Optimal Training for Time-Selective Wireless Fading Channels Using Cutoff Rate

Saswat Misra,[1, 2] Ananthram Swami,[1] and Lang Tong[2]

[1] The Army Research Laboratory, Adelphi, MD 20783, USA
[2] Department of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14850, USA

We consider the optimal allocation of resources—power and bandwidth—between training and data transmissions for single-user time-selective Rayleigh flat-fading channels under the cutoff rate criterion. The transmitter exploits statistical channel state information (CSI) in the form of the channel Doppler spectrum to embed pilot symbols into the transmission stream. At the receiver, instantaneous, though imperfect, CSI is acquired through minimum mean-square estimation of the channel based on some set of pilot observations. We compute the ergodic cutoff rate for this scenario. Assuming estimator-based interleaving and M-PSK inputs, we study two special cases in-depth. First, we derive the optimal resource allocation for the Gauss-Markov correlation model. Next, we validate and refine these insights by studying resource allocation for the Jakes model.

1. INTRODUCTION

In wireless communications employing coherent detection, imperfect knowledge of the fading channel state imposes limits on the achievable performance as measured by, for example, the mutual information, the bit-error rate (BER), or the minimum mean-square error (MMSE). Typically, a fraction of system resources—bandwidth and energy—is devoted to channel estimation techniques (known as training) which improve knowledge of the channel state. Such schemes give rise to a tradeoff between the allocation of limited resources to training on one hand and data on the other, and it is natural to seek the optimal allocation of resources between these conflicting requirements. Such optimization is of particular interest for rapidly varying channels, where the energy and bandwidth savings of an optimized scheme can be significant.

In this context, the pilot symbol assisted modulation (PSAM) [1, 2] has emerged as a viable and robust training technique for rapidly varying fading channels. In PSAM, known pilot symbols are multiplexed with data symbols for transmission through the communications channel. At the receiver, knowledge of these pilots is used to form channel estimates, which aid the detection of the data both directly (by modifying the detection rule based on the channel estimate) and indirectly (e.g., by allowing for estimator-directed modulation, power control, and media access). PSAM has been

incorporated into standards for IEEE 802.11, Global System for Mobile Communication (GSM), Wideband Code-Division Multiple-Access (WCDMA), and military protocols, and many theoretical issues are now being addressed. For example, optimized approaches to PSAM have recently been studied from the perspectives of frequency and timing offset estimation [3, 4], BER [1, 5–7], and the channel capacity or its bounds [8–11].

Most relevant to the current study are [12–14], each of which considers PSAM design for the continuously time-varying single-input single-output (SISO) time-selective Rayleigh flat-fading channel, under capacity or its bounds. In each work, the transmitter is assumed to have knowledge of the Doppler spectrum, and the receiver makes (instantaneous) MMSE estimates of the channel based on some subset of the pilot observations. In [13], three estimators (of varying complexity) are proposed and used to predict the channel state for a Gauss-Markov channel correlation model. The optimal binary inputs based on the SNR and estimator statistics are used, and it is determined that for sufficiently correlated channels (i.e., slow enough fading), PSAM provides significant gains in the achievable rates over the no pilot approach. Analysis was carried out through numerical simulation, and the optimization of energy between pilot and data symbols was not attempted. In [12, 14], the authors assume a bandlimited Doppler spectrum and derive closed-form bounds on the channel capacity, using the estimator that exploits all

past and future pilot observations. In both works the capacity and/or its bounds are seen to be parameterized by the variance of this channel estimator. Closed-form results are derived for the optimal allocation of training and bandwidth in some cases.

Here, we study optimal PSAM design for the SISO time-selective Rayleigh flat-fading channel under the cutoff rate. The cutoff rate is a lower bound on the channel capacity and provides an upper bound on the probability of block decoding error (by bounding the random coding exponent). It has been used to establish practical limits on coded performance under complexity constraints [15], and can often be evaluated in closed-form when the capacity cannot (an overview of the cutoff rate for fading channels can be found in [16]). The cutoff rate with perfect receiver channel state information (CSI) has been examined in [17] (independent fading) and in [18, 19] (temporally correlated fading), and for no CSI multiple-input multiple-output (MIMO) systems in [20]. However, we are not aware of any work in which PSAM design is considered from the cutoff rate perspective. Assuming M-PSK inputs, and a general class of MMSE estimators in which some subsets of past and future pilots are exploited at the receiver, we derive a simple expression for the interleaved cutoff rate that will be seen to facilitate analysis.

This paper is organized as follows. In Section 2 we specify the system model and derive the corresponding cutoff rate using M-PSK inputs. In Section 3 we study optimal training for the special case of the Gauss-Markov correlation model. Closed-form expressions for the optimal energy and bandwidth allocation follow in some cases. In Section 4 we validate and refine the design paradigms gained in the last section, by studying optimal training for the well-known (though less tractable) Jakes correlation model. In Section 5 we summarize our guidelines for PSAM design in rapidly fading channels, and propose future work.

Notation. We use the following (standard) notation: (a) $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \Sigma)$ denotes a complex Gaussian random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and with independent real and imaginary parts, each having covariance matrix $\Sigma/2$, (b) $\mathcal{E}_X[\cdot]$ is expectation with respect to the random variable X (the subscript X is omitted where obvious), (c) superscripts "$*$," "t," and "H" denote complex conjugation, transposition, and conjugate transposition, (d) I_N is the $N \times N$ identity matrix, and (e) $|a|$ denotes the absolute value of the scalar a, $|A|$ denotes the determinant of the matrix A, and $|\mathcal{A}|$ denotes the cardinality of the set $\mathcal{A}$ (the context will make use of $|\cdot|$ clear in each case).

2. SYSTEM MODEL

We review the channel model and PSAM-based training scheme, discuss the transmission of a codeword, and evaluate the cutoff rate.

2.1. Channel model

We consider single-user communications over a time-selective (i.e., temporally correlated) Rayleigh flat-fading channel. The sampled baseband received signal y_k (assuming perfect

timing) is given by the scalar observation equation

$$y_k = \sqrt{E_k} h_k s_k + n_k, \tag{1}$$

where k denotes discrete time, $s_k \in \mathcal{S}_M \triangleq \{e^{-j2\pi\nu/M}\}_{\nu=0}^{M-1}$ represents the M-PSK input, E_k is the energy in the kth transmission slot, $h_k \sim \mathcal{CN}(0, \sigma_h^2)$ models fading, and $n_k \sim \mathcal{CN}(0, \sigma_n^2)$ models additive white Gaussian noise (AWGN). We define the normalized channel correlation function

$$R_h(\tau) \triangleq \frac{1}{\sigma_h^2} \mathcal{E}[h_k h_{k+\tau}^*]. \tag{2}$$

2.2. Pilot symbol assisted modulation

In PSAM, the transmitter embeds known pilot symbols into the transmission stream. We consider periodic PSAM in which pilots are embedded with period T, so that $s_k = +1$ at times $k = mT$ ($m = 0, \pm 1, \ldots$). Because the allocation of energy to training versus data symbols entails a tradeoff, we allow a different energy level for each. Define

$$E_k \triangleq \begin{cases} E_P, & k = mT, \\ E_D, & k \neq mT, \end{cases} \tag{3}$$

where E_P is the pilot symbol energy and E_D the data symbol energy.[1] We define the *received* SNR in the pilot and data slots as

$$\kappa_P \triangleq E_P \frac{\sigma_h^2}{\sigma_n^2}, \qquad \kappa_D \triangleq E_D \frac{\sigma_h^2}{\sigma_n^2}. \tag{4}$$

In each time slot $k = mT + \ell$ ($m = 0, \pm 1, \ldots; 0 \leq \ell \leq T - 1$), an MMSE (i.e., conditional-mean) estimate of the channel is made at the receiver using a selection of past, current, and future pilot symbol observations. Specifically, the estimate at the ℓth lag from the most recent pilot is

$$\hat{h}_{mT+\ell} = \mathcal{E}[h_{mT+\ell} \mid \{y_{(m+n)T}\}, n \in \mathcal{N}], \tag{5}$$

where $\mathcal{N} \subseteq \mathbb{Z}$ is the subset of pilot indices used by the estimator.[2] The cardinality $|\mathcal{N}|$ denotes the number of pilots used for estimation. Since $h_{mT+\ell}$ and $\{y_{(m+n)T}\}_{n \in \mathcal{N}}$ are jointly Gaussian, the MMSE estimate of (5) is linear in the pilot observations, and therefore, also Gaussian. We get [22, pages 508–509]

$$\hat{h}_{mT+\ell} = C_{h_\ell \mathbf{y}} C_{\mathbf{yy}}^{-1} \mathbf{y}, \tag{6}$$

where $C_{h_\ell \mathbf{y}}$ is the $1 \times |\mathcal{N}|$ correlation vector between the estimate and observation, $C_{\mathbf{yy}}$ the $|\mathcal{N}| \times |\mathcal{N}|$ observation

---

[1] The current two-dimensional energy allocation problem is easily extendable to a T-dimensional one, in which each of the $T - 1$ data slots may be allocated a unique energy value. We report results from this approach in [21].

[2] Observations in the nonpilot slots could be used to further improve the channel estimate, as is done in semiblind estimators.

covariance matrix, and $\mathbf{y}$ the $|\mathcal{N}| \times 1$ observation vector, whose elements in the ith row and jth column are given by $(1 \leq i, j \leq |\mathcal{N}|)$,

$$(\mathbf{y})_{i,1} = y_{(m+\mathcal{N}_i)T},$$
$$(C_{\mathbf{yy}})_{i,j} = (\mathcal{E}[\mathbf{yy}^H])_{i,j} = E_P \sigma_h^2 R_h(|\mathcal{N}_i - \mathcal{N}_j|T) + \sigma_n^2 \delta(i-j),$$
$$(C_{h_\ell \mathbf{y}})_{1,j} = (\mathcal{E}[h_{mT+\ell}\mathbf{y}^H])_{1,j} = \sqrt{E_P}\sigma_h^2 R_h(|\ell - \mathcal{N}_j T|),$$
$$(7)$$

where $\mathcal{N}_\nu$ denotes the νth *smallest* element in $\mathcal{N}$ ($\nu = 1,\ldots,|\mathcal{N}|$), and where $\delta(\cdot)$ is the Kronecker delta. We will find it useful to write the last two equations in the form

$$C_{\mathbf{yy}} = \sigma_n^2(\kappa_P R_{\mathbf{hh}} + I_{|\mathcal{N}|}),$$
$$C_{h_\ell \mathbf{y}} = \sqrt{E_P}\sigma_h^2 R_{h_\ell \mathbf{y}},$$
$$(8)$$

where definitions of the $|\mathcal{N}| \times |\mathcal{N}|$ matrix $R_{\mathbf{hh}}$ and $1 \times |\mathcal{N}|$ vector $R_{h_\ell \mathbf{y}}$ are evident.

