given by

$$\delta^* = \sqrt{\eta(\eta+1)} - \eta \tag{13}$$

[4]Note that $\frac{E_b}{N_0}$ is the bit energy normalized by the noise power spectral level N_0.

where

$$\eta = \frac{m\,\text{SNR} + (m-1)}{m(m-2)\text{SNR}} \quad \text{and} \quad \text{SNR} = \frac{\gamma^2 P}{N_0}. \tag{14}$$

Note that SNR in (14) is the received signal-to-noise ratio. In the average power limited case, sending a single pilot is optimal because instead of increasing the number of pilot symbols, a single pilot with higher power can be used and a decrease in the duration of the data transmission can be avoided. Hence, the optimal $\mathbf{x}_d$ is an $(m-1)$-dimensional Gaussian vector. Since the above results are indeed special cases of those in [2], the details are omitted. The resulting capacity expression[5] is

$$\begin{aligned} C_L(\text{SNR}) &= \frac{m-1}{m} E_w\left\{\log\left(1 + \frac{\phi(\text{SNR})\text{SNR}^2}{\psi(\text{SNR})\text{SNR} + (m-1)}|w|^2\right)\right\} \\ &= \frac{m-1}{m} E_w\left\{\log\left(1 + f(\text{SNR})|w|^2\right)\right\} \text{ nats/symbol} \end{aligned} \tag{15}$$

where

$$\phi(\text{SNR}) = \delta^*(1-\delta^*)m^2, \tag{16}$$

$$\psi(\text{SNR}) = (1 + (m-2)\delta^*)m, \tag{17}$$

and $w \sim \mathcal{CN}(0,1)$. Note also that the expectation in (15) is with respect to the random variable w. The bit energy values in this setting are given by

$$\frac{E_{b,U}}{N_0} = \frac{\text{SNR}}{C_L(\text{SNR})} \log 2. \tag{18}$$

$\frac{E_{b,U}}{N_0}$ provides the least amount of normalized bit energy values in the worst-case scenario and also serves as an upper bound on the achievable bit energy levels of channel (8). It is shown in [4] that if the channel estimate is assumed to be perfect, and Gaussian codebooks designed for known channels are used, and scaled nearest neighbor decoding is employed at the receiver, then the generalized mutual information has an expression similar to (15) (see [4, Corollary 3.0.1]). Hence $\frac{E_{b,U}}{N_0}$ also gives a good indication of the energy requirements of a system operating in this fashion. The next result provides the asymptotic behavior of the bit energy as SNR decreases to zero.

Lemma 1: The normalized bit energy (18) grows without bound as the signal-to-noise ratio decreases to zero, i.e.,

$$\left.\frac{E_{b,U}}{N_0}\right|_{C_L=0} = \lim_{\text{SNR}\to 0} \frac{\text{SNR}}{C_L(\text{SNR})} \log 2 = \frac{\log 2}{\dot{C}_L(0)} = \infty. \tag{19}$$

Proof: In the low SNR regime, we have

$$C_L(\text{SNR}) = \frac{m-1}{m}\left(f(\text{SNR})E\{|w|^2\} + o(f(\text{SNR}))\right) \tag{20}$$

$$= \frac{m-1}{m}\left(f(\text{SNR}) + o(f(\text{SNR}))\right). \tag{21}$$

As $\text{SNR} \to 0$, $\delta^* \to 1/2$, and hence $\phi(\text{SNR}) \to m^2/4$ and $\psi(\text{SNR}) \to m + m(m-2)/2$. Therefore, it can easily be seen that

$$f(\text{SNR}) = \frac{m^2}{4(m-1)}\text{SNR}^2 + o(\text{SNR}^2) \tag{22}$$

from which we have $\dot{C}_L(0) = 0$. □

[5]Unless specified otherwise, all logarithms are to the base e.

