# Binary Adaptive Coded Pilot Symbol Assisted Modulation Over Rayleigh Fading Channels Without Feedback

Ibrahim Abou-Fayçal, *Member, IEEE*, Muriel Médard, *Senior Member, IEEE*, and Upamanyu Madhow, *Fellow, IEEE*

***Abstract*—Pilot symbol assisted modulation (PSAM) is a standard approach for transceiver design for time-varying channels, with channel estimates obtained from pilot symbols being employed for coherent demodulation of the data symbols. In this paper, we show that PSAM schemes can be improved by adapting the coded modulation strategy at the sender to the quality of the channel measurement at the receiver, without requiring any channel feedback from the receiver. We consider performance in terms of achievable rate for binary signaling schemes. The transmitter employs interleaved codes, with data symbols coded according to their distance from the nearest pilot symbols. Symbols far away from pilot symbols encounter poorer channel measurements at the receiver and are therefore coded with lower rate codes, while symbols close to pilot symbols benefit from recent channel measurements and are coded with higher rate codes. The performance benefits from this approach are quantified in the context of binary signaling over time-varying Rayleigh fading channels described by a Gauss–Markov model. The spacing of the pilot symbols is optimized to maximize the mutual information between input and output in this setting. Causal and noncausal channel estimators of varying complexity and delay are considered. It is shown that, by appropriate optimization for the spacing between consecutive pilot symbols, the adaptive coding techniques proposed can improve achievable rate, without any feedback from the receiver to the sender. Moreover, channel estimation based on the two closest pilot symbols is generally close to optimal.**

***Index Terms*—Adaptive modulation, fading channels, Gauss–Markov, pilot symbol assisted modulation, Rayleigh fading, time-varying channels.**

## I. Introduction

In order to make the best use of wireless resources, many different types of adaptive schemes are employed. Adaptive schemes seek to modify the transmission scheme used by the sender according to the state of the channel seen by the receiver. Generally, such schemes involve feedback, concerning the state of the channel, from the receiver to the sender. In an information-theoretic context, adaptive signaling is used for Markov channels with perfect sender and receiver channel side information [53], [20] or imperfect channel side information [15], [35], [50], [26]. Power control is a commonly used type of adaptive transmission. Many practical schemes consider modifying the modulation used in order to combat fading. A common means of adapting transmission is to use different types of modulation [46], [39], for instance different levels of QAM constellations [52], [54], [41], [45], [18], [27], [55], [47], [56], [19], according to the state of the channel and possibly other considerations such as multiuser interference [17].

In this paper, we consider a different type of adaptive scheme. While the scheme still adapts the transmission to the channel seen by the receiver, the transmitter does not use feedback to determine its policy. Instead, the transmitter takes into account the time-varying quality of the channel measurement available at the receiver in order to modify its transmission policy. Thus, the transmitter adapts its signaling and coding to the quality of the channel measurement, rather than to the quality of the channel (in terms of carrier to noise ratio or other metric.) The channel measurement is obtained through regularly spaced pilot symbols. For the particular set-up that we envisage, the use of irregularly-spaced symbols would not be advantageous because we obtain a single optimized trade-off between energy and time spent on PSAM, with respect to the benefits of better channel estimation. The use of regularly spaced pilot symbols is closely related to the usual arguments that require independent and identically distributed (IID) symbols to achieve maximum mutual information, when we consider a macro-symbol to the set of transmitted symbols and its neighboring pilot symbols.

The interspersion of pilot symbols among coded data symbols is generally referred to as pilot symbol assisted modulation [12], [24], [25], [42]–[44], [56]. Pilot symbols are commonly used to improve detection [22], [31] and decoding [51]. Optimization of placement and design of pilot symbols in such schemes is an active area of research for single receivers [2], [3], [4]–[6][7], [8], [32], [33], [34], [36], [37], and multiple receivers [21].

The above schemes in general consider the channel to be perfectly known at the receiver through the use of the pilot symbols. The effect of channel variation and that of estimation error at the receiver has been considered extensively in the information-theoretic literature, see, for example, [11], [29], [28], [30], and the references therein. In the case where channel estima-

Paper approved by R. Raheli, the Editor for Detection, Equalization and Codiing of the IEEE Communications Society. Manuscript received September 25, 2002; revised October 13, 2003, and September 22, 2004. This work was supported in part by the National Science Foundation CAREER Award 009 3349 and by the HP Wireless Center. This paper was presented in part at the 38th Annual Allerton Conference on Communication, Control and Computing, October, 2000, and the 36th Asilomar Conference on Signals, Systems and Computers, November, 2002.

I. Abou-Fayçal is with the Electrical and Computer Engineering Department, American University of Beirut, Beirut 1107 2020, Lebanon, and also with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: ial4@aub.edu.lb; ibrahim.abou-faycal@aub.edu.lb).

M. Médard is with the Laboratory of Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: medard@mit.edu).

U. Madhow is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA (e-mail: madhow@ece.ucsb.edu).

Digital Object Identifier 10.1109/TCOMM.2005.849998

0090-6778/$20.00 © 2005 IEEE

tion at the receiver is provided by pilot symbols and where there is no feedback to the sender, the problem becomes, as we discuss below, that of transmitting over Ricean channels without channel side information at the receiver. The issue of capacity of Ricean channels has been considered in [40], but is in general not known. In the special case of Rayleigh distribution of the channel with Independent and Identically Distributed (IID) fading statistics among symbols, the capacity has been shown to be achieved using discrete inputs [1]. In [9] and [21], the difficulty in obtaining the capacity of Ricean channels is circumvented through a lower bound in which the product of the estimation error and the transmitted signal is treated as additive noise, in a manner akin to [30].

None of the above schemes considers dynamically adapting the codes at the transmitter to the quality of the channel measurement at the receiver. The channel models in [21] and [9] are block-fading in nature. In block-fading channels, the estimate and estimate error variance are the same for all symbols within a block. In contrast, in a continuously time-varying channel, the quality of the channel estimate at the receiver at any time sample depends crucially on the distance of that time sample from the pilot symbols and on the method of channel estimation.

Our model is the following. We consider a single sender and receiver, connected by a continuously time-varying Rayleigh fading channel. The Rayleigh fading channel is modeled as a Gauss-Markov process. The sender transmits coded and modulated data. The sender has no information regarding the state of the channel and, therefore, does not adapt its transmission scheme in response to fades. Moreover, at regular intervals, the sender transmits a constant and known pilot symbol, whose purpose is to enable measurement of the channel at the receiver. The pilot symbols have energy equal to the average energy constraint. The only channel estimate available at the receiver comes from the pilot symbols, hence there is no data-directed estimation of the channel. For each time sample $k$, the receiver computes the Bayesian least-squares estimate (BLSE) of the channel [38], the quality[1] of which depends on the received pilot symbols through its position with respect to the pilot symbols. After estimation, the channel, as seen by the receiver, is composed of a specular part (the estimate) and a zero-mean Gaussian-distributed fluctuation (the error). Thus, the channel may be viewed at the receiver as a Rician channel in which the specular part is given by the estimate and the Rayleigh (shorthand for two-dimensional Gaussian component) component is the estimation error. Thus, we consider the channel estimation error explicitly rather than assume that the pilot symbols afford perfect channel side information at the receiver. Furthermore, the transmitter employs multiple codebooks that are interleaved across pilot symbols as shown in Fig. 1. Finally, we consider binary[2] signaling, since such signaling performs well at low SNRs [16], [49], [57] and achieves capacity for low SNRs for Rayleigh channels [1]. Note that binary signals have been shown to perform well at low to moderate SNRs for

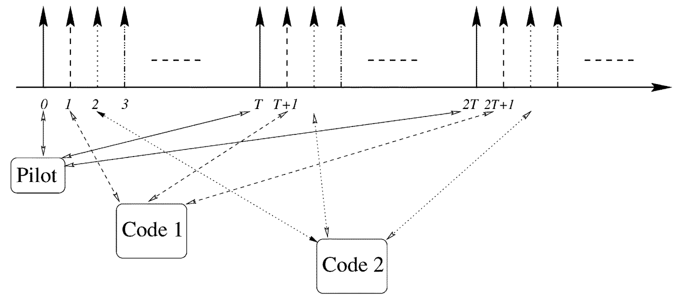


Fig. 1. Transmission scheme.

