OPTIMAL TRAINING OVER THE GAUSS-MARKOV FADING CHANNEL: A CUTOFF RATE ANALYSIS

Saswat Misra[1]*, Ananthram Swami*[1]*, and Lang Tong*[2]
[1] Army Research Lab, Adelphi, MD 20783, USA
[2] Cornell University, Ithaca, NY, USA
{smisra,aswami}@arl.army.mil, ltong@ece.cornell.edu

ABSTRACT

We consider the problem of optimal allocation of resources between training and data for transmission over a Gauss-Markov fading channel. Inaccurate channel state information (CSI) is available at the receiver through periodic training. There is no feedback so that CSI is not available at the transmitter. We study MMSE estimators that predict the current channel state based on: all past pilot observations, only the most recent pilot observation, and the most recent and next in the future pilot observations. We analyze the optimal training energy and periodicity for each of these estimators. We show that optimizing the energy and periodicity of training results in significant energy savings over a sensible, but unoptimized, approach, particularly for rapidly varying channels.

1. INTRODUCTION

Practical communication systems devote part of their power and bandwidth resources to training. Increasing the power (or bandwidth) allocated to training improves the channel estimate, but decreases the power (bandwidth) available for data transmission. We expect that if the channel varies more rapidly, the frequency of training would have to be increased, and if the channel is noisier, the training energy would have to be increased. But channel MSE translates non-linearly into uncoded bit error rate (BER), and computing the impact on coded BER is difficult. To understand these fundamental tradeoffs, we use the cutoff rate as a metric.

The channel cutoff rate R_o is an information-theoretic metric ([16], [12]) that has the following features: it is a lower bound on the channel capacity, and as such, R_o indicates a range of rates R over which reliable communication is possible, $0 < R < R_o$. Secondly, R_o gives a meaningful bound on the error performance P_e of N-length block coding (at any $R < R_o$) via the expression $P_e \leq 2^{-N(R-R_o)}$. Although rates above R_o are attainable [4], the cutoff rate is still regarded as a "practical channel capacity" for simple encoding/decoding strategies. For example, R_o has been proven to play exactly this role for sequential decoding [2].

Previous works have considered the cutoff rate of systems operating over the Rayleigh fading channel under the assumption that either perfect channel state information (CSI) or no CSI is available at the receiver. The cutoff rate with perfect CSI has been examined for i.i.d. fading in [11], and for temporally correlated fading in [5]. The cutoff rate with no CSI can be found in [10].

In the practical case of imperfect receiver CSI, one way to characterize reliable rates, and define optimized estimation parameters, is to first fix a particular channel estimation scheme (or "front end"), and then maximize the information theoretic metric of interest (e.g., cutoff rate or mutual information) over the relevant system design parameters. This simplifying approach is used in both [9] and [13] for a training-based, minimum mean square error (MMSE) channel estimation front-end. In both cases, the authors have considered the mutual information metric.

We consider a periodic training scheme as in [13]. Pilot symbols are sent periodically to provide (inaccurate) estimates of the flat-fading channel coefficient. Knowledge of the channel correlation allows us to predict the fading channel between pilot symbols. We consider MMSE channel predictors that use: (a) the last pilot; (b) all of the past pilots; (c) the last and next pilot (non-causal). In [15], we provided preliminary results for the first estimator. Using the cutoff rate as a metric, we address the following issues: What is the optimal energy allocation between data and training? What is the optimal training frequency? To illustrate the approach, we focus on a first-order Gauss-Markov model that has often been used to characterize fading channels. We quantify our results in terms of an α parameter that measures how rapidly the channel is fading ($\alpha = 0$ corresponds to i.i.d. fading, and $\alpha = 1$ to a static channel). Our key findings are summarized in Section 5.

2. SYSTEM MODEL

We describe the channel model, the training scheme, and the channel estimation algorithm.

2.1. The Gauss-Markov Channel

The channel of interest is a Rayleigh flat fading model. We consider the first order Gauss-Markov fading channel described via

$$\begin{aligned} h_k &= \alpha h_{k-1} + z_k \qquad \text{(state)}, \\ y_k &= \sqrt{E_k} h_k s_k + n_k \quad \text{(observation)}, \end{aligned} \tag{1}$$

where k denotes discrete time, $h_k \sim \mathcal{CN}(0, \sigma_h^2)$ models fading, $0 < \alpha < 1$ describes the channel correlation and is related to the normalized Doppler spread; $z_k \sim \mathcal{CN}(0, \sigma_h^2(1-\alpha^2))$ is driving noise, y_k is the signal at the receiver, E_k is the energy of the k-th symbol, $n_k \sim \mathcal{CN}(0, \sigma_N^2)$ models AWGN[1], and $s_k \in \{-1, +1\}$ is BPSK modulation. We assume that $\sigma_N^2, \sigma_h^2 \neq 0$.