Writing the system model in terms of the channel estimate $\hat{h}_k$ and estimation error $\tilde{h}_k \triangleq h_k - \hat{h}_k$, we have

$$y_k = \sqrt{E_k}\hat{h}_k s_k + \sqrt{E_k}\tilde{h}_k s_k + n_k.$$
$$(9)$$

The estimate of (6) and estimation error $\tilde{h}_{mT+\ell}$ are independent (by application of the orthogonality principle), and it follows that $\hat{h}_{mT+\ell} \sim \mathcal{CN}(0, \hat{\sigma}_{mT+\ell}^2)$ and $\tilde{h}_{mT+\ell} \sim \mathcal{CN}(0, \sigma_h^2 - \hat{\sigma}_{mT+\ell}^2)$, where $\hat{\sigma}_{mT+\ell}^2$ ($0 \leq \hat{\sigma}_{mT+\ell}^2 \leq \sigma_h^2$) is the estimator variance ℓ positions from the most recent pilot. The performance of a particular estimator will be characterized by the normalized estimator variance, termed the *CSI quality*,

and defined as

$$\omega_\ell \triangleq \frac{\hat{\sigma}_{mT+\ell}^2}{\sigma_h^2} = \frac{C_{h_\ell \mathbf{y}} C_{\mathbf{yy}}^{-1} C_{h_\ell \mathbf{y}}^H}{\sigma_h^2}$$
$$= \kappa_P R_{h_\ell \mathbf{y}} (\kappa_P R_{\mathbf{hh}} + I_{|\mathcal{N}|})^{-1} R_{h_\ell \mathbf{y}}^H.$$
$$(10)$$

Note that ω_ℓ is not a function of m (we assume steady state estimation), and that $\omega_\ell = 0$ denotes no CSI, while $\omega_\ell = 1$ denotes perfect CSI. It is assumed throughout that the transmitter has knowledge of ω_ℓ, the statistical quality of channel estimates, but *not* the instantaneous values $\hat{h}_{mT+\ell}$. (For the transmitter to acquire knowledge of ω_ℓ it must know the channel correlation $R_h(\tau)$, the estimation scheme $\mathcal{N}$, and the pilot SNR κ_P.) In the remainder of this paper we will consider two subclasses of estimators.

Causal estimation

Define the $(L, \mathbf{0})$ estimator ($L = 1, 2, \ldots$) to be the estimator which uses the last L causal pilots, $\mathcal{N} = \{-(L-1), -(L-2), \ldots, 0\}$. For example, for the last pilot $(\mathbf{1}, \mathbf{0})$ estimator we have $\mathcal{N} = \{0\}$, and from (10)

$$\omega_\ell^{(1,0)} = |R_h(\ell)|^2 \frac{\kappa_P}{1 + \kappa_P}.$$
$$(11)$$

Noncausal estimation

Define the (L, Z) estimator to be the noncausal estimator which uses the last L causal pilots and next Z noncausal ones, that is, $\mathcal{N} = \{-(L-1), \ldots, 0, \ldots, Z\}$. For example, for the $(\mathbf{1}, \mathbf{1})$ estimator, we have $\mathcal{N} = \{0, 1\}$, and

$$\omega_\ell^{(1,1)} = \frac{(\kappa_P^2 + \kappa_P)\left(|R_h(\ell)|^2 + |R_h(T-\ell)|^2\right) - 2\kappa_P^2 \operatorname{Re}\{R_h(\ell)R_h(T-\ell)R_h^*(T)\}}{(\kappa_P + 1)^2 - \kappa_P^2|R_h(T)|^2},$$
$$(12)$$

where $\operatorname{Re}\{\cdot\}$ denotes the real part.

2.3. Transmission of a codeword

The system transmits codewords of length $N' \triangleq N(T-1)$ where $N > 0$ is a positive integer. Without loss of generality, consider the codeword that starts at time $k = 0$ denoted by

$$S = \operatorname{diag}\{[s_1, \ldots, s_{T-1}, s_{T+1}, \ldots, s_{2T-1}, \ldots,$$
$$s_{(N-1)T+1}, \ldots, s_{NT-1}]\},$$
$$(13)$$

and let

$$\mathbf{h} \triangleq [h_1, \ldots, h_{T-1}, h_{T+1}, \ldots, h_{NT-1}]^t,$$
$$\hat{\mathbf{h}} \triangleq [\hat{h}_1, \ldots, \hat{h}_{T-1}, \hat{h}_{T+1}, \ldots, \hat{h}_{NT-1}]^t,$$
$$\tilde{\mathbf{h}} \triangleq [\tilde{h}_1, \ldots, \tilde{h}_{T-1}, \tilde{h}_{T+1}, \ldots, \tilde{h}_{NT-1}]^t,$$
$$(14)$$

denote the channel, the channel estimate, and the estimation error during the span of the codeword. We define normalized correlation matrices for the channel estimate and estimation error,

$$\hat{\Sigma} \triangleq \frac{1}{\sigma_h^2}\mathcal{E}[\hat{\mathbf{h}}\hat{\mathbf{h}}^H], \qquad \tilde{\Sigma} \triangleq \frac{1}{\sigma_h^2}\mathcal{E}[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H].$$
$$(15)$$

The observation of the codeword after transmission through the channel (9) is

$$\mathbf{y} = \sqrt{E_D}S\hat{\mathbf{h}} + \sqrt{E_D}S\tilde{\mathbf{h}} + \mathbf{n},$$
$$(16)$$

where $\mathbf{n} \triangleq [n_1, \ldots, n_{T-1}, n_{T+1}, \ldots, n_{NT-1}]^t$ is the noise vector. Note that the diagonal elements of $\hat{\Sigma}$ and $\tilde{\Sigma}$ are $\mathbf{1}_N \otimes [\omega_1, \ldots, \omega_{T-1}]$ and $\mathbf{1}_N \otimes [1 - \omega_1, \ldots, 1 - \omega_{T-1}]$, respectively, where $\mathbf{1}_N$ is a row-vector of N ones and where "$\otimes$" denotes the matrix Kronecker product.

The receiver employs the maximum likelihood (ML) detector which regards S as the channel input and the pair $(\mathbf{y}, \widehat{\mathbf{h}})$ as the channel output. Among all possible input symbol sequences for S, denoted by $\mathscr{S}$, the detector chooses the sequence which maximizes the posterior probability of the output, that is,

$$\max_{S \in \mathscr{S}} P(\mathbf{y}, \widehat{\mathbf{h}} \mid S), \tag{17}$$

where $P(\cdot, \cdot \mid \cdot)$ is the probability distribution function (PDF) of the channel outputs, conditioned on the channel input. Noting that $P(\mathbf{y}, \widehat{\mathbf{h}} \mid S) = P(\mathbf{y} \mid S, \widehat{\mathbf{h}})P(\widehat{\mathbf{h}})$ and using standard simplifications under Gaussian statistics, we have, from (17),

$$\max_{S \in \mathscr{S}} \frac{\exp\left\{ (\mathbf{y} - \sqrt{E_D}S\widehat{\mathbf{h}})^H (\sigma_n^2 I_{N'} + \sigma_h^2 S\widetilde{\Sigma}S^H)^{-1} (\mathbf{y} - \sqrt{E_D}S\widehat{\mathbf{h}}) \right\}}{\left| \sigma_n^2 I_{N'} + \sigma_h^2 S\widetilde{\Sigma}S^H \right|}. \tag{18}$$

2.4. Cutoff rate

The cutoff rate, measured in bits per channel use, is [23, 24] (see [18] for time-selective fading channels with perfect receiver CSI)

$$R_o = -\lim_{N \to \infty} \min_{Q(\cdot)} \frac{1}{NT}$$
$$\times \log_2 \int_{\mathbf{y}} \int_{\widehat{\mathbf{h}}} \left[\sum_{S \in \mathscr{S}} Q(S)\sqrt{P(\mathbf{y}, \widehat{\mathbf{h}} \mid S)} \right]^2 d\widehat{\mathbf{h}} \, d\mathbf{y}, \tag{19}$$

where $Q(\cdot)$ is the probability of transmitting a particular codeword. (The normalization factor is $1/NT$ (rather than $1/N'$) to account for the information-loss in pilots slots.) The cutoff rate is evaluated in the appendix and found to be

$$R_o = -\lim_{N \to \infty} \min_{Q(\cdot)} \frac{1}{NT} \log_2 \sum_{V, W \in \mathscr{S}} \frac{Q(V)Q(W)\left| I_{N'} + \kappa_D V\widetilde{\Sigma}V^H \right|^{1/2} \left| I_{N'} + \kappa_D W\widetilde{\Sigma}W^H \right|^{1/2}}{\left| I_{N'} + (1/2)\kappa_D (V\widetilde{\Sigma}V^H + W\widetilde{\Sigma}W^H) + (1/4)\kappa_D (V - W)\widehat{\Sigma}(V - W)^H \right|}. \tag{20}$$

Equation (20) is seen to match [18, equation (14)] for the special case of perfect channel estimation (i.e., $\widehat{\Sigma} = I$ and $\widetilde{\Sigma} = 0$). Equation (20) can be used to determine optimal PSAM parameters and the resulting cutoff rate, however, the ensuing analysis would be largely based on numerical techniques. In the remainder of this paper, we focus on more tractable approaches to an analysis of optimal PSAM.

2.5. Interleaving

An interleaving-deinterleaving pair [25, pages 468–469] is an integral component of many wireless communications systems. A common assumption is that of infinite depth (i.e., perfect) interleaving, in which the correlation between channel fades at any two symbols within a codeword is completely removed. For example, this assumption has been used to study the cutoff rate of the time-selective fading channel with perfect CSI in [18]. Although interleaving discards information on the channel correlation, such a step is necessary in practice since most channel codes in use have been designed for independently fading channels.(The effect of interleaving on the cutoff rate was studied in [19] for a class of block-interference channels with memory. It was shown that the cutoff rate is generally a decreasing function of the chan-

nel memory length, without or without channel state information (this represents a different behavior than known for channel capacity). An analysis of the effect of interleaving is complicated in our setting by the fact that *both* the estimated channel and effective noise term (consisting of the estimation error plus AWGN) are rendered memoryless sequences by the interleaver. Thus, there exist scenarios where interleaving may either increase or decrease R_o.)