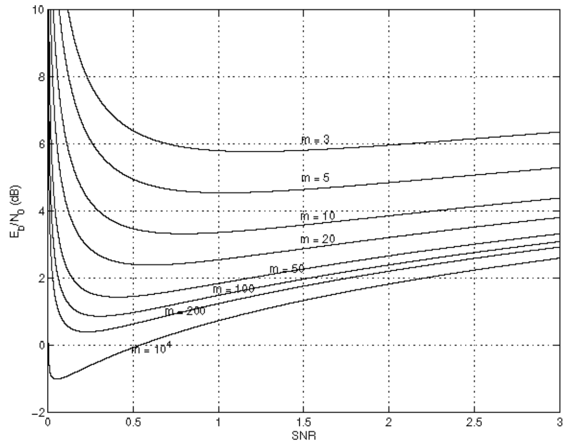

Fig. 1. Energy per bit $E_{b,U}/N_0$ vs. SNR in the worst-case scenario

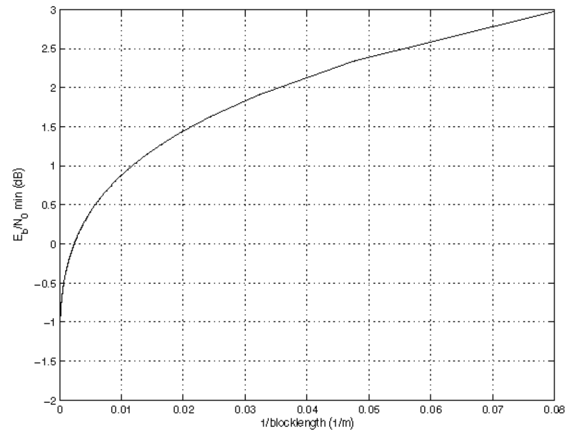

Fig. 2. Minimum energy per bit $\frac{E_{b,U}}{N_0}_{min}$ vs. $\frac{1}{m}$ in the worst-case scenario

The fact that C_L decreases as SNR^2 as SNR goes to zero has already been pointed out in [2]. Lemma 1 shows the impact of this behavior on the energy-per-bit, and indicates that it is extremely energy-inefficient to operate at very low SNR values. We further conclude that in a training-based scheme where the channel estimate is assumed to be perfect, the minimum energy per bit is achieved at a finite and nonzero SNR value. This most energy-efficient operating point can be obtained by numerical analysis. We can easily compute $C_L(\text{SNR})$ in (15), and hence the bit energy values, using the Gauss-Laguerre quadrature integration technique.

Figure 1 plots the normalized bit energy curves as a function of SNR for block lengths of $m = 3, 5, 10, 20, 50, 100, 200, 10^4$. As predicted, for each block length value, the minimum bit energy is achieved at nonzero SNR, and the bit energy requirement increases as $\text{SNR} \to 0$. It is been noted in [2] that training-based schemes, which assume the channel estimate to be perfect, perform poorly at very low SNR values, and the exact transition point below which one should not operate in this fashion is deemed as not clear. Here, we propose the SNR level at which the minimum bit energy is achieved as a transition point since operating below this point results in higher bit energy requirements. Another observation from Fig. 1 is that the minimum bit energy decreases with increasing m and is achieved at a lower SNR value. The following result sheds a light on the asymptotic behavior of the capacity as $m \to \infty$.

Theorem 1: As the block length m increases, C_L approaches to the capacity of the perfectly known channel, i.e.,

$$\lim_{m\to\infty} C_L(\text{SNR}) = E_w\{\log(1 + \text{SNR}|w|^2)\}. \tag{23}$$

Moreover, define $\chi = 1/m$. Then

$$\left.\frac{dC_L(\text{SNR})}{d\chi}\right|_{\chi=0} = -\infty. \tag{24}$$

Proof: We have

$$\lim_{m\to\infty} C_L(\text{SNR}) = \lim_{m\to\infty} E_w\left\{\log\left(1 + f(\text{SNR})|w|^2\right)\right\} \tag{25}$$

$$= E_w\left\{\lim_{m\to\infty} \log\left(1 + f(\text{SNR})|w|^2\right)\right\} \tag{26}$$

$$= E_w\left\{\log\left(1 + \text{SNR}|w|^2\right)\right\}. \tag{27}$$

(26) holds due to integrable upper bound $\left|\log(1+f(\text{SNR})|w|^2)\right| \leq 3\text{SNR}|w|^2$ and the Dominated Convergence Theorem. (24) follows again from the application of the Dominated Convergence Theorem and the fact that the derivative of $f(\text{SNR})$ with respect to $\chi = 1/m$ at $\chi = 0$ is $-\infty$. □