2.2. Channel Estimation

To obtain (imperfect) CSI at the receiver, a single pilot symbol is inserted periodically into the symbol stream with period T. Periodic pilot placement is natural given the wide-sense stationarity

[1] $\mathcal{CN}(\mu, \sigma^2)$ denotes a complex Gaussian random variable with mean μ, and independent real and imaginary parts, each with variance $\sigma^2/2$.

of the channel. Motivation for inserting a single pilot at a time, rather than several, may be found in [6], [1], and [7]. Periodic pilot placement implies that $s_{mT} = +1$, and suggests periodic energy allocation, i.e., $E_k = E_{k \bmod T}$.

We consider three different MMSE channel estimators:

E1. The $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$ estimator uses only the most recent pilot observation to predict the subsequent $T-1$ channel states before the next pilot, i.e., the channel estimate ℓ positions after the most recent pilot is given by $\widehat{h}_{mT+\ell} = \mathcal{E}\left[h_{mT+\ell}|y_{mT}\right]$.

E2. The $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator uses all past pilots to predict the current channel state, i.e., $\widehat{h}_{mT+\ell} = \mathcal{E}\left[h_{mT+\ell}|\left\{y_{pT}\right\}_{p=-\infty}^{m}\right]$. This estimator is readily implemented as a Kalman filter.

E3. The $\mathbf{Q}_{(\mathbf{1},\mathbf{1})}$ estimator is a non-causal smoother which uses the last and "next" pilot observations to predict the current channel state, i.e., $\widehat{h}_{mT+\ell} = \mathcal{E}\left[h_{mT+\ell}|y_{mT}, y_{(m+1)T}\right]$.

We can rewrite the system equation (1) explicitly in terms of the channel estimate $\widehat{h}_k$ and estimation error $\widetilde{h}_k$:

$$y_k = \sqrt{E_{\lceil k/T \rceil}}\,\widehat{h}_k\, s_k + \sqrt{E_{\lceil k/T \rceil}}\,\widetilde{h}_k\, s_k + n_k. \qquad (2)$$

The use of an MMSE estimator in E1-E3 implies that the channel estimate and estimation error are Gaussian and orthogonal, so that $\widehat{h}_k \sim \mathcal{CN}\left(0, \widehat{\sigma}_k^2\right)$ and $\widetilde{h}_k \sim \mathcal{CN}\left(0, \sigma_h^2 - \widehat{\sigma}_k^2\right)$. We assume that perfect interleaving is employed, so that $\widehat{h}_k$ and $\widetilde{h}_k$ are i.i.d. sequences in k and with respect to each other [3]. The receiver employs an ML-decoder that treats the channel estimate $\widehat{h}_k$ as if it were the true channel, thereby treating the estimation error as AWGN. Indeed, for the BPSK s_k, the estimation error is independent of s_k and manifests itself as a simple SNR loss.

It will be useful to define the *estimator quality* ω_ℓ as

$$\omega_\ell \triangleq \widehat{\sigma}_\ell^2/\sigma_h^2\,, \qquad (3)$$

which is a function of the particular estimator used and of the time index $(\ell \bmod T)$ relative to the last pilot; it completely captures the estimator performance. The cases of $\omega_\ell = 0$ and $\omega_\ell = 1$ correspond to no CSI and perfect CSI respectively. Let

$$\kappa_0 \triangleq \sigma_h^2 E_0/\sigma_N^2 \quad \text{and} \quad \kappa_\ell \triangleq \sigma_h^2 E_\ell/\sigma_N^2\,,$$

denote the faded pilot and data energies. Then the estimator quality, $\omega_\ell^{(x,y)}$, for the estimator $\mathbf{Q}_{(\mathbf{x},\mathbf{y})}$ is given by