Since channel realizations occurring exactly ℓ ($1 \le \ell \le T - 1$) slots from the last pilot have the same estimator statistic ω_ℓ, we assume that these slots are interleaved only among each other (preserving the marginal statistics of the channel estimate and error). Further, it is assumed that the interleaver uses a different interleaving scheme in each sub-channel, so that the correlation between any two codeword symbols is zero. Perfect interleaving renders $\widehat{\Sigma}$ and $\widetilde{\Sigma}$ diagonal, so that

$$\widehat{\Sigma} = I_N \otimes \text{diag}\left\{ [\omega_1, \ldots, \omega_{T-1}] \right\},$$
$$\widetilde{\Sigma} = I_N \otimes \text{diag}\left\{ [1 - \omega_1, \ldots, 1 - \omega_{T-1}] \right\}. \tag{21}$$

Each of the matrices in (20) is now diagonal. The cutoff rate simplifies to

$$R_o = -\frac{1}{T} \sum_{\ell=1}^{T-1} \min_{Q_\ell(\cdot)} \log_2 \sum_{s_\ell, v_\ell \in \mathscr{S}_M} Q_\ell(s_\ell)Q_\ell(v_\ell) \frac{\sqrt{1 + \kappa_D(1 - \omega_\ell)|s_\ell|^2}\sqrt{1 + \kappa_D(1 - \omega_\ell)|v_\ell|^2}}{1 + (\kappa_D/2)(1 - \omega_\ell)(|s_\ell|^2 + |v_\ell|^2) + (\kappa_D/4)\omega_\ell|s_\ell - v_\ell|^2}, \tag{22}$$

where $Q_\ell(\cdot)$ is the probability distribution ℓ slots from the last pilot ($1 \le \ell \le T - 1$). The communications channel is symmetric in its input (M-PSK), and so the cutoff rate is maximized by the equiprobable distribution $Q_\ell(\cdot) = 1/M$. Evaluating the double sum and invoking the constant modulus property of M-PSK yields

$$R_o = -\frac{1}{T} \sum_{\ell=1}^{T-1} \log_2 \left[\frac{1}{M} \sum_{m=0}^{M-1} \frac{1 + \kappa_D(1 - \omega_\ell)}{1 + \kappa_D(1 - \omega_\ell \cos^2(\pi m/M))} \right]. \tag{23}$$

Equation (23) can be interpreted as follows: the ℓth term in the above sum represents the cutoff rate of the ℓth data subchannel (conceptually consisting of all transmissions occurring ℓ slots after a pilot). Thus, (23) represents the cutoff rate of $T - 1$ parallel subchannels, normalized by the factor $1/T$ to account for pilot transmissions. Because the temporal-correlation of the channel is exploited for channel estimation *before* deinterleaving, the cutoff rate depends on the CSI quality $\{\omega_\ell\}_{\ell=0}^{T-1}$. If estimation is perfect ($\omega_\ell = 1$, for all ℓ), (23) matches [18, equation (16)], as it must. Equation (23) represents the M-PSK cutoff rate under perfect interleaving for an arbitrary channel correlation $R_h(\tau)$, estimation scheme $\mathcal{N}$ and power and bandwidth allocation (κ_P, κ_D, T). It is the basis for the subsequent analysis.

3. OPTIMAL TRAINING FOR THE GAUSS-MARKOV MODEL

In this section we determine optimal PSAM parameters under energy and bandwidth constraints for the Gauss-Markov (GM) channel model, whose correlation is described by a first-order autoregressive (AR) process. It is known that second- and third-order AR models provide excellent fits to the Jakes model [26], but they are not as tractable. The GM model has previously been used to characterize the effect of imperfect channel knowledge on the performance of decision-feedback equalization [27], mutual information [28], and minimum mean-square estimation error [6] of time-selective fading channels. The correlation is given by

$$R_h(\tau) = \alpha^{|\tau|} \quad (0 < \alpha < 1), \tag{24}$$

where the α parameter is related to the normalized Doppler spread of the channel and is typically within the range $0.9 \le \alpha < 0.99$ [13, 28]. It will be seen that the GM model provides simple, closed-form, and intuitive expressions for the CSI quality of many estimators of interest (including those of infinite length) and leads to simple design rules for the optimal allocation of resources between training and data, motivating its study in this section.

3.1. Energy allocation

In one period of transmission, the total energy consumed is $\kappa_P + (T-1)\kappa_D$ (without ambiguity, we use received energies), and an energy constraint requires that

$$\kappa_P + \kappa_D(T - 1) \le \kappa_{av} T, \tag{25}$$

where $\kappa_{av} > 0$ is the allowable average energy per transmission (averaged over pilots and data). The inequality in the constraint will be met with equality since R_o is increasing in both κ_P and κ_D. We consider causal and noncausal estimators separately in the following.

(1) Causal estimation

For causal $(L, \mathbf{0})$ estimators, it can be shown that the cutoff rate optimizing pilot energy $\kappa_P^\star$ is given by the following one dimensional optimization problem involving only the CSI quality in the pilot slot ω_0

$$\kappa_P^\star = \arg \max_{0 \le \kappa_P \le \kappa_{av} T} \left[\frac{\kappa_{av} T - \kappa_P}{\kappa_{av} T - \kappa_P + (T - 1)} \omega_0(\kappa_P) \right], \quad (26)$$

where $\omega_0(\kappa_P)$ emphasizes dependency on κ_P. The proof follows by substituting for κ_D in terms of the energy constraint into (23), and uses the fact that $\omega_\ell = \alpha^{2\ell}\omega_0$ for any causal estimator.[3]

The optimal pilot energy $\kappa_P^\star$ is implicit in (26), as a particular estimator has not been specified (explicit expressions will be given in the examples below). However, when $|\mathcal{N}|$ is finite, it is clear from the last equality in (10) that ω_0 is a ratio of polynomials in κ_P. Consequently, maximization of (26) involves polynomial rooting. We can write

$$\kappa_P^\star = \{\kappa_P : a_0 + a_1\kappa_P + a_2\kappa_P^2 + \cdots + a_U\kappa_P^U = 0, \ 0 < \kappa_P \le \kappa_{av} T\}, \tag{27}$$

where $a_0, \ldots, a_U$ are coefficients to be determined. A sufficient condition for a closed-form solution is $U \le 4$. Next, we derive the optimal training energy at low and high SNR.

Low SNR

To study the low SNR setting, we start from (10):

$$\begin{aligned} \omega_0 &= \kappa_P R_{h_0\mathbf{y}} \left(\kappa_P R_{\mathbf{hh}} + I_{|\mathcal{N}|}\right)^{-1} R_{h_0\mathbf{y}}^H \\ &\approx \kappa_P R_{h_0\mathbf{y}} \left(I_{|\mathcal{N}|} - \kappa_P R_{\mathbf{hh}}\right) R_{h_0\mathbf{y}}^H \\ &\approx \kappa_P R_{h_0\mathbf{y}} R_{h_0\mathbf{y}}^H = \kappa_P \frac{1 - \alpha^{2TL}}{1 - \alpha^{2T}}, \end{aligned} \tag{28}$$

where the approximations hold as $\kappa_P \to 0$. Substitution of (28) into (26) yields

$$\lim_{\kappa_{av} \to 0} \frac{\kappa_P^\star}{\kappa_{av} T} = \frac{1}{2}, \tag{29}$$

which states that *half* of the total energy per period should be allocated to the pilot symbol.

---

[3] To prove this fact, note that under the GM model, we have $(C_{h_\ell \mathbf{y}})_{1,j} = \sqrt{E_P}\sigma_h^2 \alpha^{|\ell - \mathcal{N}_j T|}$. For causal estimators $\mathcal{N}_j \le 0$, and therefore, $(C_{h_\ell \mathbf{y}})_{1,j} = \sqrt{E_P}\sigma_h^2 \alpha^{\ell - \mathcal{N}_j T} = \alpha^\ell (C_{h_0\mathbf{y}})_{1,j}$. Therefore, $C_{h_\ell \mathbf{y}} = \alpha^\ell C_{h_0\mathbf{y}}$, and from (10), $\omega_\ell = \alpha^{2\ell}\omega_0$.

TABLE 1: The optimal fractional training energy $\kappa_P^\star / \kappa_{av} T$ for arbitrary causal and noncausal estimators under the Gauss-Markov channel.

	Causal $(\mathbf{L}, \mathbf{0})$ estimators	Noncausal $(\mathbf{L}, \mathbf{Z})$ estimators
$\dfrac{\kappa_P^\star}{\kappa_{av} T}$	see (26)	see (35)
$\kappa_{av} \longrightarrow 0$	$= 1/2$	$= 1/2$
$\kappa_{av} \longrightarrow \infty$	$= \dfrac{1}{1 + \sqrt{T-1}}$	$\dfrac{1}{1 + \sqrt{2(T-1)}} \le (\cdot) \le \dfrac{1}{1 + \sqrt{T-1}}$

High SNR

At high SNR, the performance of *any* causal estimator converges to that of the $(\mathbf{1}, \mathbf{0})$ estimator. To see this, start from (10)

$$
\begin{aligned}
\omega_\ell &= \kappa_P \alpha^{2\ell} R_{h_\ell \mathbf{y}} (\kappa_P R_{\mathbf{hh}} + I_{|\mathcal{N}|})^{-1} R_{h_\ell \mathbf{y}}^H \\
&\approx \alpha^{2\ell} R_{h_\ell \mathbf{y}} R_{\mathbf{hh}}^{-1} R_{h_\ell \mathbf{y}}^H = \alpha^{2\ell},
\end{aligned}
\tag{30}
$$

where the approximation holds as $\kappa_P \to \infty$, and where we have exploited the specific tridiagonal structure of $R_{\mathbf{hh}}^{-1}$ to arrive at the last equality. Clearly, (30) matches (11) with (24) at high SNR. Intuitively, the channel state in the most recent pilot transmission $k = mT$ is learnt *perfectly* at high SNR, and this renders older pilots $k = (m-1)T, (m-2)T, \dots$ irrelevant for prediction in the Markov model of (24).

The fractional training energy for any causal estimator at high SNR can now be found by substituting (11) with (24) into (26). We find that

$$
\lim_{\kappa_{av} \to \infty} \frac{\kappa_P^\star}{\kappa_{av} T} = \frac{1}{1 + \sqrt{T-1}}.
\tag{31}
$$

The general properties of $\kappa_P^\star$ for causal estimators are summarized in the left half of Table 1.