Ricean channels [13], [23] under any peak power constraint. The distributions illustrated in [13] and [23], even for high SNRs have 4 or fewer points, but are far from symmetric for 4 points. Thus, not only is binary optimal for low to moderate SNRs, but QAM does not appear to be a capacity-achieving distribution for Ricean channels.

The codes we consider are mutual information-achieving binary random codes. Thus, our optimization is done directly in terms of mutual information. The maximization of mutual information subsumes the optimization of both the modulation and the coding. We select the binary signaling scheme that optimizes mutual information for our channel condition and energy constraints.

Our purpose is to assess the benefits of performing adaptive modulation and coding to take into account the quality of the estimate obtained at the receiver through pilot tones. We consider three cases. First, we consider the case where no pilot symbols are transmitted and the channel is not estimated at the receiver. The channel at the receiver is then a Rayleigh channel. Since we do not perform any data-directed estimation, and we are inherently ignoring the correlation of the fading coefficients within a codeword, the system we consider is equivalent to one in which all the channel samples were mutually independent (by averaging over very long times). For low to moderate SNRs, binary signaling is optimal. The purpose of considering the Rayleigh channel with no pilot symbols is to establish a basis of comparison for the other two cases, which do employ pilot symbols. While the benefit of obtaining channel measurements, for instance via pilot symbols, is clear, the costs must also be considered explicitly. Indeed, the first question we pose is whether it is preferable to forego pilot symbols and channel estimation altogether and devote to coded data the time allocated to pilot symbols. The computed rates in this scenario correspond to the achievable rates of our scheme as the sounding interval increases to infinity.

Second, we consider the case where we use pilot symbols to aid in the detection and decoding at the receiver but do not modify the distribution of the transmitted signal. We term this scheme the nonadaptive scheme with pilot tones, since the scheme is adaptive at the receiver but not at the sender. The sender uses a fixed distribution, specifically the one that is optimal for transmission when the channel is block-faded. In effect, the sender behaves as though the channel estimate did not vary between pilot symbols. Since we look at mutual information, we may consider this case to be indicative of the performance of schemes where neither the distribution nor

[1]Measured in mean-squared error.

[2]More precisely, we use a discrete random input with two mass-points, i.e., we employ a two-level modulated signal with levels freely chosen to maximize the achievable rates. Throughout the paper, we will refer to this modulation scheme as "binary" for convenience.

the coding is modified according to the distance between the transmitted signal and the pilot symbols.

Third, we consider the case where we use pilot symbols as well as adaptive signaling and coding, but maintain the average per symbol power constant in the coded data. We consider the case where the estimation of the channel is causally performed either using the last pilot symbol or all the past pilots, and the case where the estimation is noncausally based on either the last transmitted pilot along with the pilot transmitted next, or all past and future pilots. At each time sample, the sender changes its transmission policy according to the distance to pilot symbols. We compare the performance of the second and third cases in order to establish the benefit of adaptive transmission. Note here that, since we are studying information rates, we are considering codes with block lengths increasing to infinity, and any estimation done in a noncausal fashion is nonrestrictive. For instance, a reasonable noncausal estimation procedure would be to use all the pilot tones within a codeword. We show that, for our adaptive signaling and coding schemes, there exists, for each channel and SNR level, a single maximal point for the achievable rate as a function of the inter-pilot symbol interval. Thus, to determine the best inter-pilot symbol interval, we need only search for a maximal point for the achievable rate as a function of the inter-pilot symbol interval.

In Section II, we present our channel model and the principles of noncausal and causal estimation. In Section III, we discuss in detail the different receiver estimation procedures that are considered. In Section IV we discuss the nonadaptive and adaptive schemes used, and in Section V, we present our numerical results. These results allow us to optimize numerically the spacing between pilot symbols for adaptive and nonadaptive schemes, since such an optimization cannot be obtained in closed form. Using the optimized spacing between pilot symbols for the different schemes, we can evaluate the benefit of adaptive schemes. Finally, conclusions and directions for future work are presented in Section VI.

## II. Channel Model

We consider the following discrete-time model for the Rayleigh fading channel $\mathsf{Y}_k = \mathsf{R}_k\mathsf{X}_k + \mathsf{N}_k$, where $k$ is the time index,[3] $\mathsf{X}_k$ is the channel input[4] at time $k$, $\mathsf{Y}_k$ the output, and $\mathsf{R}_k$ and $\mathsf{N}_k$ independent complex circular Gaussian random variables with mean zero and variance $\sigma_R^2$ and $\sigma_N^2$ the amplitude of the fading coefficient $\mathsf{R}_k$ is Rayleigh distributed and its phase is uniform. The input is average power limited: $E[|\mathsf{X}_k|^2] \leq P$.

We assume that the fading process is a first-order Gauss-Markov process

$$\mathsf{R}_k = \alpha\mathsf{R}_{k-1} + \mathsf{Z}_k \quad (1)$$

where $k$ denotes the time index and the $\{\mathsf{Z}_k\}_k$'s are IID circular Gaussian distributed with zero mean and variance equal to $(1-|\alpha|^2)\sigma_R^2$. In order to guarantee ergodicity, we rule out the case $\alpha = 1$ and consider only values of $\alpha \in [0, 1)$.

[3]Time indexes will be exclusively denoted using subscripts throughout this paper.

[4]Random variables are denoted using capital letters, while their sample values are denoted using lower-case letters.

A similar channel model was introduced in [30], where the relation between this channel and the coherence time is established: for a coherence time of $T_c$ and transmission over a bandwidth of $W, \alpha$ is determined by $\alpha^{T_c W} = \phi$ where $\phi$ is the level of decorrelation we deem necessary in our definition of coherence time. In the literature, the correlation coefficient is taken to vary from 0.9 [14] to 0.37 [10] for a time separation of $T_c$. For bandwidths in the 10 kHz range, and Doppler spreads of the order of 100 Hz, $\alpha$ will typically range between 0.9 and 0.99. For instance, for a Doppler spread of 100 Hz, $W = 10^4$ $\phi = 0.1$, we have $\alpha = 0.977$, for a Doppler spread of 200 Hz, $\alpha = 0.955$, and for a Doppler spread of 50 Hz, $\alpha = 0.989$.

When no pilot signals are used and the correlation between fading coefficients is ignored, the channel behaves as a memoryless Rayleigh fading channel. The capacity of this channel was studied in [1] and the optimizing input distribution was found to be discrete with a finite number of mass points, one of which is located at the origin. Furthermore, a binary distribution was found to be optimal at low and moderate values of the SNR, motivating the use of binary input strategies in this paper.