$$\omega_\ell^{(1,0)} = \alpha^{2\ell}\frac{\kappa_0}{1+\kappa_0}, \qquad (4)$$

$$\omega_\ell^{(\infty,0)} = \alpha^{2\ell}\frac{\kappa_0 - 1 + \sqrt{(1+\kappa_0)^2 + 4\kappa_0\frac{\alpha^{2T}}{1-\alpha^{2T}}}}{\kappa_0 + 1 + \sqrt{(1+\kappa_0)^2 + 4\kappa_0\frac{\alpha^{2T}}{1-\alpha^{2T}}}}\,, \qquad (5)$$

$$\omega_\ell^{(1,1)} = (\Gamma_{(\ell)}^2 + \Gamma_{(T-\ell)}^2)(\kappa_0^2 + \kappa_0) + 2\kappa_0^2\Gamma_{(\ell)}\Gamma_{(T-\ell)}\alpha^T \qquad (6)$$

$$\text{where } \Gamma_{(k)} \triangleq \frac{\alpha^k(\kappa_0+1) - \alpha^{2T-k}\kappa_0}{(\kappa_0+1)^2 - \kappa_0^2\alpha^{2T}}.$$

Space limitations preclude us from including proofs, which may be found in [14]. The quality of the $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$ and $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimators decreases monotonically with ℓ, i.e., distance from the last pilot; the quality of the $\mathbf{Q}_{(\mathbf{1},\mathbf{1})}$ estimator is symmetric, with worst performance midway between the two pilots.

3. CUTOFF RATE

The cutoff rate for equiprobable BPSK modulation with perfect interleaving. is given by (we consider the general input, uninterleaved case in [14]):

$$R_o = -\frac{1}{T}\sum_{\ell=1}^{T-1}\log_2\frac{1}{4}\sum_{\mathcal{S}\in\{-1,1\}}\sum_{\mathcal{V}\in\{-1,1\}} \times \mathcal{E}_{\widehat{h}_\ell}\left[\int_{y_\ell}\sqrt{P(y_\ell|\mathcal{S},\widehat{h}_\ell)P(y_\ell|\mathcal{V},\widehat{h}_\ell)}dy_\ell\right], \qquad (7)$$

where $\mathcal{E}_{\widehat{h}_\ell}$ denotes expectation w.r.t. the pdf of the channel estimate, $\widehat{h}_\ell$, and $P(y_\ell|s_\ell,\widehat{h}_\ell)$ is the pdf of the received signal, conditioned upon the transmitted signal and the channel estimate.

Evaluating (7), we find the cutoff rate to be

$$R_o = -\frac{1}{T}\sum_{\ell=1}^{T-1}\log_2\left\{\frac{1}{2} + \frac{1}{2}\left[\frac{1+\kappa_\ell(1-\omega_\ell)}{1+\kappa_\ell}\right]\right\}. \qquad (8)$$

For simplicity, we consider the case where all the data slots have the same energy, $E_\ell = E_1$, $\ell = 1, ..., T-1$; the general case is considered in [14]. The cutoff rate (8) is valid for each of the estimators in E1-E3 by using the appropriate expression for the estimator quality as given in (4) to (6). More generally, we show in [14] that (8) is valid for *all* MMSE estimators of the form $\widehat{h}_{mT+\ell} = \mathcal{E}\left[h_{mT+\ell}|\left\{y_{pT}\right\}_{p\in\mathcal{P}}\right]$, where $\mathcal{P} \subset \mathcal{Z}$, and where the associated estimator quality ω_ℓ must be determined.

Since R_o is an (increasing) function of ω_ℓ, it is useful to compare the estimator quality expressions in (4) to (6). It can be verified that $\omega_\ell^{(1,0)} \le \omega_\ell^{(\infty,0)}$, that $\omega_\ell^{(1,0)} \le \omega_\ell^{(1,1)}$, and that no similar inequality can be given between $\omega_\ell^{(\infty,0)}$ and $\omega_\ell^{(1,1)}$. In the high SNR regime ($\kappa_0 \to \infty$), $\omega_\ell^{(1,0)} \to \omega_\ell^{(\infty,0)}$. This is because the channel is learnt *perfectly* in each pilot slot, and so additional past pilots do not improve the estimator. For a rapidly fading channel, as $\alpha \to 0$, $\omega_\ell^{(1,0)} \to \omega_\ell^{(\infty,0)}$, since it is the most recent pilot that provides most of the information about the channel state. For a nearly static channel, i.e., as $\alpha \to 1$, $\omega_\ell^{(\infty,0)} > \omega_\ell^{(1,1)}$. This is because the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator provides an infinite number of noisy looks at the static channel, whereas the $\mathbf{Q}_{(\mathbf{1},\mathbf{1})}$ estimator provides only two noisy looks. Further, $\omega_\ell^{(\infty,0)} < \omega_\ell^{(1,1)}$ if the channel is varying rapidly or the SNR is large : When $\alpha \to 0$, it is the closest pilots that contain channel information; the $\mathbf{Q}_{(\mathbf{1},\mathbf{1})}$ provides two "close" pilots. As $\kappa_0 \to \infty$, the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator converges to the $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$ estimator, which is outperformed by the $\mathbf{Q}_{(\mathbf{1},\mathbf{1})}$ estimator. Lastly, we note that all three estimators become equivalent for high-SNR static channels, i.e., as $\alpha \to 1$ *and* $\kappa_0 \to \infty$.