The first part of Theorem 1 is not surprising and is expected because reference [7] has already shown that as the block length grows, the perfect knowledge capacity is achieved even if no channel estimation is performed. This result agrees with our observation in Fig. 1 that -1.59 dB is approached at lower SNR values as m increases. However, the rate of approach is very slow in terms of the block size, as proven in the second part of Theorem 1 and evidenced in Fig. 2. Due to the infinite slope[6] observed in the figure, approaching -1.59dB is very demanding in block length.

B. Flash Training and Transmission

One approach to improve the energy efficiency in the low SNR regime is to increase the peak power of the transmitted signals. This can be achieved by transmitting ν fraction of the time with power P/ν. Note that training also needs to be performed only ν fraction of the time. This type of training and communication, called flash transmission scheme, is analyzed in [8] where it is shown that the minimum bit energy of -1.59dB can be achieved if the block length m increases at a certain rate as SNR decreases. In the setting we consider, flash transmission scheme achieves the following rate:

$$C_{fL}(\text{SNR}, \nu) = \nu(\text{SNR}) C_L\left(\frac{\text{SNR}}{\nu(\text{SNR})}\right) \tag{28}$$

where $0 < \nu(\cdot) \leq 1$ is the duty cycle which in general is a function of the SNR. First, we show that flash transmission using peaky Gaussian signals does not improve the minimum bit energy.

Proposition 2: For any duty cycle function $\nu(\cdot)$,

$$\inf_{\text{SNR}} \frac{\text{SNR}}{C_{fL}(\text{SNR}, \nu)} \geq \inf_{\text{SNR}} \frac{\text{SNR}}{C_L(\text{SNR})}. \tag{29}$$

[6]Note that Theorem 1 implies that the slope of $\frac{\text{SNR}}{C_L(\text{SNR})}$ at $\chi = \frac{1}{m} = 0$ is ∞.

Proof: Note that for any SNR and $\nu(\text{SNR})$,

$$\frac{\text{SNR}}{C_{fL}(\text{SNR},\nu)} = \frac{\frac{\text{SNR}}{\nu(\text{SNR})}}{C_L\left(\frac{\text{SNR}}{\nu(\text{SNR})}\right)} = \frac{\tilde{\text{SNR}}}{C_L(\tilde{\text{SNR}})} \geq \inf_{\text{SNR}} \frac{\text{SNR}}{C_L(\text{SNR})} \quad (30)$$

where $\tilde{\text{SNR}}$ is defined as the new SNR level. Since the inequality in (30) holds for any SNR and $\nu(\cdot)$, it also holds for the infimum of the left-hand side of (30), and hence the result follows. □

We classify the duty cycle function into three categories:

1) $\nu(\cdot)$ that satisfies $\lim_{\text{SNR}\to 0} \frac{\text{SNR}}{\nu(\text{SNR})} = 0$
2) $\nu(\cdot)$ that satisfies $\lim_{\text{SNR}\to 0} \frac{\text{SNR}}{\nu(\text{SNR})} = \infty$
3) $\nu(\cdot)$ that satisfies $\lim_{\text{SNR}\to 0} \frac{\text{SNR}}{\nu(\text{SNR})} = a$ for some constant $a > 0$.

Next, we analyze the performance of each category of duty cycle functions in the low-SNR regime.