For an adaptively coded system, we consider the case where a pilot signal of power $P$ is transmitted once at the beginning of each interval of length $T$ (i.e., at time indexes $k = lT, l = 1, 2, \ldots$), enabling the receiver to estimate the fading coefficients governing the channel statistics. The resulting channel outputs are $\mathsf{Y}_{lT} = \mathsf{R}_{lT}\sqrt{P} + \mathsf{N}_{lT}, l \in \mathbb{N}$. In order to simplify notation, we define the following quantities. $\mathsf{R}_l$ is the column vector of length $T$ with entry $j$ given by $\mathsf{R}_{lT+T-j}, \mathsf{Z}_l$ is the column vector of length $T$ with entry $j$ given by $\mathsf{Z}_{lT+T-j}, A$ is a $T \times T$ matrix, with row $j$ being $(\cdots 0)$, matrix $G$ is a $T \times T$ matrix where row $j$ consists of $j - 10$ s, followed by a 1, followed by $\alpha, \alpha^2 \ldots$, and $C = (0, 0, \ldots, 0, \sqrt{P})$. Using these definitions, a state-space representation of the system is given by $\mathsf{R}_{l+1} = A\mathsf{R}_l + G\mathsf{Z}_{l+1}$ and $\mathsf{Y}_{lT} = C\mathsf{R}_l + \mathsf{N}_{lT}$. Based on the observation of $\{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}}$, BLSE estimates of the fading coefficients $\mathsf{R}_l$ are obtained, where $\mathcal{S}$ is the set of integers $l$ such that pilots symbols at time indexes $lT$ are used for channel estimation at time $k \in [lT, (l+1)T - 1]$.

### A. Bayesian Least-Squares Estimation (BLSE)

Let $\mathcal{S}_k$ be the set of integers $l$ such that pilots symbols at time indexes $lT$ are used for channel estimation at time $k$. Conditioned on the pilot signals $\{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k}$, note first that the conditional distribution of the output $\mathsf{Y}_k$ given the input $\mathsf{X}_k$ is Gaussian $k \neq lT$)

$$\begin{aligned}
&p_{\mathsf{Y}_k \mid \mathsf{X}_k, \{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k}}(y_k \mid x_k, \{y_{lT}\}_{l\in\mathcal{S}_k}) \\
&= \mathcal{N}(y_k; \mathsf{E}[\mathsf{Y}_k \mid \mathsf{X}_k = x_k, \{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k} = \{y_{lT}\}_{l\in\mathcal{S}_k}], \\
&\quad \mathsf{Var}(\mathsf{Y}_k \mid \mathsf{X}_k = x_k, \{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k} = \{y_{lT}\}_{l\in\mathcal{S}_k})) \\
&= \mathcal{N}\left(y_k; x_k\mathsf{E}[\mathsf{R}_k \mid \{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k} = \{y_{lT}\}_{l\in\mathcal{S}_k}], \right. \\
&\quad \left. |x_k|^2\mathsf{Var}\left(\mathsf{R}_k \mid \{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k} = \{y_{lT}\}_{l\in\mathcal{S}_k}\right) + \sigma_N^2\right) \\
&= \frac{1}{\pi\left(v_k|x_k|^2 + \sigma_N^2\right)} \exp\left\{-\frac{|y_k - \hat{\mathsf{R}}_k x_k|^2}{v_k|x_k|^2 + \sigma_N^2}\right\}, \\
&\qquad lT < k < (l+1)T \quad (2)
\end{aligned}$$

where $\hat{\mathsf{R}}_k$ is the BLSE estimate of the value of $\mathsf{R}_k$ and $v_k$ is the variance of the estimation error based on the observation of

$\{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k}$, and where the last equality is a property of BLSE in a joint Gaussian setting [38].

Given the pilot signal, the channel's BLSE and its corresponding mean-squared error constitute the equivalent Rician channel as seen by the receiver. Thus, from an information theoretic point of view, the BLSE and the variance of the mean-squared error determine the characteristics of the channel transition probability distribution of the channel, from which maximum mutual information may be determined. Stated differently, the BLSE constitutes a sufficient statistic. In conclusion, mutual information computations inherently assume maximum-likelihood decoding which involve both quantities and any other estimation strategy and decoding rule that is not the BLSE would correspond to a possibly sub-optimal decoding rule and the achievable rates will possibly be lower.

Finally note that, owing to the joint Gaussianity of the fading coefficients and the additive noise terms, the BLSE is linear, and hence, is also the linear least-squares estimate (LLSE) [38], and a LLSE procedure is optimal with respect to the sets $\{\mathcal{S}_k\}$.

Using (2), given its estimate, the receiver experiences a Ricean channel where the fading coefficient has mean $\hat{\mathsf{R}}_k$ and variance $v_k$. Therefore, under the assumption of binary signaling, for a given received sequence $\{y_{lT}\}_{l\in\mathcal{S}_k}$ the mutual information $I(\mathsf{X}_k;\mathsf{Y}_k\,|\,\{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k}=\{y_{lT}\}_{l\in\mathcal{S}_k})$ at time $k\in\{lT+1,\ldots,(l+1)T-1\}$ is given by

$$\begin{aligned}
&I(\mathsf{X}_k;\mathsf{Y}_k\,|\,\{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k}=\{y_{lT}\}_{l\in\mathcal{S}_k})\\
&\quad=p_k(1)\int p_{\mathsf{Y}_k\,|\,\mathsf{X}_k,\{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k}}(y\,|\,x_k(1),\{y_{lT}\}_{l\in\mathcal{S}_k})\\
&\qquad\times\ln\frac{p_{\mathsf{Y}_k\,|\,\mathsf{X}_k,\{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k}}(y\,|\,x_k(1),\{y_{lT}\}_{l\in\mathcal{S}_k})}{p_{\mathsf{Y}_k\,|\,\{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k}}(y\,|\,\{y_{lT}\}_{l\in\mathcal{S}_k})}dy\\
&\qquad+(1-p_k(1))\int p_{\mathsf{Y}_k\,|\,\mathsf{X}_k,\{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k}}(y\,|\,x_k(2),\{y_{lT}\}_{l\in\mathcal{S}_k})\\
&\qquad\times\ln\frac{p_{\mathsf{Y}_k\,|\,\mathsf{X}_k,\{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k}}(y\,|\,x_k(2),\{y_{lT}\}_{l\in\mathcal{S}_k})}{p_{\mathsf{Y}_k\,|\,\{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k}}(y\,|\,\{y_{lT}\}_{l\in\mathcal{S}_k})}dy
\end{aligned}$$

where the probability mass function of the binary input $\{x_k(1),x_k(2)\}$ is $\{p_k(1),(1-p_k(1))\}$, and is independent of the received sequence $\{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k}$ because there is no feedback from the receiver to the transmitter.

Furthermore, since we are eliminating the possibility of data-directed estimation, the achievable rates[5] are given by

$$\begin{aligned}
&\mathsf{E}\left[\frac{1}{T}\sum_{k=lT+1}^{(l+1)T-1}I(\mathsf{X}_k;\mathsf{Y}_k\,|\,\{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k})\right]\\
&\qquad=\frac{1}{T}\sum_{k=lT+1}^{(l+1)T-1}\mathsf{E}[I_k(\mathsf{X}_k;\mathsf{Y}_k\,|\,\{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k})] \quad (3)
\end{aligned}$$

where the expectation is over the (correlated) random variables $\{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k}$ which are circular Gaussian distributed with mean zero and variance $\sigma_R^2P+\sigma_N^2$.

## III. Estimation at the Receiver

Overall we consider two causal estimation procedures and two noncausal ones. We denote by $\mathrm{E}(p,f)$ the estimator that uses $p$ past pilots, and $f$ future pilots, counting out from the symbol of interest.

1) $\mathrm{E}(1,0)$ The channel parameters estimation is causal and based on the most recently sent pilot tone. Thus $\mathcal{S}_k=\{\lfloor k/T\rfloor\}$.
2) $\mathrm{E}(\infty,0)$ The estimation is performed using a causal Kalman filter. Here $\mathcal{S}_k=\{0,1,\ldots,\lfloor k/T\rfloor\}$.
3) $\mathrm{E}(1,1)$ The estimation is noncausal, based on the most recently transmitted pilot, as well as on the pilot transmitted next. Thus $\mathcal{S}_k=\{\lfloor k/T\rfloor,\lfloor k/T\rfloor+1\}$.
4) $\mathrm{E}(\infty,\infty)$ Finally, the channel parameters are estimated noncausally using a fixed-time Kalman filter smoother. Thus $\mathcal{S}_k=\mathbb{N}$.