4. OPTIMAL ENERGY ALLOCATION AND TRAINING PERIOD

We study the optimal energy allocation (κ_0^*, κ_1^*) and training period T^* for the estimators E1-E3. For a meaningful analysis, we impose a total energy constraint, $\kappa_0 + (T-1)\kappa_1 = \kappa_{\text{av}}T$. First, we find (κ_0^*, κ_1^*) for a fixed T. Then, we consider the optimal value of T. The cutoff rate is given by (8), where w_ℓ is given by (4), (5), or (6). Space limitations preclude us from including proofs, which may be found in [14].

4.1. Optimal Energy Allocation

Assume that the training period T is fixed, and let $\kappa_{\text{tot}} := \kappa_{\text{av}} T$. For the $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$ estimator, we have previously found κ_1^* as [15]

$$\kappa_1^* = \Gamma - \sqrt{\Gamma^2 - \frac{\kappa_{\text{tot}}}{T-1}\Gamma}, \;\; \Gamma = \frac{\kappa_{\text{tot}}+1}{T-2}, \tag{9}$$

for $T > 2$. For $T = 2$, $\kappa_1^* = \kappa_0^* = \kappa_{\text{tot}}/2$. Note that κ_1^* does not depend on α, however $R_o(\kappa_1^*)$ does. In the low energy regime ($\kappa_{\text{tot}} \to 0$), (9) leads to $\kappa_0^* = \frac{\kappa_{\text{tot}}}{2}$, i.e., half of the available energy should be allocated to the pilot symbol. In the high energy regime, ($\kappa_{\text{tot}} \to \infty$), we find that $\kappa_0^* = \kappa_{\text{tot}} \left[\frac{\sqrt{T-1}-1}{T-2}\right]$. For large T, the energy allocated to the training symbol decays as $T^{-1/2}$.

For the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator, the optimal training energy κ_0^* is given implicitly by (for $T > 2$)

$$\kappa_0^* = \max_{0 \le \kappa_0 \le \kappa_{\text{tot}}} \left[\frac{\kappa_{\text{tot}} - \kappa_0}{\kappa_{\text{tot}} - \kappa_0 + (T-1)}\right] \times \frac{\kappa_0 - 1 + \sqrt{(1+\kappa_0)^2 + 4\kappa_0 \frac{\alpha^{2T}}{1-\alpha^{2T}}}}{\kappa_0 + 1 + \sqrt{(1+\kappa_0)^2 + 4\kappa_0 \frac{\alpha^{2T}}{1-\alpha^{2T}}}}. \tag{10}$$

This implicit solution provides useful insights. In the low energy regime, (10) states that $\kappa_0^* = \frac{\kappa_{\text{tot}}}{2}$. In the high energy regime, $\kappa_0^* = \kappa_{\text{tot}} \left[\frac{\sqrt{T-1}-1}{T-2}\right]$. In these two limiting cases, the $\mathbf{Q}_{(\infty,\mathbf{0})}$ and $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$ estimators have the same optimal energy allocation, which is independent of α. In general, κ_0^* decreases as α (which is a measure of channel predictability) increases: As $\alpha \to 1$, $\kappa_0^* = 0$. This is because, the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator provides us with an infinite number of (noisy) observations of the nearly time-invariant channel. Each observation requires only a minuscule amount of energy, in order to make use of the infinite diversity gain. As $\alpha \to 0$, κ_0^* converges to the κ_0^* of (9) : for a rapidly fading channel, the most recent pilot provides all the information about the channel. For a general κ_{tot} and α, $\kappa_{0,(\infty,0)}^* \le \kappa_{0,(1,0)}^*$; the estimator of higher quality requires less training energy.