Theorem 3: If $\nu(\cdot)$ is chosen from either Category 1 or 2,

$$\left.\frac{E_{b,U}}{N_0}\right|_{C_{fL}=0} = \lim_{\text{SNR}\to 0} \frac{\text{SNR}}{C_{fL}(\text{SNR},\nu)} \log 2 = \infty. \quad (31)$$

If $\nu(\cdot)$ is chosen from Category 3,

$$\left.\frac{E_{b,U}}{N_0}\right|_{C_{fL}=0} = \frac{m}{m-1} \frac{a}{E_w\{\log_2(1+f(a)|w|^2)\}}. \quad (32)$$

Proof: We first note that by Jensen's inequality,

$$\frac{C_{fL}(\text{SNR},\nu)}{\text{SNR}} \leq \frac{m-1}{m} \frac{\nu(\text{SNR})}{\text{SNR}} \log\left(1+f\left(\frac{\text{SNR}}{\nu(\text{SNR})}\right)\right) \quad (33)$$

$$\stackrel{\text{def}}{=} \zeta(\text{SNR},\nu). \quad (34)$$

First, we consider category 1. In this case, as $\text{SNR} \to 0$, $\frac{\text{SNR}}{\nu(\text{SNR})} \to 0$. As shown before, the logarithm in (33) scales as $\frac{\text{SNR}^2}{\nu(\text{SNR}^2)}$ as $\text{SNR} \to 0$, and hence $\zeta(\text{SNR},\nu)$ scales as $\frac{\text{SNR}}{\nu(\text{SNR})}$ leading to

$$\lim_{\text{SNR}\to 0} \frac{C_{fL}(\text{SNR},\nu)}{\text{SNR}} \leq \lim_{\text{SNR}\to 0} \zeta(\text{SNR},\nu) = 0. \quad (35)$$

In category 2, $\frac{\text{SNR}}{\nu(\text{SNR})}$ grows to infinity as $\text{SNR} \to 0$. Since the $\log(\cdot)$ function on the right hand side of (33) increases only logarithmically as $\frac{\text{SNR}}{\nu(\text{SNR})} \to \infty$, we can easily verify that

$$\lim_{\text{SNR}\to 0} \frac{C_{fL}(\text{SNR},\nu)}{\text{SNR}} \leq \lim_{\text{SNR}\to 0} \zeta(\text{SNR},\nu) = 0. \quad (36)$$

In category 3, $\nu(\text{SNR})$ decreases at the same rate as SNR. In this case, we have

$$\lim_{\text{SNR}\to 0} \frac{C_{fL}(\text{SNR},\nu)}{\text{SNR}} = \lim_{n\to\infty} \frac{C_{fL}\left(\frac{1}{n},\nu\right)}{\frac{1}{n}} \quad (37)$$

$$= \frac{\frac{m-1}{m} E_w\{\lim_{n\to\infty} \log\left(1+f(\frac{1}{n\nu})|w|^2\right)\}}{a} \quad (38)$$

$$= \frac{\frac{m-1}{m} E_w\{\log\left(1+f(a)|w|^2\right)\}}{a} \quad (39)$$

(38) is justified by invoking the Dominated Convergence Theorem and noting the integrable upper bound

$$\left|\log\left(1+f\left(\frac{1}{n\nu}\right)|w|^2\right)\right| \leq 3\frac{1}{n\nu}|w|^2 \leq \frac{3}{\nu}|w|^2 \text{ for } n \geq 1.$$

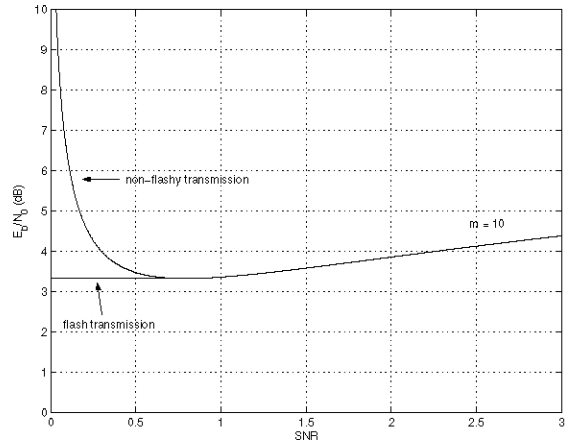

Fig. 3. Energy per bit $E_{b,U}/N_0$ vs. SNR with flash transmission

The above upper bound is given in the proof of Theorem 1. Finally, (39) follows from the continuity of the logarithm. □