### A. One Pilot Based Causal Estimation

The first method we investigate is where the receiver performs a causal estimation of the channel parameters based on the most recently sent pilot tone. For each interval, based on the observation of $y_{lT}$, estimates of the fading coefficients $\mathsf{R}_l$ are obtained by a standard Bayesian least square estimation procedure [38]. By Gaussianity, the estimates are the linear least squares estimates: $\hat{\mathsf{R}}_l=\Lambda_{\mathsf{R}_l\mathsf{Y}_{lT}}\Lambda_{\mathsf{Y}_{lT}}^{-1}y_{lT}$ and $\Lambda_e=\Lambda_{\mathsf{R}_l}-\Lambda_{\mathsf{R}_l\mathsf{Y}_{lT}}\Lambda_{\mathsf{Y}_{lT}}^{-1}\Lambda_{\mathsf{R}_l\mathsf{Y}_{lT}}^{*}$, where $\Lambda_{\mathsf{R}_l\mathsf{Y}_{lT}}$ is the cross correlation matrix of $\mathsf{R}_l$ and $\mathsf{Y}_{lT}$, and $\Lambda_{\mathsf{Y}_{lT}}$, $\Lambda_{\mathsf{R}_l}$, and $\Lambda_e$ are the covariance matrices of $\mathsf{Y}_{lT}$, $\mathsf{R}_l$ and the error respectively. We may make direct use of (1) to obtain, after manipulation:

$$\hat{\mathsf{R}}_k=\frac{\sqrt{P}\sigma_R^2}{P\sigma_R^2+\sigma_N^2}\alpha^{(k-lT)}y_{lT} \quad (4)$$

$$v_k=\sigma_R^2-\frac{P\sigma_R^4}{P\sigma_R^2+\sigma_N^2}\left|\alpha^{(k-lT)}\right|^2,\quad lT\le k<(l+1)T. \quad (5)$$

### B. Causal Kalman Filter

Using this approach, the receiver performs a maximum-likelihood estimation of the channel parameters using all the previous data available from the pilot tones. Practically speaking, the receiver is designed to run continuously (in a recursive manner) a Kalman filter on the data received from the pilots transmissions.

Recall that the state-space representation of the system is given by $\mathsf{R}_{l+1}=A\mathsf{R}_l+G\mathsf{Z}_{l+1}$ and $\mathsf{Y}_{lT}=C\mathsf{R}_l+\mathsf{N}_{lT}$. Based on the observation of $\{\mathsf{Y}_{lT}\}_{l\in\mathcal{S}_k}$, LLSE estimates of the fading coefficients $\mathsf{R}_l$ are sought after and, in the causal case, a standard Kalman filter can be used:

The update step is given by the following equations

$$\begin{aligned}
\hat{\mathsf{R}}(l\,|\,l)&=\hat{\mathsf{R}}(l\,|\,l-1)+\frac{y_{lT}-\sqrt{P}\hat{\mathsf{R}}_{lT}}{C\Phi(l\,|\,l-1)C^*+\sigma_N}\Phi(l\,|\,l-1)C^*\\
\Phi(l\,|\,l)&=\Phi(l\,|\,l-1)-\frac{1}{C\Phi(l\,|\,l-1)C^*+\sigma_N}\\
&\qquad\times\Phi(l\,|\,l-1)C^*C\Phi(l\,|\,l-1)
\end{aligned}$$

where $\hat{\mathsf{R}}(l\,|\,l)$ and $\Phi(l\,|\,l)$ represent the estimate and error covariance matrix of $\mathsf{R}$ at time $l$ given observation up to and including time $l$. Similarly, $\hat{\mathsf{R}}(l\,|\,l-1)$ and $\Phi(l\,|\,l-1)$ represent the estimate and error covariance matrix of $\mathsf{R}$ at time $l$ given observation up to and including time $l-1$. The prediction step is obtained in the following manner $\hat{\mathsf{R}}(l+1\,|\,l)=A\hat{\mathsf{R}}(l\,|\,l)$ and

[5]Note that the mutual information at times $k=lT$ is zero.

$\Phi(l+1\,|\,l) = A\Phi(l\,|\,l)A^* + GG^*\sigma_N^2$. The two sets of equations can hence be combined to yield

$$\hat{\mathsf{R}}(l\,|\,l) = A\hat{\mathsf{R}}(l-1\,|\,l-1) + \frac{y_{lT} - \sqrt{P}\alpha\mathsf{R}(l-1\,|\,l-1)_1}{C\Phi(l\,|\,l-1)C^* + \sigma_N}\Phi(l\,|\,l-1)C^*. \quad (5')$$

Furthermore, since the pair $(A, \sigma_R\sqrt{(1-|\alpha|^2)}G)$ is reachable and $(A, C/\sigma_N)$ is detectable, the Kalman filter error covariance $\Phi(l\,|\,l)$ converges, as $l$ increases to infinity, to a matrix $E_\infty$ that is the unique positive definite solution of the discrete-time algebraic Riccati equation [48]

$$E_\infty = AE_\infty A^* + \sigma_R^2(1-|\alpha|^2)GG^* - AE_\infty C^*\left[CE_\infty C^* + \sigma_N^2\right]^{-1}CE_\infty A^* \quad (6)$$

which we solve numerically using standard tools provided by *Matlab*. Using the estimation error obtained the calculation of the average mutual information can be performed in the standard manner.

### C. Two Pilots Based Non-Causal Estimation

Third, we consider the case where the estimation is noncausal and based on the most recently sent pilot tone along with the one transmitted next, i.e., $\hat{\mathsf{R}}_k$ is based on the values of $y_{lT}$ $y_{(l+1)T}$ whenever $lT < k < (l+1)T$. In this case

$$\hat{\mathsf{R}}_k = \frac{\sqrt{P}\sigma_R^2}{\left(P\sigma_R^2+\sigma_N^2\right)^2 - P^2\alpha^{2T}} \times \Big[\left[\alpha^{(k-lT)}\left(P\sigma_R^2+\sigma_N^2\right) - P\alpha^{2T-(k-lT)}\right]y_{lT} + \left[\alpha^{T-(k-lT)}\left(P\sigma_R^2+\sigma_N^2\right) - P\alpha^{T+(k-lT)}\right]y_{(l+1)T}\Big] \quad (7)$$

$$v_k = \sigma_R^2 - \frac{P\sigma_R^4}{\left(P\sigma_R^2+\sigma_N^2\right)^2 - P^2\alpha^{2T}} \times \left[\left(P\sigma_R^2+\sigma_N^2\right)\left[\alpha^{2(k-lT)} + \alpha^{2[T-(k-lT)]}\right] - 2P\alpha^{2T}\right]. \quad (8)$$

Note that (7) and (8) are extensions to (4) and (5), where one additional observation: $y_{(l+1)T}$ is considered.

### D. Non-Causal Fixed-Point Smoother

In order to quantify the highest possible gain from using an adaptive technique as described above, we consider the case where the receiver performs a maximum-likelihood estimation of the channel parameters using all of the data available from the pilot tones. In essence, the receiver is assumed to perform a smoothing operation where the channel parameters are estimated given data that extends to times beyond those at which these estimates are sought. We will further assume that data is available from an infinite time in the past as well as infinite time in the future. On a practical level, the analysis in this section corresponds to the limiting behavior of a receiver that is designed to run a finite delay fixed point Kalman smoother on the data received from the pilots transmissions. The computations can be done recursively according the well-known Kalman filter equations [48], but most importantly, when considering an infinite window, the improvement due to smoothing is given by $E_\infty - E_\infty^s = E_\infty M E_\infty$, where $E_\infty^s$ is the error covariance of the smoother, $E_\infty$ is the solution to (6), $M$ is the positive definite solution of

$$M = [A(I-K_\infty C)]^* M [A(I-K_\infty C)] + C^*\left[C^*E_\infty C + \sigma_N^2\right]^{-1}C$$

and $K_\infty$ is the steady state gain of the filter $K_\infty = E_\infty C^*[C^*E_\infty C + \sigma_N^2]^{-1}$, which is computed numerically using *Matlab*.