Lastly, we give κ_0^* for the $\omega_\ell^{(1,1)}$ estimator. For simplicity, we consider the case where $T > 2$. Because no closed form expression for κ_0 exists in general, we will focus on the low energy and high energy regimes. The optimal training energy κ_0^* in the low SNR regime can again be shown to be given by $\kappa_0^* = \frac{\kappa_{\text{tot}}}{2}$. In the high SNR regime, we consider separately the cases of large and small α. For $\alpha \ll 1$, we find once again that $\kappa_0^* = \kappa_{\text{tot}} \left[\frac{\sqrt{T-1}-1}{T-2}\right]$. However, when $\alpha \approx 1$, it can be shown that

$$\gamma_1 + \sqrt{\gamma_1^2 - \gamma_1} \le \;\; \kappa_0^*/\kappa_{\text{tot}} \;\; \le \gamma_2 + \sqrt{\gamma_2^2 - \gamma_2}, \tag{11}$$

where $\gamma_1 = \frac{T^2 + \mod(T,2)}{-2T^3+3T^2+\mod(T,2)}$ and $\gamma_2 = \frac{T^2-2T+2}{-T^3+2T^2-2T+2}$. Comparing the RHS of (11) and the preceding expression for κ_0^* when α is small, it is seen that, at high SNR, κ_0^* decreases as the channel predictability α increases. This is verified by numerical evaluation for all values of SNR. We can show that as T becomes large, the upper bound becomes tight, and the optimal training energy at high SNR (for any α) is again given by $\kappa_0 = 1/\sqrt{T}$. Lastly, we note that $\kappa_{0,(1,1)}^* \le \kappa_{0,(1,0)}^*$; i.e., the better estimator requires less training energy.

	$T^*_{(1,0)}$	$T^*_{(\infty,0)}$	$T_{B,(\infty,0)}$ $T_{B,(1,0)}$	$T^*_{(1,1)}$	$T_{B,(1,1)}$
$\alpha = 0.80$					
$\kappa_{\text{av}} = 1$	3	3	**3**	4	**4**
$\kappa_{\text{av}} = 10$	3	3	**3**	4	**4**
$\kappa_{\text{av}} = 100$	3	3	**3**	4	**4**
$\alpha = 0.95$					
$\kappa_{\text{av}} = 1$	8	5	**5**	10	**7**
$\kappa_{\text{av}} = 10$	5	5	**5**	7	**7**
$\kappa_{\text{av}} = 100$	5	5	**5**	7	**7**
$\alpha = 0.99$					
$\kappa_{\text{av}} = 1$	20	11	**9**	29	**15**
$\kappa_{\text{av}} = 10$	11	10	**9**	17	**15**
$\kappa_{\text{av}} = 100$	9	9	**9**	15	**15**

Table 1. Comparing the optimal training period $T^*_{(.,.)}$ to the lower bound $T_{B,(.,.)}$ for several different value of α and κ_{av}.

4.2. Optimal Training Period

The preceding analysis gives insights into the optimal energy allocation (κ_0^*, κ_1^*) for a fixed training period T. Below, we consider the optimal training period T^* for each estimator.

A lower bound on the optimal value of the training period T_B can be found by considering the high energy regime ($\kappa_{\text{av}} \to \infty$), which is equivalent to assuming that the channel is known *perfectly* in the relevant training slots. The expression for T_B is given below for each of the three estimators.

For the $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$ and $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimators (and more generally, for *all* causal estimators, see [14]), the lower bound is the same, depends only on α, and is given by:

$$T_{B,(1,0)} = T_{B,(\infty,0)} = \arg\min_T \prod_{\ell=1}^{T-1} \left[1 - \frac{\alpha^{2\ell}}{2}\right]^{1/T}.$$

Although the high-SNR bounds are equal, it can be shown that, in general, $T^*_{(1,0)} \ge T^*_{(\infty,0)}$. The better estimation scheme requires *more frequent* pilots. This was noted heuristically in [13]. Here, it can be proved analytically by substituting both (4) and (5) into (8) and comparing the results. Note however that the better estimator achieves a higher cutoff rate, as would be expected. For the $\mathbf{Q}_{(\mathbf{1},\mathbf{1})}$ estimator, the lower bound is

$$T_{B,(1,1)} = \arg\max_T \prod_{\ell=1}^{T-1} \left[1 + \frac{1}{2}\frac{2\alpha^{2T} - \alpha^{2\ell} - \alpha^{2(T-\ell)}}{1-\alpha^{2T}}\right]^{-1/T}.$$