Example 1. If $U \le 4$ then closed-form expressions for the optimal training energy (over all SNR) exist. Of particular interest are the $(\mathbf{1}, \mathbf{0})$ and $(\infty, \mathbf{0})$ estimators which represent the limiting cases of causal estimation in our study. For the $(\mathbf{1}, \mathbf{0})$ estimator, the CSI quality ω_0 is given by (11) with (24). Substitution into (26) yields

$$
\kappa_P^\star = \begin{cases} \dfrac{\sqrt{(T-1)(\kappa_{av} T + 1)\left[(\kappa_{av} + 1) T - 1 \right]} - (\kappa_{av} + 1) T + 1}{T - 2}, & T > 2, \\[2mm] \dfrac{1}{2} \kappa_{av} T, & T = 2, \end{cases}
\tag{32}
$$

which agrees with Table 1 in the limiting cases, $\kappa_{av} \to 0$ and $\kappa_{av} \to \infty$, as it must. When $T = 2$, energy is equally divided between pilot and data, as it is in typical transmit-reference schemes.

For the $(\infty, \mathbf{0})$ estimator, the CSI quality is found from (10) to be

$$
\omega_\ell^{(\infty,0)} = \alpha^{2\ell} \frac{\kappa_P - 1 + \sqrt{(1 + \kappa_P)^2 + 4\kappa_P (\alpha^{2T}/1 - \alpha^{2T})}}{\kappa_P + 1 + \sqrt{(1 + \kappa_P)^2 + 4\kappa_P (\alpha^{2T}/1 - \alpha^{2T})}}, \tag{33}
$$

where inversion of the infinite-dimension $C_{\mathbf{yy}}$ matrix has been carried out using the spectral factorization technique [29]. Substituting (33) into (26), it can be verified that as $\alpha \to 1$, the optimal training energy $\kappa_P^\star \to 0$. This is because the $(\infty, \mathbf{0})$ estimator provides an *infinite* number of noisy observations of the time-invariant (in the $\alpha \to 1$ limit) channel. Each observation requires only a minuscule amount of energy in order to exploit the infinite (in the limit) diversity gain. As $\alpha \to 0$, $\kappa_P^\star$ converges to the $\kappa_P^\star$ of the $(\mathbf{1}, \mathbf{0})$ estimator

in (32) (this follows since $\omega_\ell^{(\infty,0)}$ converges to $\omega_\ell^{(1,0)}$): for a rapidly fading channel, only the most recent pilot proves useful. For arbitrary α, the optimal training energy is found by solving (26) with (33). For brevity, we use the coefficient notation of (27), for which we get

$$
a_0^{(\infty,0)} = -\kappa_{av}^2 T^2 \left[(\kappa_{av} + 1) T - 1 \right]^2,
$$

$$
a_1^{(\infty,0)} = 2\kappa_{av} T \left[(\kappa_{av} + 1) T - 1 \right] \left[(2\kappa_{av} + 1) T - 1 \right],
$$

$$
a_2^{(\infty,0)} = -6\kappa_{av} T \left[(\kappa_{av} + 1) T - 1 \right], \tag{34}
$$

$$
a_3^{(\infty,0)} = \frac{4\alpha^{2T}}{1 - \alpha^{2T}} (T - 1)^2 + 2T(T - 1) + 4\kappa_{av} T,
$$

$$
a_4^{(\infty,0)} = (T - 2) T.
$$

Note that $U = 4$, ensuring a closed form solution. Properties of the $(\mathbf{1}, \mathbf{0})$ and $(\infty, \mathbf{0})$ estimators, representing the limiting

TABLE 2: The optimal fractional training energy $\kappa_P^\star/\kappa_{\mathrm{av}}T$ for the $(1,0)$ and $(\infty,0)$ causal estimators, and the $(1,1)$ and (∞,∞) noncausal estimators, under the Gauss-Markov channel.

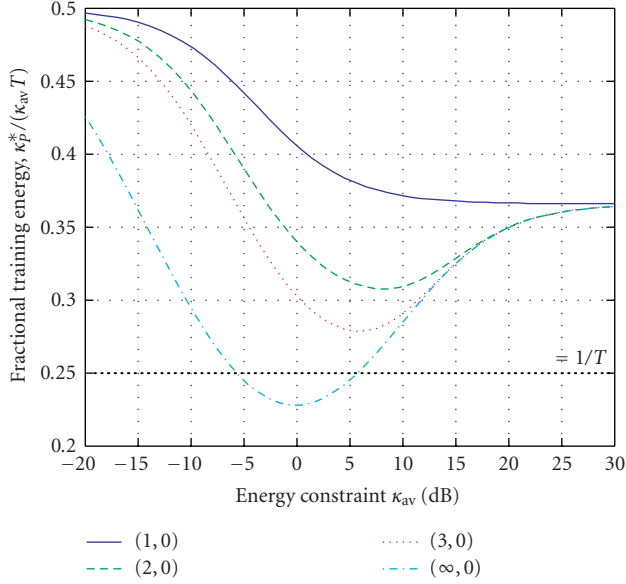|	$(1,0)$	$(\infty,0)$	$(1,1)$	(∞,∞)
$\dfrac{\kappa_P^\star}{\kappa_{\mathrm{av}}T}$	See (32)	See (34)	= N/A	= N/A
$\alpha \longrightarrow 0$	No dependency	$\longrightarrow \dfrac{\kappa_P^\star}{\kappa_{\mathrm{av}}T}$ of $(1,0)$	$\longrightarrow \dfrac{1}{1+\sqrt{(T-1)(\kappa_{\mathrm{av}}T+1)/(T(\kappa_{\mathrm{av}}+1)-1)}}$	$\longrightarrow \dfrac{\kappa_P^\star}{\kappa_{\mathrm{av}}T}$ of $(1,1)$
$\alpha \longrightarrow 1$	No dependency	$\longrightarrow 0$	$\longrightarrow \dfrac{1}{1+\sqrt{(T-1)(2\kappa_{\mathrm{av}}T+1)/(T(\kappa_{\mathrm{av}}+1)-1)}}$	$\longrightarrow 0$

FIGURE 1: The optimal fractional training energy $\kappa_P^\star/(\kappa_{\mathrm{av}}T)$ for several causal estimators when $\alpha = 0.99$, $M = 8$, and $T = 4$. The dashed line $= 1/T$ is the fractional training energy under a static (constant) energy allocation. At high SNR, the fractional energy saturates to $1/(1+\sqrt{T-1})$.

cases of causal estimation, are summarized on the left side of Table 2.

In Figure 1, we plot the fractional training energy for the $(1,0)$, $(2,0)$, $(3,0)$, and $(\infty,0)$ estimators as a function of the energy constraint κ_{av} for $M = 8$, $T = 4$, and $\alpha = 0.99$.[4] It is seen that as more pilots are exploited, less training energy is required. The *fractional* training energy is nonmonotonic in κ_{av} for the multipilot estimators, though $\kappa_P^\star$ is monotonic.[5]

---

[4] A closed-form solution for $\kappa_P^\star$ under the $(2,0)$ estimator also exists (i.e., $U \leq 4$), but it has been omitted for brevity. For the $(3,0)$ estimator, a sixth-order polynomial in $\kappa_P^\star$ ensues.

[5] Using the Kuhn-Tucker conditions, it can be shown that the fractional energy allocation is nonmonotonic when the channel estimation is better (when more pilots are used, for larger α, and/or for smaller T). For example, for the $(\infty,0)$ estimator, it can be shown that the fractional energy allocation is nonmonotonic according to

$$\frac{1+\alpha^{2T}}{1-\alpha^{2T}} - \sqrt{T-1} \underset{\substack{\text{monotonic}}}{\overset{\substack{\text{non-monotonic}}}{\underset{<}{\gtreqless}}} 0.$$

(2) Noncausal estimation

The optimal energy allocation is generally not available in closed-form for noncausal $(\mathbf{L},\mathbf{Z})$ estimators. In general, it can be expressed as

$$\kappa_P^\star = \arg \max_{\kappa_P + \kappa_D(T-1)=\kappa_{\mathrm{av}}T} R_o. \qquad (35)$$

We start by considering $\kappa_P^\star$ in the limiting SNR cases. We obtain a closed-form solution at low SNR, and simple, but useful, bounds at high SNR.

Low SNR

At low SNR, the CSI quality (10) is simplified using a technique similar to that used in (28) for causal estimators. We find that

$$\omega_\ell \approx \frac{1}{1-\alpha^{2T}}\left[\alpha^{2\ell}\left(1-\alpha^{2TL}\right) + \frac{\alpha^{2T}\left(1-\alpha^{2TZ}\right)}{\alpha^{2\ell}}\right]\kappa_P, \quad (36)$$

where the approximation holds as $\kappa_P \to 0$. Although this expression depends on ℓ, substitution into (35) nevertheless yields a closed-form expression for $\kappa_P^\star$. After taking the limit, we get

$$\lim_{\kappa_{\mathrm{av}}\to 0} \frac{\kappa_P^\star}{\kappa_{\mathrm{av}}T} = \frac{1}{2}, \qquad (37)$$

implying once again that half of the available energy per period should be allocated to the pilot symbol at low SNR.

High SNR

At high SNR, the performance of any noncausal estimator converges to that of the $(1,1)$ estimator (the proof is similar to the one used to derive (30) for causal estimators). Using this fact, we substitute (12) with (24) into (35), and consider the limiting cases of rapid ($\alpha \to 0$) and slow ($\alpha \to 1$) fading, which provide upper and lower bounds on $\kappa_P^\star$. We get

$$\frac{1}{1+\sqrt{2(T-1)}} \leq \frac{\kappa_P^\star}{\kappa_{\mathrm{av}}T} \leq \frac{1}{1+\sqrt{T-1}}, \qquad (38)$$

where the lower bound is met with equality as $\alpha \to 1$, and the upper bound as $\alpha \to 0$ (the technique used to evaluate these limits will be made clear shortly, in the arguments leading to (42)). Comparison of (38) to (31) reveals that a noncausal
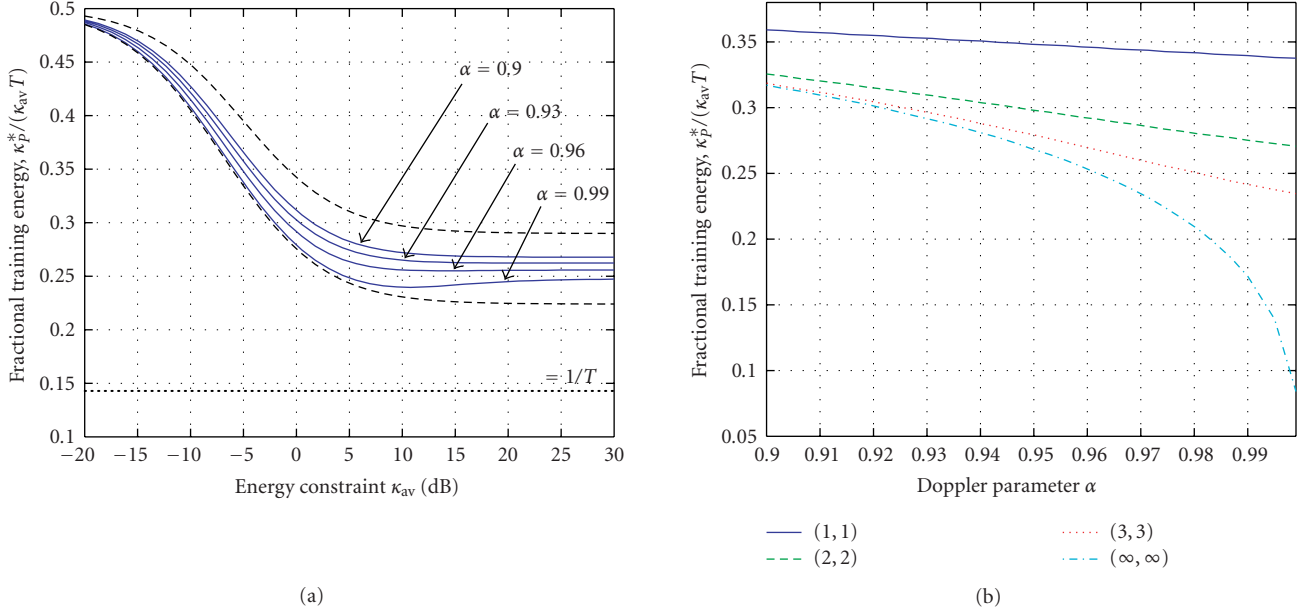
(a)