## IV. Nonadaptive and Adaptive Coding

First we consider the scheme where the transmitter does not adapt its transmission strategy to the statistics of the channel estimates used at the receiver. More precisely, we compute the achievable rates when the transmitter is using a single fixed input distribution at all times. For this case, we assume that the transmitter considers the channel to be block-faded, i.e., the fading coefficient is assumed constant over intervals of length $T$ and changing independently from one interval to the next. In order to evaluate the performance of such a scheme for our model, we find first the optimal input distribution for the block-faded system, and then compute the average mutual information under this distribution for different values of $\alpha$. This will allow us to quantify the performance of a system when the transmitter does not adapt its coding strategy and uses instead one fixed codebook independently of the time index.

Next we consider the scheme where, without any channel state information, the transmitter takes into consideration the statistics of the channel estimates at the receiver. It adapts accordingly its modulation and coding to maximize the rates that can be reliably transmitted over the channel. While no optimal power allocation is performed here (a constant amount of power is used instead), at each time step the transmitter uses a "good" codebook achieving the highest mutual information of the Ricean channel the receiver sees.

Equivalently, one can think of the problem as that of finding the best input strategy that maximizes the expected mutual information $\mathsf{E}[I_k(\mathsf{X}_k;\mathsf{Y}_k\,|\,\{\mathsf{Y}_{lT}\})]$ for each time step between $lT+1$ and $(l+1)T-1$. For these computations, we considered the estimation methods described in Section II, i.e., for the pair $(\{\hat{\mathsf{R}}_k\},\{v_k\})$ given by either set of (4) and (5), or of (5′) and (6), or of (7) and (8).

Since no closed form expression can be obtained for the optimal input distribution, we use standard *Matlab* tools to optimize, for each time period $k \in \{1,\ldots,T-1\}$, the expected mutual information over the input probability distribution. As mentioned previously, the input alphabet is restricted to consist of only two points. The corresponding optimal distribution yields of course the highest achievable rates depending on how far the transmission is occurring with respect to the pilot signals. Sending pilot tones frequently clearly reduces the rates as a significant portion of the time and power is used to estimate the channel and no information is conveyed from the transmitter to the receiver. On the other hand, when the pilots are used very infrequently, the channel estimates at the receiver are of poor quality and the information rates are low.

It is worth mentioning that the numerical results confirm what one expects regarding the optimal input distribution. Namely,

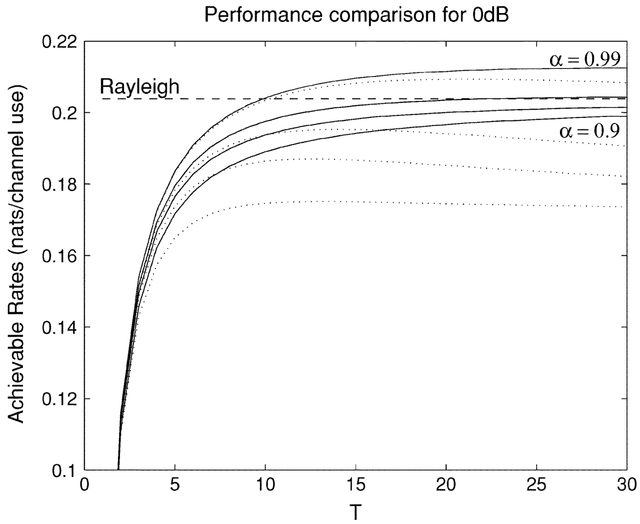


Fig. 2. Best achievable rates at 0 dB function of $T$ when $\alpha = 0.9, 0.95, 0.97$ and $0.99$ for adaptive *(full lines)* and nonadaptive *(dotted lines)* coding using $\mathrm{E}(1,0)$. The dashed line represents the capacity of the IID Rayleigh fading channel.

the solution lies between the extremes of on-off keying (optimal for the IID Rayleigh fading case [1]) and antipodal signaling (optimal for a perfectly known channel). Indeed, the optimal input distribution consists of two nonzero masses, the first of which is negative located between $-\sqrt{P}$ and zero, and the second, positive greater than $\sqrt{P}$.

One of the questions that we will try to answer in the following section is: what is the optimal value of $T$ that would yield the best compromise?

## V. Numerical Results

Figs. 2–4 show, for SNRs of 0, 3, and 5 dB, the results for applying $\mathrm{E}(1,0)$, the causal estimation based on one pilot tone. The line with long dashes corresponds to the case where we send no pilot tones and hence have no estimate of the channel at the receiver. The dashed curves correspond to the case where we do not vary the input distribution, but rather use a single distribution for all symbols other than pilot symbols. The transmitter's signal is designed as though the channel were constant in between pilots. This corresponds to the common block-fading model. The full curves consider signaling and codes that adapt to the location from a coded symbol to the last pilot symbol. Thus, symbols closer to the last pilot symbol are coded with higher rate codes, while codes for symbols far from the last pilot symbol are coded with lower rate codes. This scheme is readily implemented by using interleaved codes where the interleaving period is the period of the sounding interval as illustrated in Fig. 1. The curves are parameterized by values of $\alpha$, for $\alpha = 0.9, 0.95, 0.97, 0.99$, with higher curves corresponding to higher values of $\alpha$. The optimal value for a particular scheme and SNR occurs at the maximum of the curve. Note that, for low values of $\alpha$, i.e., for rapidly varying channels, or for schemes that do not adapt the signaling and codes to the proximity of a pilot symbol, sounding through pilot symbols may not be beneficial. As is apparent from Figs. 2–4, in the considered ranges of the SNR, there is no benefit gained from using pilot symbol

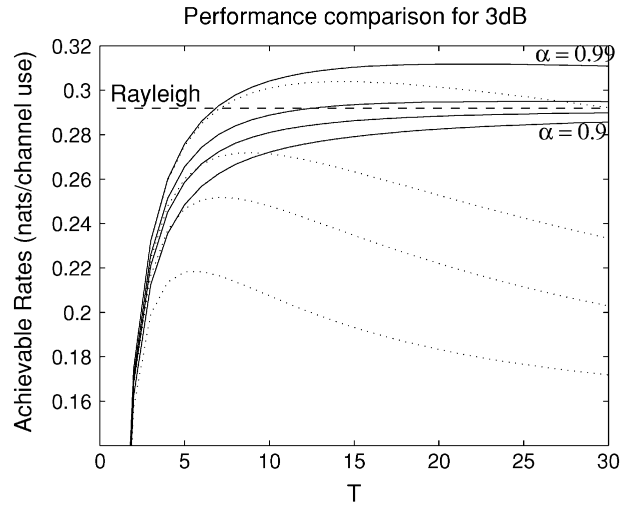


Fig. 3. Best achievable rates at 3 dB function of $T$ when $\alpha = 0.9, 0.95, 0.97$ and $0.99$ for adaptive *(full lines)* and nonadaptive *(dotted lines)* coding using $\mathrm{E}(1,0)$. The dashed line represents the capacity of the IID Rayleigh fading channel. The circles indicate the values of $T$ that maximize achievable rate.