It can be shown that $T_{B,(1,1)} \ge T_{B,(\infty,0)} = T_{B,(1,0)}$ and that $T^*_{(1,0)} \le T^*_{(1,1)}$. Unlike the previous case, we find that the better estimation scheme requires less frequent training. This apparent conflict is resolved by noting that the training period is not determined by the quality ω_ℓ of the estimation scheme, but rather, by how quickly ω_ℓ "falls off" as ℓ is increased. We remark that the metric of interest is the average cutoff rate, not the frequency of training, and that our metric takes into account the training overhead. Table 1 compares T_B to T^* for each estimator, for several values of κ_{av} and α. In general, the bound is accurate for high SNR (larger κ_{av}) rapidly varying channels (smaller α). From the table, when $\alpha = 0.80$, the lower bound is exact for all $\kappa_{\text{av}} \ge 1$. For $\alpha = 0.95$, the lower bound is exact for $\kappa_{\text{av}} \ge 10$. For $\alpha = 0.99$, the lower bound is tight for $\kappa_{\text{av}} \ge 10$, and exact for $\kappa_{\text{av}} \ge 100$.

4.3. Cutoff Rate with Optimized Training

First, we analyze how using each of the three proposed estimators affects the *unoptimized* cutoff rate; see Figure 1. We use the same

energy κ_{av} in each transmission slot, and pick a fixed, but reasonable, value for the training period ($T = 10$) based on Table 1 when $\alpha = 0.99$. At low SNR, there is a 3 dB gain in the cutoff rate for the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator over the $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$. As expected, this gain diminishes as SNR is increased. Also as expected, the $\mathbf{Q}_{(\mathbf{1},\mathbf{1})}$ estimator outperforms the other two at high SNR. The gain in using the $\mathbf{Q}_{(\mathbf{1},\mathbf{1})}$ estimator over the other two is as much as 3 dB at low to moderate SNR. As SNR $\rightarrow \infty$, the cutoff rate for the $\mathbf{Q}_{(\mathbf{1},\mathbf{1})}$ estimator saturates to 0.8514; a value that exceeds the saturation cutoff rate of 0.7829 for either of the other two estimators.

Next, we analyze how using each of three proposed estimators affects the *optimized* cutoff rate. In Figure 2, we plot the cutoff rate, optimized over the energy allocation (κ_0, κ_1) *and* training period T. Note that optimizing the training period effectively narrows the gain of the more complicated estimators over the $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$ estimator. In particular, the $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$ and $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimators perform within 1 dB of each other. At high SNR, the optimized energy allocation provides no gain; even a "sloppy" energy allocation will allow the cutoff rate to saturate to its maximum value. A poor choice of T will result in a large loss in the cutoff rate relative to the optimized value, even at high SNR.

From Figs. 1-2, we determine the gain in cutoff rate attained by using optimized training parameters in place of the unoptimized (but reasonable) parameters. The gain is typically between $3 \sim 4$ dB at low SNR for the $\mathbf{Q}_{(\mathbf{1},\mathbf{0})}$ estimator, $2 \sim 3$ dB for the $\mathbf{Q}_{(\mathbf{1},\mathbf{1})}$ estimator, and ~ 0.5 dB for the $\mathbf{Q}_{(\infty,\mathbf{0})}$ estimator. The gain in using optimized parameters diminishes as the estimation scheme uses more pilot symbols. In this case, as more pilot observations are exploited, the less the cutoff rate benefits from an optimized energy allocation. For more rapidly varying channels, the gain of the (1,1) estimator is even larger, even with optimized parameters. Lastly, we emphasize that the gain would have been even more dramatic, had a "poor" value of T been chosen.