Theorem 3 shows that if the rate of the decrease of the duty cycle is faster or slower than SNR as $\text{SNR} \to 0$, the bit energy requirement still increases without bound in the low-SNR regime. This observation is tightly linked to the fact that the capacity curve C_L has a zero slope as both $\text{SNR} \to 0$ and $\text{SNR} \to \infty$. For improved performance in the low-SNR regime, it is required that the duty cycle scale as SNR. A particularly interesting choice is

$$\nu(\text{SNR}) = \frac{1}{a^*}\text{SNR}$$

where a^* is equal to the SNR level at which the minimum bit energy is achieved in a non-flashy transmission scheme. Fig. 3 plots the normalized bit energy $\frac{E_{b,U}}{N_0}$ as a function of SNR for block size $m = 10$. The minimum bit energy is achieved at $\text{SNR} = 0.8$. For $\text{SNR} < 0.8$, flash transmission is employed with $\nu(\text{SNR}) = 1/0.8\,\text{SNR}$. As observed in the figure, the minimum bit energy level can be maintained for lower values of SNR at the cost of increased peak-to-average power ratio.

C. Peak Power Constraint on the Pilot

Heretofore, we have assumed that there are no peak power constraints imposed on either the data or pilot symbols. Recall that the power of the pilot symbol is given by

$$|x_t|^2 = \delta^* mP = \sqrt{\xi(\xi+mP)} - \xi \quad (40)$$

where $\xi = \frac{m\gamma^2 P + (m-1)N_0}{(m-2)\gamma^2}$. We immediately observe from (40) that the pilot power increases at least as $\sqrt{m}$ as m increases. For large block sizes, such an increase in the pilot power may be prohibitive in practical systems. Therefore, it is of interest to impose a peak power constraint on the pilot in the following form:

$$|x_t|^2 \leq \kappa P. \quad (41)$$

Since the average power is uniformly distributed over the data symbols, the average power of a data symbol is proportional to P and is at most $(1-\delta^*)2P$ for any block size. Therefore, κ can be seen as a limitation on the peak-to-average power ratio. Note that we will allow Gaussian signaling for data transmission.

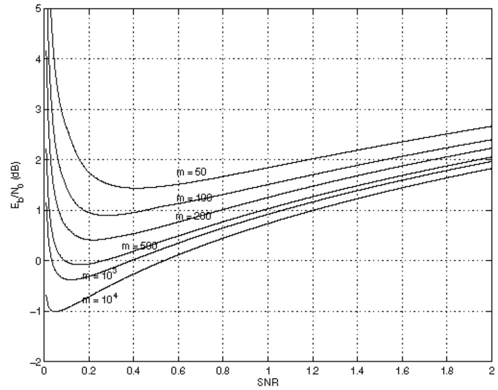

Fig. 4. Energy per bit $E_{b,U}/N_0$ vs. SNR for block sizes of $m = 50, 100, 200, 10^3, 10^4$. The pilot peak power constraint is $|x_t|^2 \leq 10P$.

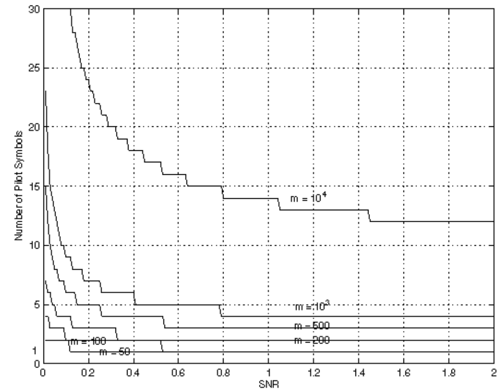

Fig. 5. Number of pilot symbols per block vs. SNR

TABLE I

	$\frac{E_{b,U}}{N_0}_{\min}$ (dB)	# of pilots	SNR	$\frac{E_{b,U}}{N_0}_{\min}$ (dB) (no peak constraints)
$m = 50$	1.441	1	0. 41	1.440
$m = 100$	0.897	2	0.28	0.871
$m = 200$	0.413	3	0.22	0.404
$m = 500$	-0.079	5	0.16	- 0.085
$m = 10^3$	-0.375	9	0.12	-0.378
$m = 10^4$	-1.007	44	0.05	-1.008