(b)

FIGURE 2: (a) The fractional training energy for the $(1, 1)$ estimator as a function of the energy constraint κ_{av} for several values of α when $M = 8$ and $T = 7$. Also shown (dashed lines) are the lower and upper bounds on $\kappa_P^\star/\kappa_{av}T$ as determined from (42). (b) The fractional training energy $\kappa_P^\star/\kappa_{av}T$ versus the Doppler parameter α for the $(1, 1)$, $(2, 2)$, $(3, 3)$, and (∞, ∞) estimators at an SNR of $\kappa_{av} = 0$ dB when $M = 4$ and $T = 4$.

estimator never uses more training energy than a causal one at high SNR (for fixed T). General properties of $\kappa_P^\star$ for non-causal estimators are summarized in the right half of Table 1.

Example 2. We start with an analysis of the $(1, 1)$ estimator which is valid for all SNR. Simplifying (12) for the Gauss-Markov model, we get

$$\omega_\ell^{(1,1)} = \frac{\kappa_P^2(\alpha^{2\ell} + \alpha^{2(T-\ell)} - 2\alpha^{2T}) + \kappa_P(\alpha^{2\ell} + \alpha^{2(T-\ell)})}{(\kappa_P + 1)^2 - \kappa_P^2\alpha^{2T}}. \tag{39}$$

Next, we evaluate the CSI quality under rapid and slow fading. For rapid fading, we get

$$\lim_{\alpha \to 0} \omega_\ell^{(1,1)} = \frac{\kappa_P}{1 + \kappa_P} \max(\alpha^{2\ell}, \alpha^{2(T-\ell)}), \tag{40}$$

and for slow fading we get

$$\lim_{\alpha \to 1} \omega_\ell^{(1,1)} = \frac{\kappa_P}{1/2 + \kappa_P}, \quad \forall \ell. \tag{41}$$

Substitution of (40) and (41) into (35) yields closed-form solutions. We get

$$\frac{1}{1 + \sqrt{(T-1)(2\kappa_{av}T+1)/(T(\kappa_{av}+1)-1)}}$$
$$\leq \frac{\kappa_P^\star}{\kappa_{av}T} \tag{42}$$
$$\leq \frac{1}{1 + \sqrt{(T-1)(\kappa_{av}T+1)/(T(\kappa_{av}+1)-1)}}.$$

In Figure 2(a) we plot the fractional training energy for the $(1, 1)$ estimator as a function of the energy constraint κ_{av} for several values of α when $M = 8$ and $T = 7$. Also shown (dashed lines) are the lower and upper bounds on $\kappa_P^\star/\kappa_{av}T$ from (42). Although the upper bound was derived for the condition $\alpha \to 0$, it is seen to be useful for the practical range of α.

Next, we consider the (∞, ∞) estimator. The CSI quality is found to be

$$\omega_\ell^{(\infty,\infty)} = 1 - \frac{1 + \kappa_P + \alpha^{2T}(\kappa_P - 1) - \kappa_P(\alpha^{2\ell} + \alpha^{2(T-\ell)})}{\sqrt{(1 - \alpha^{2T})[(\kappa_P + 1)^2 - (\kappa_P - 1)^2\alpha^{2T}]}}, \tag{43}$$

which follows from (10) after applying spectral factorization. To determine bounds on the optimal training energy, we again consider the cases of slow and rapid fading. For slow fading, we apply L'Hôpital's rule to (43), and obtain

$$\lim_{\alpha \to 1} \omega_\ell^{(\infty,\infty)} = 1, \qquad \kappa_P > 0, \tag{44}$$

and it follows from (35) that $\kappa_P^\star \to 0$. For rapid fading $(\alpha \to 0)$, it is seen that $\omega_\ell^{(\infty,\infty)}$ converges to $\omega_\ell^{(1,1)}$ (i.e., to the expression on the right hand side of (40)). Therefore, $\kappa_P^\star$ converges to the $\kappa_P^\star$ of the $(1, 1)$ estimator. In Figure 2(b) we plot the fractional training energy $\kappa_P^\star/\kappa_{av}T$ versus Doppler α for the $(1, 1)$, $(2, 2)$, $(3, 3)$ and (∞, ∞) estimators at an SNR of $\kappa_{av} = 0$ dB when $M = 4$ and $T = 4$. For smaller values of α,

the $(2, 2)$ estimator provides most of the reduction in the required training energy, and gains saturate with more sophisticated estimators. For large α, the (∞, ∞) estimator takes advantage of the high-order diversity gain available over the slowly varying channel, and requires considerably less energy than the competing estimators. Properties of the $(1, 1)$ and (∞, ∞) estimators, which represent the limiting cases of noncausal estimation, are summarized on the right side of Table 2.

3.2. Training period

In this section we consider the optimal period (equivalently, frequency) with which pilot symbols should be inserted into the symbol stream. The optimal value of T depends on the normalized Doppler α, the cardinality of the input M, the energy constraint κ_{av}, the energy allocation (e.g., the optimal allocation as in Section 3.1 or a static allocation $\kappa_D = \kappa_P = \kappa_{av}$), and the particular estimator employed at the receiver. However, we will see that the analysis simplifies greatly in the high SNR setting. We will again find it convenient to distinguish between the cases of causal and noncausal estimation.

(1) Causal estimation

At high SNR, the optimal training period for any causal estimator is found from (23). Taking the argmax in T and letting $\kappa_{av} \to \infty$ we get

$$T_C \triangleq \arg \max_{2 \leq x < \infty, x \in \mathbb{N}} \left[\prod_{\ell=1}^{x-1} \frac{1}{M} \sum_{m=0}^{M-1} \frac{1 - \alpha^{2\ell}}{1 - \alpha^{2\ell} \cos^2(\pi m/M)} \right]^{-1/x}, \tag{45}$$

where we have again used the convergence of all causal estimators to the $(1, 0)$ estimator at high SNR. Equation (45) depends only on M and α; it is independent of the particular

estimator used and the energy allocation strategy. Although motivated by the high SNR setting, it will be seen that (45) provides good approximation to the optimal training period over a wide range of SNR.

Example 3. We study the applicability of the training period rule of (45) to $(1, 0)$ and $(\infty, 0)$ estimators at *finite* values of SNR. A comparison is given in Table 3 for QPSK (i.e., $M = 4$). The second and third columns are the optimal training period for the $(1, 0)$ estimator under the static and optimal energy allocations, respectively (determined numerically). The fourth and fifth columns are the training period for the $(\infty, 0)$ estimator under static and optimal energy allocations (determined numerically), and the sixth column is the optimal training period at high SNR determined from (45). The optimal training period for either estimators, under either energy allocation strategy, is seen to converge to T_C as the SNR increases, which is expected. It is seen that convergence occurs sooner when the fading becomes more rapid. For example, for $\alpha = 0.80$, the training period predicted by (45) is correct for SNRs as small as 0 dB (for either the $(1, 0)$ or $(\infty, 0)$ estimators and under either energy allocation strategy). For $\alpha = 0.95$, T_C is exact for SNRs as low as 10 dB, and for $\alpha = 0.99$, T_C is correct to an SNR of 20 dB. For a fixed estimator, it is seen that the optimal training period can vary greatly depending on the energy allocation strategy—at least for smaller κ_{av} and larger α. For example, when $\alpha = 0.99$ and $\kappa_{av} = 0$ dB, the optimal training period varies from 10 (under constant allocation) to 20 (under optimal allocation).

(2) Noncausal estimation

Similarly, we find the optimal training period for any noncausal estimator by considering the high SNR setting. Letting $\kappa_{av} \to \infty$ in (23), we get

$$T_{NC} \triangleq \arg \max_{2 \leq x < \infty, x \in \mathbb{N}} \left[\prod_{\ell=1}^{x-1} \frac{1}{M} \sum_{m=0}^{M-1} \frac{(1 - \alpha^{2\ell})(1 - \alpha^{2(x-\ell)})}{1 - \alpha^{2x} - (\alpha^{2\ell} + \alpha^{2(x-\ell)} - 2\alpha^{2x}) \cos^2(\pi m/M)} \right]^{-1/x}. \tag{46}$$

Example 4. The right side of Table 3 illustrates the training period for noncausal estimators. The seventh and eighth columns of the table are the optimal training period for the $(1, 1)$ estimator under static and optimal energy allocations, respectively (determined numerically). The ninth and tenth columns are the training period for the (∞, ∞) estimator under static and optimal energy allocations (determined numerically), and the eleventh column is the optimal training period at high SNR determined from (46). Again, we note that T_{NC} provides good approximation to the optimal training period for larger SNR and for more rapid fading. The table reflects intuition: the more predict-

able the channel (larger α), the less frequently training is required (larger T). However, the table generally indicates that more sophisticated estimators (e.g., the (∞, ∞)) require more frequent training symbols than simpler ones (e.g., the $(1, 1)$). To explain this result we refer to (23), which shows that the optimal T is determined not directly by the quality of the estimator, but rather by how quickly the cutoff rate in the ℓth subchannel diminishes in ℓ ($1 \leq \ell \leq \lceil T - 1/2 \rceil$ for noncausal estimators). If the better estimator causes the biased sum of (23) to degrade more quickly in ℓ, then T will be smaller for the better estimator.