5. SUMMARY

We have given an expression for the cutoff rate R_o of the interleaved Gauss-Markov fading channel when partial CSI is available at the receiver via periodic training and MMSE channel estimation. We have analyzed the optimal training energy κ_0^* and training period T^* for three different MMSE estimators. Certain facts apply to all estimators: κ_0^* is always less than or equal to one-half of the total energy available, with equality at low SNR. As SNR increases, the required training energy diminishes. At high SNR, and for a long training period T, $\kappa_0^* = \kappa_{\text{av}}\sqrt{T}$, where κ_{av} is the *average* energy per slot. For intermediate values of SNR, the optimal training energy depends on the Doppler spread α in a way that is unique to each estimator. In general, as α increases, the required training energy diminishes. We have found a lower bound on the optimal training period that is exact at high SNR (i.e., when the channel prediction error dominates). We have also shown an unexpected result, that the estimator which uses the infinite past requires more frequent training than the one that just uses the last pilot (although it achieves a larger cutoff rate). The non-causal estimator that uses both the last and the next pilot requires the least frequent training among the three estimators. Finally, we have shown that optimizing the training parameters (energy allocation, and training period) results in significant energy savings (up to 4 dB) versus a sensible, but unoptimized approach. Extensions of interest include the analysis of superimposed training, *no* training, general binary modulation schemes, and matrix channels.

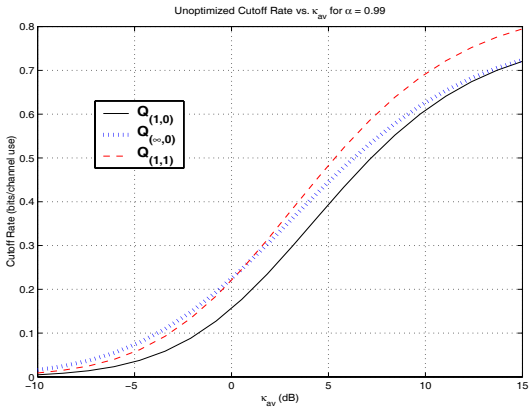

Fig. 1. The cutoff rate for three different estimators for $T = 10$ for equi-energy transmission slots, and for $\alpha = 0.99$.

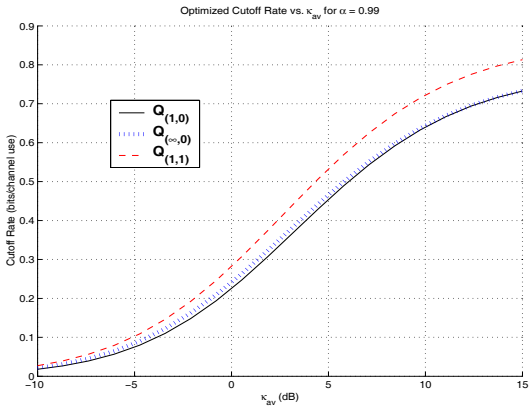

Fig. 2. The cutoff rate for three different PSAM estimators, optimized over the training period T and energy κ_0 for $\alpha = 0.99$.

6. REFERENCES

[1] S. Adireddy, L. Tong, H. Viswanathan, *IEEE Trans. Info. Theory*, Aug 2002.

[2] E. Arikan, *IEEE Trans. Info. Theory*, Jan 1988.

[3] J. Baltersee, G. Fock, H. Meyr, *IEEE Trans. Commun.*, Dec 2001.

[4] C. Berrou, *IEEE Commun. Mag.*, Aug 2003.

[5] K. Leeuwin-Boulle and J.C. Belfiore, *IEEE Trans. Info. Theory*, Mar 1993.

[6] J.K. Cavers, *IEEE Trans. Veh. Tech.*, 40(4), 686-693, Nov 1991.

[7] M. Dong, L. Tong, and B. Sadler, *Proc. ICASSP*, 2002.

[8] R. Gallager, *Information Theory and Reliable Communication*, 1968.

[9] B. Hassibi and B. Hochwald, *IEEE Trans. Info. Theory*, April 2003.

[10] A.O. Hero and T.L. Marzetta, *IEEE Trans. Info. Theory*, Sept 2001.

[11] S. Jamali, and T. Le-Ngoc, *Coded-Modulation Techniques for Fading Channels*, 1994.

[12] J. Massey, In *Proc. 1974 Int. Zurich Seminar., Digital Communication*, March 1974, pp.E2.1-E2.4.

[13] M. Medard, I. Abou-Faycal, and U. Madhow, *Proc. 38th Annual Allerton Conf. on Commun., Control, and Computing*, Oct 2002.

[14] S. Misra, A. Swami, and L. Tong, "Cutoff Rate Analysis of Correlated Fading Channels with Partial Receiver CSI," Journal Version, In preparation.

[15] S. Misra, A. Swami, and L. Tong, *Proc. IEEE SPAWC*, June 2003.

[16] J.M. Wozencraft and I.M. Jacobs, *Principles of Communication Engineering*, 1965.