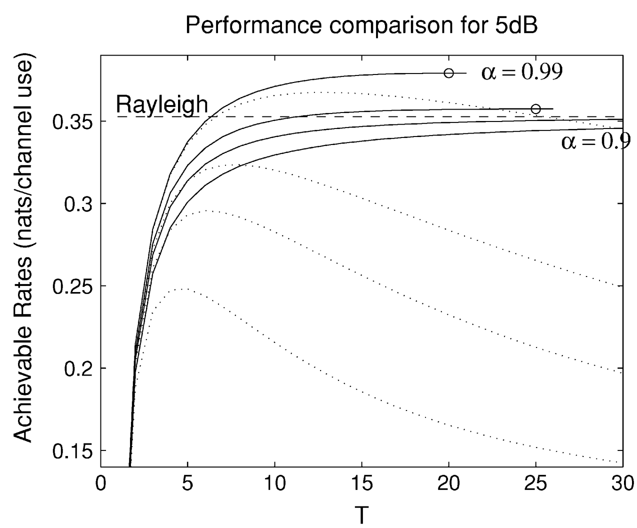


Fig. 4. Best achievable rates at 5 dB function of $T$ when $\alpha = 0.9, 0.95, 0.97$ and $0.99$ for adaptive *(full lines)* and nonadaptive *(dotted lines)* coding using $\mathrm{E}(1,0)$. The dashed line represents the capacity of the IID Rayleigh fading channel. The circles indicate the values of $T$ that maximize achievable rate.

assisted modulation for $\alpha = 0.9, 0.95$. Indeed, for these low correlation models, not sending a pilot tone and coding instead to a memoryless Rayleigh fading channel outperforms the adaptive and nonadaptive causal scheme we have described. However, for $\alpha = 0.97, 0.99$, the figures show not only that an improvement is possible, but also that there is a trade-off between infrequent transmission of pilots and the channel estimate quality.

As shown in the Appendix, there for every SNR is a single maximal point for our feasible rate with respect to $T$. This holds for both the adaptive and nonadaptive schemes. Thus, once an optimal value of $T$ is obtained, we our search can stop. The optimal value for $T$ is apparent in the figures.

Finally, note from Fig. 3 and 4 that the adaptive coding technique proposed can yield up to 42% improvement in achievable rate over the nonadaptive one, without any feedback from the receiver to the sender. For a given value of $\alpha$, the larger the SNR

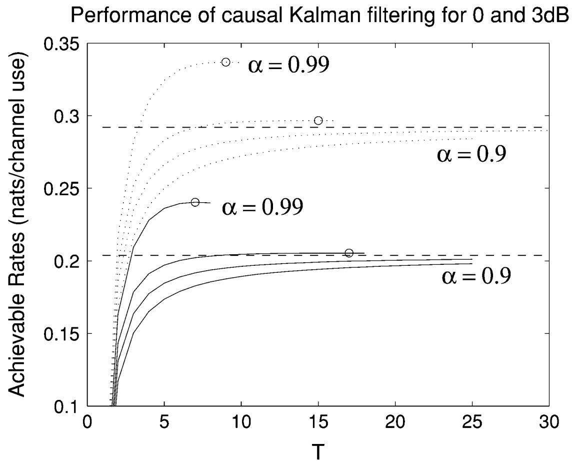


Fig. 5. Best achievable rates at 0 dB *(full lines)* and 3 dB *(dotted lines)* function of $T$ when $\alpha = 0.9, 0.95, 0.97$ and $0.99$ for adaptive coding using $\mathrm{E}(\infty, 0)$. The dashed lines represents the capacity of the IID Rayleigh fading channel for 0 dB (lower line) and 3 dB (upper line). The circles indicate the values of $T$ that maximize achievable rates.

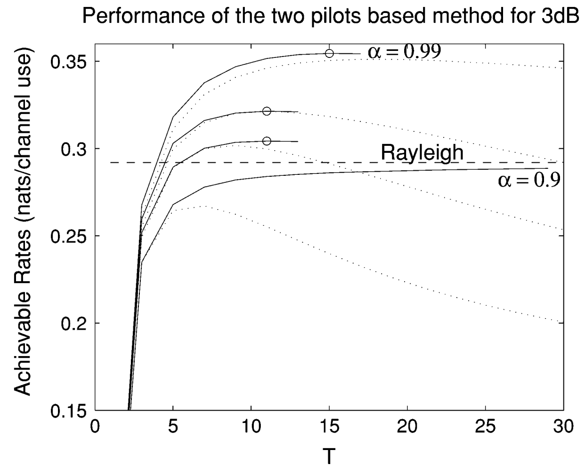


Fig. 7. Best achievable rates at 3 dB function of $T$ when $\alpha = 0.9, 0.95, 0.97$ and $0.99$ for adaptive *(full lines)* and nonadaptive *(dotted lines)* coding using $\mathrm{E}(1, 1)$. The dashed line represents the capacity of the IID Rayleigh fading channel. The circles indicate the values of $T$ that maximize achievable rates.

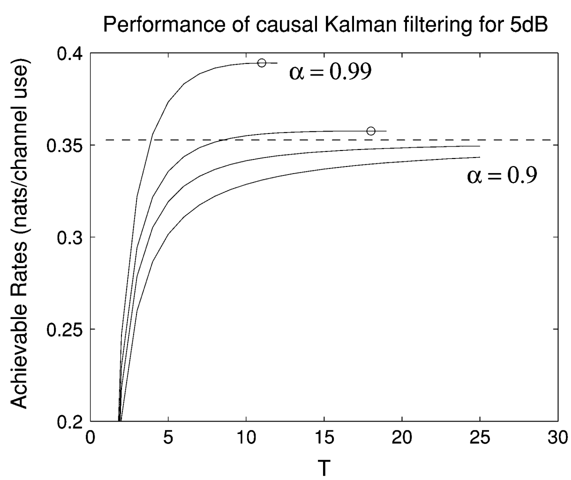


Fig. 6. Best achievable rates at 5 dB function of $T$ when $\alpha = 0.9, 0.95, 0.97$ and $0.99$ for adaptive coding using $\mathrm{E}(\infty, 0)$. The dashed line represents the capacity of the IID Rayleigh fading channel. The circles indicate the values of $T$ that maximize achievable rates.

the larger the gain, and similarly, for a given SNR, the lower the $\alpha$ the larger the gain.

Similar results hold for $\mathrm{E}(\infty, 0)$, in which all past pilot symbols are used to estimate the channel at the receiver. Figs. 5 and 6 show the performance of $\mathrm{E}(\infty, 0)$ schemes for 0, 3, and 5 dB. While there is an advantage to considering all past pilot symbols, $\mathrm{E}(\infty, 0)$ does not provide significant improvement over $\mathrm{E}(1, 0)$. The improvement is limited by the fact that pilot symbols before the last pilot symbol are fairly weakly correlated with the current channel realization. For a high correlation among channel realizations during different pilot tone transmissions, we would require a very low $T$, which would be inefficient. The optimal value of $T$, however, is reduced with respect to $\mathrm{E}(1, 0)$. The benefit of sounding more frequently springs from the fact that pilot symbols provide more benefit than for $\mathrm{E}(1, 0)$, since pilot symbols are used in many channel estimates rather than in a single one.

The above discussion for $\mathrm{E}(\infty, 0)$ motivates considering $\mathrm{E}(1, 1)$, in which the two closest pilot symbols are used for estimation. Note that using a noncausal estimator is not detrimental to delay from an implementation point of view, since we are considering interleaved codes, which themselves require decoding over several sounding intervals. Thus, the extra delay required by having to consider a noncausal implementation will in general be negligible in comparison to the delay due to interleaving. For the sake of brevity, we do not show results for all SNRs and values of $\alpha$. Fig. 7 shows the results for an SNR of 3 dB. Comparing for $\alpha = 0.95$ our results with those for the causal methods, $\mathrm{E}(1, 0)$ and $\mathrm{E}(\infty, 0)$, we see that even for nonadaptive signaling and coding, sounding is preferable over no pilot symbols. Note also that the difference between the performance of the adaptive and nonadaptive schemes is lower for $\mathrm{E}(1, 1)$ than for the causal methods, particularly when $\alpha = 0.99$. The reason for this reduced difference in performance is the following. In the case of two-pilot noncausal estimation, the variance of the estimation error does not fluctuate throughout the interval between consecutive pilot symbols as much as for causal methods. For causal methods, the channel estimate is poor for symbols far from the last sounding symbol. Indeed, the worst estimation error variance is for the coded symbol immediately preceding the next pilot symbol. In the noncausal case, the proximity to the next pilot symbol is used to provide a good estimate.