TABLE 3: The optimal training period under the Gauss-Markov fading channel for QPSK ($M = 4$). The left half of the table is a study for causal estimators: the $(1, 0)$ estimator under the static (1st column) and optimal (2nd) energy allocations, the $(\infty, 0)$ estimator under static (3rd) and optimal (4th) energy allocations, and the optimal training period at high SNR (5th column) determined from (45). The right half of the table is a study for noncausal estimators: the $(1, 1)$ estimator under the static (6th column) and optimal (7th) energy allocations, the (∞, ∞) estimator under static (8th) and optimal (9th) energy allocations, and the optimal training period at high SNR (10th column) determined from (46).

	Causal estimation					Noncausal estimation				
	$T_{(1,0)}$	$T^\star_{(1,0)}$	$T_{(\infty,0)}$	$T^\star_{(\infty,0)}$	T_C	$T_{(1,1)}$	$T^\star_{(1,1)}$	$T_{(\infty,\infty)}$	$T^\star_{(\infty,\infty)}$	T_{NC}
				$\alpha = \mathbf{0.80}$						
$\kappa_{av} = 0\,\text{dB}$	3	3	3	3	**3**	3	4	3	3	**3**
$\kappa_{av} = 10\,\text{dB}$	3	3	3	3	**3**	3	3	3	3	**3**
$\kappa_{av} = 20\,\text{dB}$	3	3	3	3	**3**	3	3	3	3	**3**
				$\alpha = \mathbf{0.95}$						
$\kappa_{av} = 0\,\text{dB}$	5	8	4	5	**4**	6	10	5	7	**5**
$\kappa_{av} = 10\,\text{dB}$	4	4	4	4	**4**	5	6	5	6	**5**
$\kappa_{av} = 20\,\text{dB}$	4	4	4	4	**4**	5	5	5	5	**5**
				$\alpha = \mathbf{0.99}$						
$\kappa_{av} = 0\,\text{dB}$	10	20	6	10	**7**	12	28	8	16	**10**
$\kappa_{av} = 10\,\text{dB}$	8	9	7	8	**7**	11	14	10	13	**10**
$\kappa_{av} = 20\,\text{dB}$	7	7	7	7	**7**	10	11	10	11	**10**

3.3. Performance analysis

We now examine the effect of optimal training on the cutoff rate. In Figure 3(a) we plot the QPSK cutoff rate under the $(1, 0)$ estimator for Doppler values $\alpha = \{0.90, 0.95, 0.99\}$. For fixed α, we plot the cutoff rate under (a) optimization over the energy allocation and training period, $\kappa_P = \kappa_P^\star, \kappa_D = \kappa_D^\star$, and $T = T^\star$, (b) optimization over the training period but not the energy allocation, $T = T^\star, \kappa_D = \kappa_P = \kappa_{av}$,[6] and (c) the unoptimized case, $\kappa_D = \kappa_P = \kappa_{av}$, $T = T_C$ (the training period is fixed at the high SNR optimal value determined from (45)). The merits of optimal allocation increase with the channel predictability: when $\alpha = 0.99$ there is a ~ 2 dB gain at $\kappa_{av} = 0$ dB, but when $\alpha = 0.9$ the gain is only a fraction of a dB. In each case, we find that it is the energy allocation, not assignment of the training period, that provides most of the gain in optimized training. This is due in part to our choice of $T = T_C$, which is optimal at high SNR. In Figure 3(b) we plot the impact of an arbitrary choice of T on the cutoff rate under constant energy allocation, $\kappa_D = \kappa_P = \kappa_{av} = 20$ dB. The degradation may be significant when T is chosen suboptimally.

To determine the merits of more sophisticated estimators, we compare the cutoff rate under the simplest and the most complex causal ($(1, 0)$ and $(\infty, 0)$) and noncausal

($(1, 1)$ and (∞, ∞)) estimators in Figure 4(a) for $\alpha = 0.98$ and a constant energy allocation with unoptimized choice of T (we choose $T = T_C$ for the causal estimators or $T = T_{NC}$ for the noncausal estimators). Therefore, the curves in the figure represent the largest increase in cutoff rate due to the use of a sophisticated estimator in place of a simpler one. At small SNR there is a ~ 2 dB gain in using more sophisticated estimators. However, this gain is seen to diminish at high SNR (as expected for the GM model). We repeat the figure, but with optimized energy and training assignments, in Figure 4(b). Remarkably, the energy saving for using the $(\infty, 0)$ estimator in place of the $(1, 0)$ (or the (∞, ∞) in place of the $(1, 1)$) is seen to be a fraction of a dB over the entire SNR range. Energy optimization reduces the need for sophisticated estimators in the GM model.

4. OPTIMAL TRAINING FOR JAKES MODEL

In this section we study optimized training for the Jakes channel correlation [30]. While the GM model studied in the last section provides straightforward analytic results, the Jakes model is known to be an accurate and experimentally validated model in dense scattering environments. The analysis in this section will be used to validate and refine the design paradigms derived in the last section. For the Jakes model we have

---

[6] Note that $T^\star$ will generally be different in (a) and (b) since the energy allocation strategy is different in each case. Nevertheless, we use the notation $T = T^\star$ to denote the optimal training period for both.
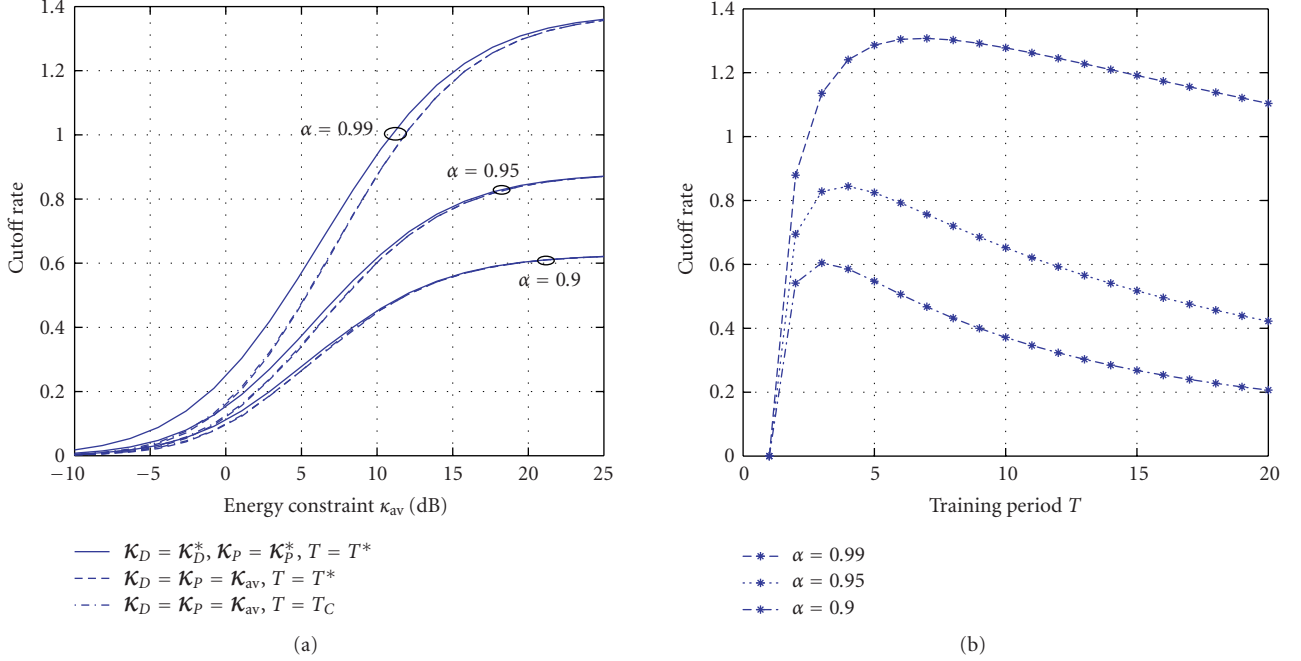
$$R_h(\tau) = J_0\big(2\pi f_D T_D \tau\big), \qquad (47)$$

FIGURE 3: The QPSK ($M = 4$) cutoff rate under the $(\mathbf{1}, \mathbf{0})$ estimator for Doppler values $\alpha = \{0.90, 0.95, 0.99\}$. (a) For fixed α, the solid line represents optimization over the energy allocation and training period $\kappa_P = \kappa_P^\star, \kappa_D = \kappa_D^\star$, and $T = T^\star$, the dashed line represents optimization over the training period, but not the energy allocation, $\kappa_D = \kappa_P = \kappa_{av}$ and $T = T^\star$, and the dashed-dotted line represents the unoptimized case, $\kappa_D = \kappa_P = \kappa_{av}$, and $T = T_C$ (the training period is fixed at the high SNR value determined from (45)). (b) The effect of a suboptimal assignment of the training period T on cutoff rate under the constant energy allocation, $\kappa_D = \kappa_P = \kappa_{av} = 20$ dB.

where $J_0(\cdot)$ is the zero-order Bessel function of the first kind, and $f_D T_D > 0$ is the normalized Doppler parameter. The cutoff rate is given by (23) with CSI quality (10) as before, but now under the channel correlation of (47).

In the following simulation we consider mobile speeds of $\{12, 120\}$ Km/h which correspond to the Doppler parameters $f_D T_D = \{1/1000, 1/100\}$ at a carrier frequency of 900 MHz and symbol period of $T_D = 10\,\mu$-sec [31, pages 141–143]. In Figure 5, we plot the cutoff rate for both values of $f_D T_D$ under (a) optimized energy and training period assignments, $\kappa_P = \kappa_P^\star, \kappa_D = \kappa_D^\star, T = T^\star$, and under the optimized training period only, $\kappa_P = \kappa_D = \kappa_{av}, T = T^\star$. Behavior is seen to be qualitatively similar to that in Figure 3(a) for the GM model: energy-optimized training is seen to provide a noticeable increase in the cutoff rate for slower fading channels, but not for rapidly fading channels. Further, optimized training provides the largest savings at low SNR (~ 3 dB at $\kappa_{av} = 0$ dB), but is of diminishing benefit as SNR increases. The vehicle speeds tested represent extreme cases. For intermediate speeds (i.e., values of $f_D T_D$) performance increases smoothly both in the cutoff rate and in the gains realized by optimal resource allocation.