Hence, there are no hard peak power limitations on data signals, but flashy signals are precluded. This approach will enable us to work with a closed-form capacity expression. Although Gaussian signals can theoretically assume large values, the probability of such values is decreasing exponentially.

If the optimal power allocated to a single pilot exceeds κP, i.e., $\delta^* m P > \kappa P \Rightarrow \delta^* m > \kappa$, the peak power constraint on the pilot becomes active. In this case, more than just a single pilot may be needed for optimal performance.

In this section, we address the optimization of the number of pilot symbols when each pilot symbol has fixed power $|x_{t,i}|^2 = \kappa P \; \forall i$. If the number of pilot symbols is $l < m$, then $\|\mathbf{x}_t\|^2 = l\kappa P$ and, as we know from Section III,

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

Similarly as before, when the estimate error is assumed to be another source of additive noise and overall additive noise is assumed to be Gaussian, the input-output mutual information achieved by Gaussian signaling is given by

$$I_{L,p} = \frac{m-l}{m} E_w \left\{\log\left(1 + g(\text{SNR}, l)|w|^2\right)\right\} \tag{42}$$

where $w \sim \mathcal{CN}(0,1)$ and

$$g(\text{SNR}, l) = \frac{l\kappa(m - l\kappa)\text{SNR}^2}{(m - l\kappa + (m-l)l\kappa)\text{SNR} + m - l}. \tag{43}$$

The optimal value of the training duration l that maximizes $I_{L,p}$ can be obtained through numerical optimization. Fig. 4 plots the normalized bit energy values $\frac{\text{SNR}\log 2}{I_{L,p}}$ in dB obtained with optimal training duration for different block lengths. The peak power constraint imposed on a pilot symbol is $|x_t|^2 \leq 10P$. Fig. 5 gives the optimal number of pilot symbols per block. From Fig. 4, we observe that the minimum bit energy, which is again achieved at a nonzero value of the SNR, decreases with increasing block length and approaches to the fundamental limit of -1.59dB. We note from Fig. 5 that the number pilot symbols per block increases as the block length increases or as SNR decreases to zero. When there are no peak constraints, $\delta^* m \to m/2$ as SNR $\to 0$. Hence, we need to allocate approximately half of the available total power mP to the single pilot signal in the low-power regime, increasing the peak-to-average power ratio. In the limited peak power case, this requirement is translated to the requirement of more pilot symbols per block at low SNR values.

Table I lists, for different values m, the minimum bit energy values, the required number of pilot symbols at this level, and the SNR at which minimum bit energy is achieved. It is again assumed that $\kappa = 10$. The last column of the table provides the minimum bit energy attained when there are no peak power constraints on the pilot signal. As the block size increases, the minimum bit energy is achieved at a lower SNR value while a longer training duration is required. Furthermore, comparison with the last column indicates that the loss in minimum bit energy incurred by the presence of peak power constraints is negligible.

REFERENCES

[1] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmission," *IEEE Signal Processing Mag.*, pp. 12-25, Nov. 2004.
[2] 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.
[3] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for frequency-selective block-fading channels," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2338-2353, Aug. 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] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1319-1343, June 2002.
[6] M. C. Gursoy, H. V. Poor, and S. Verdú, "The noncoherent Rician fading channel – Part II : Spectral efficiency in the low power regime," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2207-2221, Sept. 2005.
[7] 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.
[8] L. Zheng, D. N. C. Tse, and M. Médard "Channel coherence in the low SNR regime," submitted to *IEEE Trans. Inform. Theory*, Aug. 2005.