Note also that the optimal values of $T$ appear to be smaller than for $\mathrm{E}(1, 0)$. Indeed, the gain from estimating using $\mathrm{E}(1, 1)$ is considerably greater than that with $\mathrm{E}(1, 0)$, making the more frequent transmission of pilots worthwhile.

Figs. 8 and 9 show the performance of $\mathrm{E}(\infty, \infty)$ for SNRs of 0 and 3 dB, and 5 dB, respectively. As expected, the method performs better than any of the previous methods. The behavior is generally comparable to that of $\mathrm{E}(1, 1)$. A joint comparison of all methods for adaptive coding with $\alpha = 0.99$ is shown in Figs. 10 and 11, for SNRs of 0 and 3 dB, and 5 dB respectively. For SNRs above 0 dB, $\mathrm{E}(1, 1)$ outperforms $\mathrm{E}(\infty, 0)$. Moreover,

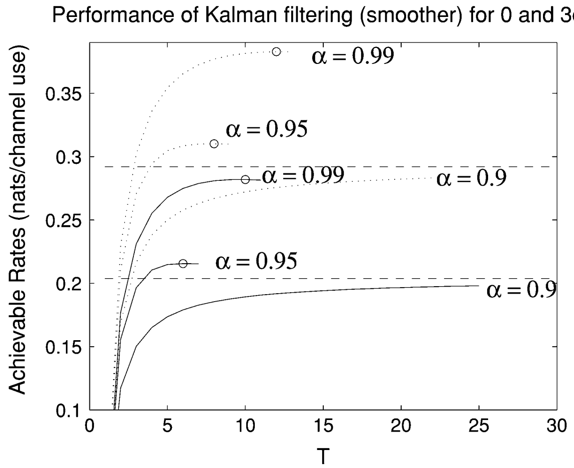


Fig. 8. Best achievable rates at 0 dB *(full lines)* and 3 dB *(dotted lines)* function of $T$ when $\alpha = 0.9, 0.95$ and $0.99$ for adaptive coding using $\mathrm{E}(\infty, \infty)$. The dashed lines represents the capacity of the IID Rayleigh fading channel for 0 dB (lower line) and 3 dB (upper line). The circles indicate the values of $T$ that achievable rates.

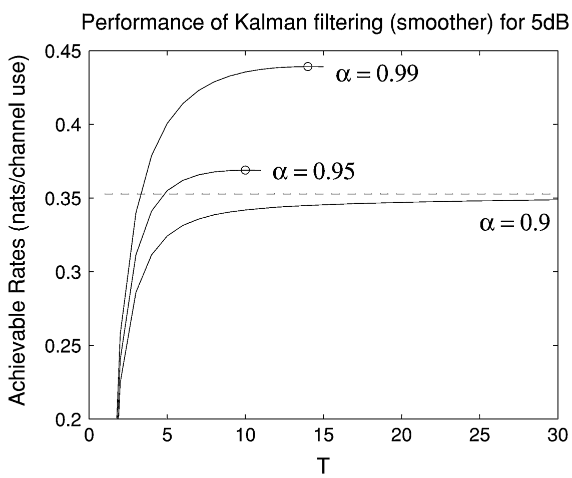


Fig. 9. Best achievable rates at 5 dB function of $T$ when $\alpha = 0.9, 0.95$ and $0.99$ for adaptive coding using $\mathrm{E}(\infty, \infty)$. The dashed line represents the capacity of the IID Rayleigh fading channel. The circles indicate the values of $T$ that maximize achievable rates.

the difference in performance among the methods tends to be reduced as the SNR increases. While $\mathrm{E}(\infty, \infty)$ requires in theory an infinite number of past and future symbols, in practice it can be implemented with a finite window. The performance of any finite window noncausal scheme would fall between that of $\mathrm{E}(1, 1)$ and $\mathrm{E}(\infty, \infty)$. The bulk of the benefit is derived from nearby pilot symbols and our experiments indicate rapid convergence to the infinite case with a finite set of pilot symbols in the past and future.

## VI. Conclusions and Future Work

We have investigated the use of adaptive and nonadaptive sender strategies for time-varying channels with pilot symbol assisted modulation and no feedback from the receiver to the sender. We have shown that, depending on the rate of change of the channel, adaptive sender strategies, properly optimized for spacing between consecutive pilot tones, can improve achievable rates. This improvement comes at no cost in term of computation at the sender, since the codes are pre-computed. The benefit is particularly marked when we use more than a single pilot symbol to perform channel estimation at the receiver.

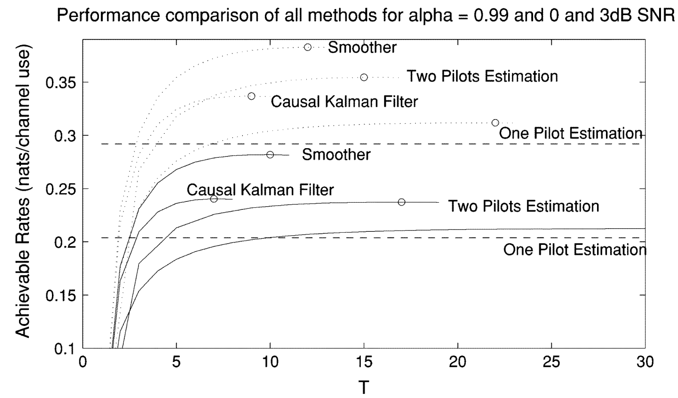


Fig. 10. Comparison of adaptive coding for all Methods for $\alpha = 0.99$ at SNRs of 0 dB *(full lines)* and 3 dB *(dotted lines)*. The dashed lines represents the capacity of the IID Rayleigh fading channel for 0 dB (lower line) and 3 dB (upper line). The circles indicate the values of $T$ that maximize achievable rates.

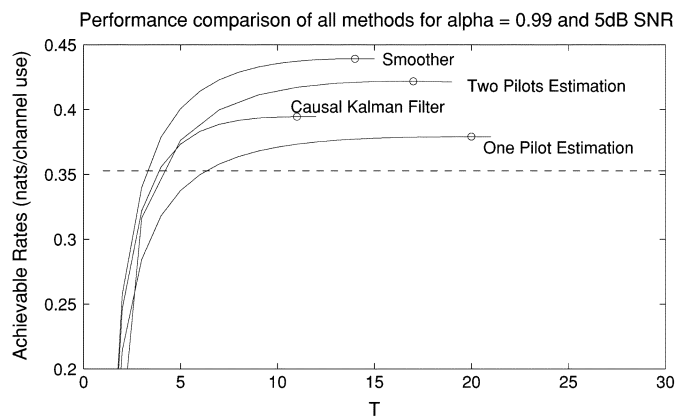


Fig. 11. Comparison of adaptive coding for all Methods for $\alpha = 0.99$ at an SNR of 5 dB. The dashed line represents the capacity of the IID Rayleigh fading channel. The circles indicate the values of $T$ that maximize achievable rates.

Several further research topics naturally arise. The most immediate one is the consideration of the number and location of pilot symbols used for estimation. For instance, we may consider estimation based on the closest $n$ pilot symbols rather than all past and present symbols, used in the smoother. $\mathrm{E}(1, 1)$ is the special case where $n = 2$. Another possible estimation technique is a fixed lag Kalman filter. The recursive nature of channel estimation for our model means that considering all past pilot symbols is not computationally onerous, because estimates are continually updated. Considering future pilot symbols, however, may have implications in terms of delay when implementation issues are taken into account.