In Section 3 we made use of the Markov property of the channel in several instances, exploiting the convergence of the $(\mathbf{L}, \mathbf{0})$ estimator to the $(\mathbf{1}, \mathbf{0})$ estimator (alt., the $(\mathbf{L}, \mathbf{Z})$ estimator to the $(\mathbf{1}, \mathbf{1})$ estimator) at high SNR. To test the degree to which this property holds under the Jakes model, we plot the cutoff rate for QPSK input and optimized training

(both the energy and training period have been optimized) in Figure 6 when $f_D T_D = 1/100$. In comparison to Figure 4(b), the counterpart figure for the GM model, we notice a similar qualitative behavior at low SNR. However, we notice differences at high SNR: more sophisticated estimators are seen to be useful at high SNR under optimized training (for both causal and noncausal estimation). For example, there is now a ~ 3 dB gain in going from the $(\mathbf{1}, \mathbf{0})$ estimator to the $(\mathbf{2}, \mathbf{0})$ estimator at an SNR of $\kappa_{av} = 15$ dB. Further, performance of the $(\mathbf{L}, \mathbf{0})$ (alt., $(\mathbf{L}, \mathbf{Z})$) estimator does not converge to that of the $(\mathbf{1}, \mathbf{0})$ (alt., $(\mathbf{1}, \mathbf{1})$) estimator at high SNR. In general it is seen that the largest gain is achieved in going from the $(\mathbf{1}, \mathbf{0})$ estimator to the $(\mathbf{2}, \mathbf{0})$ estimator (alt., from the $(\mathbf{1}, \mathbf{1})$ to the $(\mathbf{2}, \mathbf{2})$), after which adding more pilots provides diminishing returns to the cutoff rate.

5. DISCUSSION AND FUTURE WORK

We have considered cutoff rate optimal training within a PSAM framework for time-selective Rayleigh flat-fading channels. For M-PSK inputs, we have derived in (23) a simple expression for the interleaved cutoff rate that is parameterized by an arbitrary channel correlation, channel estimation scheme, and the allocation of power and bandwidth to training symbols. Using this expression, we have derived analytic rules for the optimal allocation of resources for the Gauss-Markov fading channel, the basic properties of which are summarized in Tables 1 and 2. For example, it was seen
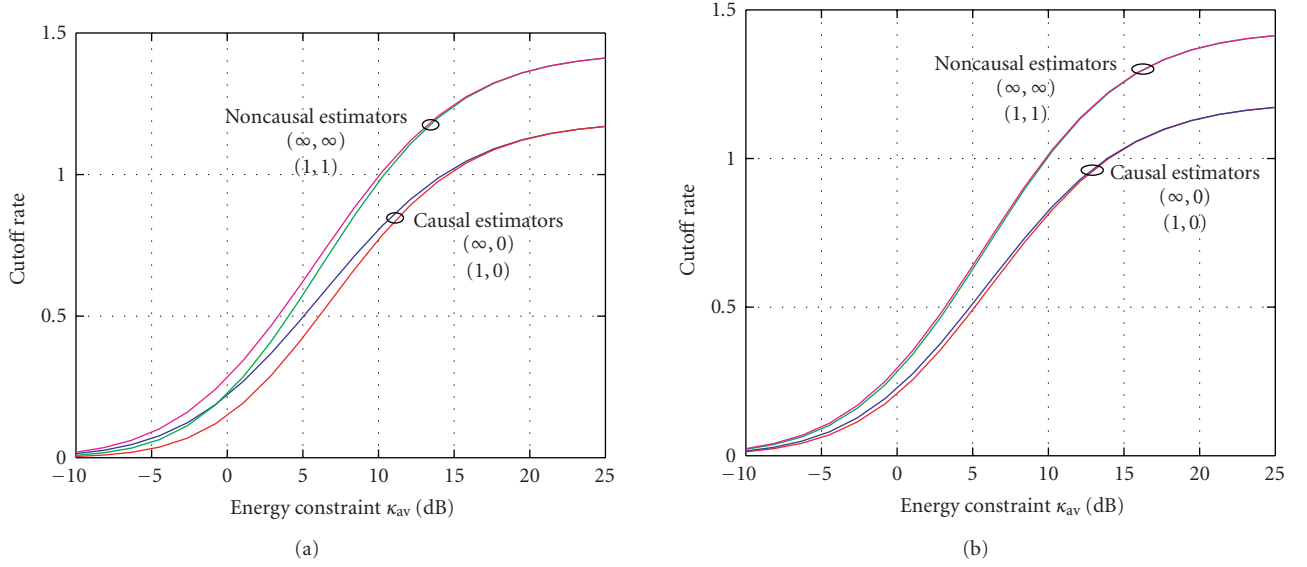
FIGURE 4: The QPSK ($M = 4$) cutoff rate for $\alpha = 0.98$ for the $(\mathbf{1}, \mathbf{0})$, $(\infty, \mathbf{0})$, $(\mathbf{1}, \mathbf{1})$, and (∞, ∞) estimators under (a) the constant energy allocation ($\mathcal{K}_D = \mathcal{K}_P = \mathcal{K}_{\text{av}}$) and unoptimized training period (we choose $T = T_C$ for the causal estimators, and $T = T_{\text{NC}}$ for noncausal estimators), and (b) with optimization over both the energy allocation and training period, $\mathcal{K}_D = \mathcal{K}_D^\star$, $\mathcal{K}_P = \mathcal{K}_P^\star$, and $T = T^\star$.
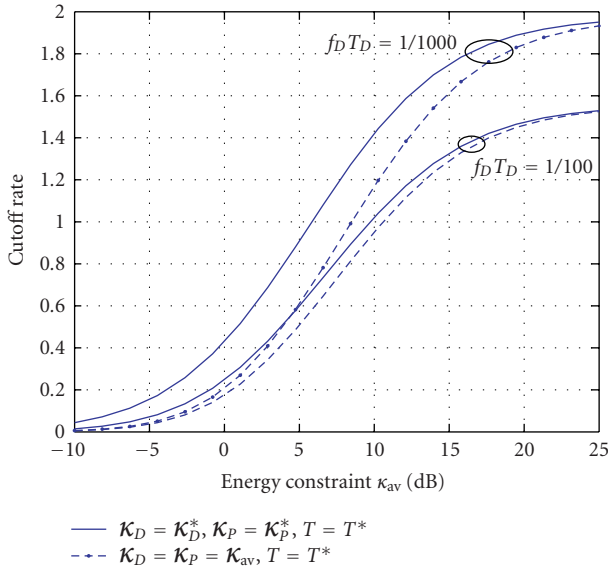

FIGURE 5: The QPSK ($M = 4$) cutoff rate under the $(\mathbf{1}, \mathbf{0})$ estimator for Doppler values, $f_D T_D = \{1/1000, 1/100\}$. For fixed $f_D T_D$, the solid line represents optimization over the energy allocation and training period, $\mathcal{K}_P = \mathcal{K}_P^\star$, $\mathcal{K}_D = \mathcal{K}_D^\star$, and $T = T^\star$, and the dashed line represents optimization over the training period but not the energy allocation, $\mathcal{K}_D = \mathcal{K}_P = \mathcal{K}_{\text{av}}$ and $T = T^\star$.
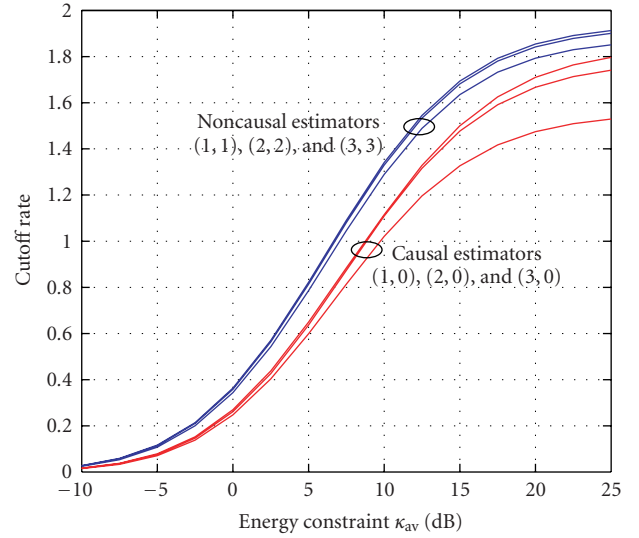
FIGURE 6: The QPSK ($M = 4$) cutoff rate for $f_D T_D = 0.01$ for the $(\mathbf{1}, \mathbf{0})$, $(\mathbf{2}, \mathbf{0})$, and $(\mathbf{3}, \mathbf{0})$ causal estimators, and the $(\mathbf{1}, \mathbf{1})$, $(\mathbf{2}, \mathbf{2})$, and $(\mathbf{3}, \mathbf{3})$ noncausal estimators, under optimization over both the energy allocation and training period, $\mathcal{K}_D = \mathcal{K}_D^\star$, $\mathcal{K}_P = \mathcal{K}_P^\star$, and $T = T^\star$.

that at low SNR, half the total available energy should go to training, while at high SNR, a noncausal estimator never uses more energy than a causal one (with equality when the fading is rapid). We have provided expressions for the training period that, while optimal at high SNR, were seen to provide solid guidance across a wide range of SNR and Doppler values (see (45), (46), and Table 3). Next, we studied optimal training for the Jakes model. It was seen that insights derived from the Gauss-Markov model were predictive at low SNR and for simpler estimators, but were not useful at higher

SNR with more sophisticated multi-pilot estimators. Using the Jakes model, it was seen that multi-pilot estimators were indeed useful at high SNR, with the largest gains coming with the addition of the first few pilots. Both the Gauss-Markov and Jakes models indicated that while optimized energy allocation is useful at low SNR, it is not useful at high SNR.

Among related work on optimal training for point-to-point time-selective Rayleigh flat-fading channels, only [14] offers analytic results for the optimal training period and energy allocation in a similar framework. The authors consider Gaussian inputs and derive a lower bound on channel capacity that relates estimation error directly to system performance in similar fashion to (23) (however, the capacity formulation requires an expectation over the random channel that cannot be evaluated directly). The (∞, ∞) estimator is used. As a representative example, we compare the optimal energy allocation result of [14] at high SNR for any (bandlimited) Doppler spectrum to our result at high SNR for the GM model. Our result is given by (38) where the lower and upper bounds are achieved in limit of static and i.i.d. fading (resp.). Reference [14] derives

$$\left(\frac{\kappa_P^\star}{\kappa_{\mathrm{av}} T} \right)_{[14]} = \frac{1}{1 + \sqrt{\beta(T/T - 1)}}, \tag{48}$$

where $\beta > 0$ is a constant. This represents a different behavior for large T, as $\kappa_P^\star / \kappa_{\mathrm{av}} T$ saturates to some value > 0 for [14], but not for our model. A comparison of optimal training period results is not as straightforward, as [14] considers pilots to be samples of an underlying continuous time channel: an interpretation that is excluded here.

It is of further interest to study cutoff rate optimal training under nonsymmetric inputs such as M-QAM (an analysis for perfect CSI appears in [32]), and generalizations to the MIMO setting and the case where the transmitter has only statistical knowledge of the Doppler spectrum.

APPENDIX

DERIVATION OF THE CUTOFF RATE, (20)

We start from (19). Note that

$$\int_{\mathbf{y}} \int_{\hat{\mathbf{h}}} \left[\sum_{S \in \mathcal{S}} Q(S) \sqrt{P(\mathbf{y}, \hat{\mathbf{h}} \mid S)} \right]^2 d\hat{\mathbf{h}} \, d\mathbf{y}$$
$$= \sum_{V, W \in \mathcal{S}} Q(V) Q(W) \mathcal{E}_{\hat{\mathbf{h}}} \tag{A.1}$$
$$\times \left[\int_{\mathbf{y}} \sqrt{P(\mathbf{y} \mid V, \hat{\mathbf{h}}) P(\mathbf{y} \mid W, \hat{\mathbf{h}})} d\mathbf{y} \right].$$

Note that $\mathbf{y} \mid V, \hat{\mathbf{h}} \sim \mathcal{CN}(\mathbf{u}_V, \Sigma_V)$ and that $\mathbf{y} \mid W, \hat{\mathbf{h}} \sim \mathcal{CN}(\mathbf{u}_W, \Sigma_W)$, where $\mathbf{u}_W \triangleq \sqrt{E_D} W \hat{\mathbf{h}}$, $\mathbf{u}_V \triangleq \sqrt{E_D} V \hat{\mathbf{h}}$, $\Sigma_V \triangleq \sigma_h^2 E_D V \tilde{\Sigma} V^H + \sigma_n^2 I_{N'}$, and $\Sigma_W \triangleq \sigma_h^2 E_D W \tilde{\Sigma} W^H + \sigma_n^2 I_{N'}$. We get

$$\int_{\mathbf{y}} \sqrt{P(\mathbf{y} \mid V, \hat{\mathbf{h}}) P(\mathbf{y} \mid W, \hat{\mathbf{h}})} d\mathbf{y}$$
$$= e^{-(1/2)(\mathbf{u}_V - \mathbf{u}_W)^H (\Sigma_V + \Sigma_W)^{-1}(\mathbf{u}_V - \mathbf{u}_W)} \frac{|\Sigma_V|^{1/2} |\Sigma_W|^{1/2}}{|(\Sigma_V + \Sigma_W)/2|}. \tag{A.2}$$

Next, we take the expectation of (A.2) with respect to $\hat{\mathbf{h}} \sim \mathcal{CN}(0, \sigma_h^2 \hat{\Sigma})$. We get

$$\frac{|\sigma_n^2 I_{N'} + \sigma_h^2 E_D V \tilde{\Sigma} V^H|^{1/2} |\sigma_n^2 I_{N'} + \sigma_h^2 E_D W \tilde{\Sigma} W^H|^{1/2}}{|\sigma_n^2 I_{N'} + (1/2)\sigma_h^2 E_D (V \tilde{\Sigma} V^H + W \tilde{\Sigma} W^H) + (1/4)\sigma_h^2 E_D (V - W)\hat{\Sigma}(V - W)^H|}. \tag{A.3}$$

Dividing the numerator and denominator by σ_n^2 and substituting the result into (19) yields (20).

REFERENCES

[1] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels [Mobile Radio]," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 4, pp. 686–693, 1991.

[2] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions," *IEEE Signal Processing Magazine*, vol. 21, no. 6, pp. 12–25, 2004.

[3] M. J. Garcia and J. M. Páez-Borrallo, "Tracking of time misalignments for OFDM systems in multipath fading channels," *IEEE Transactions on Consumer Electronics*, vol. 48, no. 4, pp. 982–989, 2002.

[4] W. Kuo and M. P. Fitz, "Frequency offset compensation of pilot symbol assisted modulation in frequency flat fading," *IEEE Transactions on Communications*, vol. 45, no. 11, pp. 1412–1416, 1997.

[5] X. Cai and G. Giannakis, "Adaptive PSAM accounting for channel estimation and prediction errors," *IEEE Transactions on Wireless Communications*, vol. 4, no. 1, pp. 246–256, 2005.

[6] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading

channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 5, pp. 1403–1418, 2004.

[7] X. Dong and L. Xiao, "Symbol error probability of two-dimensional signaling in Ricean fading with imperfect channel estimation," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 2, pp. 538–549, 2005.

[8] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[9] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for frequency-selective block-fading channels," *IEEE Transactions on Information Theory*, vol. 48, no. 8, pp. 2338–2353, 2002.

[10] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Transactions on Signal Processing*, vol. 51, no. 5, pp. 1351–1366, 2003.

[11] S. Ohno and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 2138–2145, 2004.

[12] J. Baltersee, G. Fock, and H. Meyr, "An information theoretic foundation of synchronized detection," *IEEE Transactions on Communications*, vol. 49, no. 12, pp. 2115–2123, 2001.

[13] M. Medard, I. Abou-Faycal, and U. Madhow, "Adaptive coded modulation without channel feedback for pilot symbol assisted modulation," in *Proceedings of the 38th Annual Allerton Conference on Communication, Control and Computing*, Monticello, Ill, USA, October 2002.

[14] S. Ohno and G. Giannakis, "Average-rate optimal PSAM transmissions over time-selective fading channels," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 712–720, 2002.

[15] E. Arikan, "Upper bound on the cutoff rate of sequential decoding," *IEEE Transactions on Information Theory*, vol. 34, no. 1, pp. 55–63, 1988.

[16] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2619–2692, 1998.

[17] S. Jamali and T. Le-Ngoc, *Coded-Modulation Techniques for Fading Channels*, Kluwer Academic, Dordrecht, The Netherlands, 1994.

[18] K. Leeuwin-Boulle and J. C. Belfiore, "Cutoff rate of time correlated fading channels," *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 612–617, 1993.

[19] R. J. McEliece and W. E. Stark, "Channels with block interference," *IEEE Transactions on Information Theory*, vol. 30, no. 1, pp. 44–53, 1984.

[20] A. O. Hero and T. L. Marzetta, "Cutoff rate and signal design for the quasi-static Rayleigh-fading space-time channel," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2400–2416, 2001.

[21] S. Misra, A. Swami, and L. Tong, "Cutoff rate of the Gauss-Markov channel with adaptive energy allocation," in *Processing of IEEE Workshop on Signal Processing Advances in Wireless Communications*, Rome, Italy, June 2003.

[22] S. Kay, *Fundamentals of Statistical Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.

[23] J. Massey, "Coding and moulation in digital communications," in *Proceedings of International Zurich Seminar on Digital Communications*, March 1974.

[24] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, John Wiley & Sons, New York, NY, USA, 1965.

[25] J. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 2001.

[26] P. H.-Y. Wu and A. Duel-Hallen, "Multiuser detectors with disjoint Kalman channel estimators for synchronous CDMA mobile radio channels," *IEEE Transactions on Communications*, vol. 48, no. 5, pp. 752–756, 2000.

[27] M. Stojanovic, J. G. Proakis, and J. A. Catipovic, "Analysis of the impact of channel estimation errors on the performance of a decision-feedback equalizer in fading multipath channels," *IEEE Transactions on Communications*, vol. 43, no. 2, pp. 877–886, 1995.

[28] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 933–946, 2000.

[29] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*, Prentice-Hall, Englewood Cliffs, NJ, USA, 2000.

[30] W. C. Jakes Jr., *Microwave Mobile Communication*, John Wiley & Sons, New York, NY, USA, 1974.

[31] T. Rappaport, *Wireless Communications: Principles & Practice*, Prentice-Hall, Upper Saddle River, NJ, USA, 1996.

[32] E. Baccarelli, "Bounds on the symmetric cutoff rate for QAM transmissions over time-correlated flat-faded channels," *IEEE Communications Letters*, vol. 2, no. 10, pp. 279–281, 1998.

Saswat Misra was born in College Park, Md, USA, in 1978. He received the B.S. in electrical engineering from the University of Maryland at College Park in 2000, and the M.S. in electrical engineering from the University of Illinois at Urbana-Champaign in 2002. Since 2002, he has been a Research Scientist at the Army Research Laboratory (ARL) in Adelphi, Md, in the Communications and Network Systems division. He has previously worked on optimal training design for wireless communication systems; an area in which he has published several papers and holds two patents (pending). Since Fall 2005, he has been a Ph.D. candidate at Cornell University. He is currently studying routing and security issues in wireless military networks.

Ananthram Swami is a Senior Research Scientist and Fellow of the Army Research Laboratory in Adelphi, Md, USA, where he works in the broad area of signal processing for communications. He received the B.S. degree from the Indian Institute of Technology, Bombay, India, the M.S. degree from Rice University, Houston, Tex, and the Ph.D. degree from the University of Southern California, Los Angeles, all in electrical engineering. He has held positions with Unocal Corporation, the University of Southern California, CS-3, and Malgudi Systems. He was a Statistical Consultant to the California Lottery, and has held visiting faculty positions at INP, Toulouse, France. He is a chair of the IEEE Signal Processing Society's Technical Committee on Signal Processing for Communications, an Associate Editor (AE) of the IEEE Transactions on Wireless Communications, and an AE of the IEEE Transactions on Signal Processing. He was coorganizer and cochair of a 1999 ASA-IMA Workshop on Heavy-Tailed Phenomena, and of a 2002 ONR/NSF/ARO/CTA Workshop on Future Challenges to Wireless Communications and Networking. He is a coguest Editor of an upcoming IEEE Signal Processing

Magazine special issue on "Distributed Signal Processing in Sensor Networks" and a EURASIP JWCN special issue on "Wireless Mobile Ad Hoc Networks."

Lang Tong is a Professor in the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA. He received the B.E. degree from Tsinghua University, Beijing, China, in 1985, and M.S. and Ph.D. degrees in electrical engineering in 1987 and 1991, respectively, from the University of Notre Dame, Notre Dame, Ind, USA. He was a Postdoctoral Research Affiliate at the Information Systems Laboratory, Stanford University, in 1991. He was also the 2001 Cor Wit Visiting Professor at Delft University of Technology. He received the Young Investigator Award from the Office of Naval Research in 1996, the Outstanding Young Author Award from the IEEE Circuits and Systems Society, the 2004 Best Paper Award (with M. Dong) from the IEEE Signal Processing Society, and the 2005 Leonard G. Abraham Prize Paper Award (with P. Venkitasubramaniam and S. Adireddy) from the IEEE Communications Society. His areas of interest include statistical signal processing, wireless communications, communication networks and sensor networks, and information theory.