Another natural direction for further research is to investigate to what extent adaptive power modulation can further improve performance. One way of doing this is to consider varying the energy of the pilot symbols. Another approach is to vary the energy of the modulated signal according to its distance to the closest pilot. Establishing an optimal power allocation is difficult, however, because we do not have a closed form for capacity. Indeed, one may show that heuristic methods akin to waterfilling, which attempt to maintain the interference constant, do not generally lead to improvements over constant power allocations.

Finally, while we have considered no feedback and hence no adaptation at the sender to the state of the channel, our scheme can be applied when there is channel side information at the sender. For instance, power control when there is adaptive coding of the type described in this paper may yield more power allocation to better SNR realization than when there is no adaptation to the channel estimation error variance at the receiver. The benefits of the latter adaptation we have shown to increase with SNR [58]. Therefore, the benefit of high SNR over low SNR would, in turn, increase through the use of coding adaptation to channel estimation error variance.

## APPENDIX

In this Appendix, we establish the fact that the achievable rates that we compute have a unique maximum in $T$.

$\mathrm{E}(1,0)$*:* In the $\mathrm{E}(1,0)$ case, for a period of $T$ of pilot signaling, the achievable rates may be represented as $(1/T)\sum_{k=1}^{T-1} f(k)$, where $f(k) = \mathsf{E}[I_k(\mathsf{X}_k;\mathsf{Y}_k\,|\,\mathsf{Y}_0)]$. First, note that $f(k)$ is decreasing for positive values of $k$. If the $T_m$ represents the value of $T$ for which the first maximum is reached, i.e., for $T < T_m f(k) \le (1/T_m)\sum_{k=1}^{T_m-1} f(k) = R_m$, and $(1/(T_m+1))\sum_{k=1}^{T_m} f(k) \le (1/T_m)\sum_{k=1}^{T_m-1} f(k) = R_m$ which implies that

$$\begin{aligned}\frac{1}{T_m+1}\sum_{k=1}^{T_m} f(k) &= \frac{1}{T_m-1}f(k) + \frac{1}{T_m+1}f(T_m)\\ &= \frac{T_m}{T_m+1}R_m + \frac{1}{T_m+1}f(T_m)\\ &\le \frac{1}{T_m}\sum_{k=1}^{T_m-1} f(k) = R_m,\end{aligned}$$

and hence, $f(T_m) \le R_m$. Now for any $T > T_m$

$$\begin{aligned}\frac{1}{T}\sum_{k=1}^{T-1} f(k) &= \frac{1}{T_m}f(k) + \frac{1}{T}\sum_{k=T_m+1}^{T-1} f(k)\\ &= \frac{T_m+1}{T}\frac{1}{T_m+1}\sum_{k=1}^{T_m} f(k) + \frac{1}{T}\sum_{k=T_m+1}^{T-1} f(k)\\ &\le \frac{T_m+1}{T}R_m + \frac{1}{T}\sum_{k=T_m+1}^{T-1} f(k)\\ &\le \frac{T_m+1}{T}R_m + \frac{T-(T_m+1)}{T}f(T_m+1),\end{aligned}$$

where we have used the fact that $f(k)$ is decreasing for positive $k$. Furthermore, since $f(T_m+1) \le f(T_m) \le R_m$

$$\frac{1}{T}\sum_{k=1}^{T-1} f(k) = \frac{T-(T_m+1)}{T}R_m = R_m.$$

$E(\infty,0)$*:* For the case of $E(\infty,0)$, the argument is similar and is as follows: for a period of $T$ of pilot signaling, the achievable rates may be represented as $(1/T)\sum_{k=1}^{T-1} f(k,T)$, where $f(k,T) = \mathsf{E}[I_k(\mathsf{X}_k;\mathsf{Y}_k\,|\,\{\mathsf{Y}_{lT}\})]$. Similarly to before, for a fixed value of $T, f(k,T)$ is a decreasing function for positive values of $k$. Furthermore, for a fixed $k, f(k,T)$ is a decreasing function of $T$ for values strictly larger than $k$. Let $T_m$ represent the value of $T$ for which the first maximum is reached, i.e., for $T < T_m \sum_{k=1}^{T-1} f(k,T) \le (1/T_m)\sum_{k=1}^{T_m-1} f(k,T_m) = R_m$, and $(1/(T_m+1))\sum_{k=1}^{T_m} f(k,(T_m+1)) \le (1/T_m)\sum_{k=1}^{T_m-1} f(k,T_m) = R_m$, which implies that

$$\begin{aligned}&\frac{1}{T_m+1}\sum_{k=1}^{T_m} f(k,(T_m+1))\\ &\quad= \sum_{k=1}^{T_m-1} f(k,(T_m+1)) + \frac{1}{T_m+1}f(T_m,(T_m+1))\\ &\quad\le \frac{T_m}{T_m+1}\frac{1}{T_m}\sum_{k=1}^{T_m-1} f(k,T_m) + \frac{1}{T_m+1}f(T_m,(T_m+1))\\ &\quad= \frac{T_m}{T_m+1}R_m + \frac{1}{T_m+1}f(T_m,(T_m+1))\\ &\quad\le \frac{1}{T_m}\sum_{k=1}^{T_m-1} f(k,T_m) = R_m\end{aligned}$$

and hence, $f(T_m,(T_m+1)) \le R_m$.

Now for any $T > T_m$

$$\begin{aligned}&\frac{1}{T}\sum_{k=1}^{T-1} f(k,T)\\ &\quad= \frac{1}{T_m}f(k,T) + \frac{1}{T}\sum_{k=T_m+1}^{T-1} f(k,T)\\ &\quad= \frac{T_m+1}{T}\frac{1}{T_m+1}\sum_{k=1}^{T_m} f(k,T) + \frac{1}{T}\sum_{k=T_m+1}^{T-1} f(k,T)\\ &\quad\le \frac{T_m+1}{T}\frac{1}{T_m+1}\sum_{k=1}^{T_m} f(k,(T_m+1))\\ &\qquad+ \frac{1}{T}\sum_{k=T_m+1}^{T-1} f(k,(T_m+1))\\ &\quad\le \frac{T_m+1}{T}R_m + \frac{1}{T}\sum_{k=T_m+1}^{T-1} f(k,(T_m+1))\\ &\quad\le \frac{T_m+1}{T}R_m + \frac{T-(T_m+1)}{T}f(T_m+1,(T_m+1)).\end{aligned}$$

Furthermore, since $f(T_m+1,(T_m+1)) \le f(T_m,(T_m+1)) \le R_m$

$$\frac{1}{T}\sum_{k=1}^{T-1} f(k,T) = \frac{T-(T_m+1)}{T}R_m = R_m.$$

## REFERENCES


[1] I. C. Abou-Fayçal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.

[2] S. Adireddy and L. Tong, "Detection with embedded known symbols: Optimal symbol placement and equalization," in *Proc. 2000 Int. Conf. Acoustics Speech and Signal Processing*, vol. V, Jun. 2000, pp. 2541–2543.

[3] ——, "Optimal embedding of known symbols," in *Proc. Conf. Information Sciences and Systems*, Mar. 2000.

[4] ——, "Optimal placement of known symbols for nonergodic broadcast channels," *IEEE Trans. Inf. Theory*, submitted for publication.
[5] ——, "Optimal placement of training for unknown channels," *IEEE Trans. Inf. Theory*, submitted for publication.
[6] ——, "Repeatable quasi-periodic placement schemes," presented at the Int. Symp. Information Theory, Jun. 2001.
[7] S. Adireddy, L. Tong, and H. Viswananthan, "Optimal embedding of known symbols for OFDM," in *Proc. 2000 Int. Conf. Acoustics Speech and Signal Processing*, Salt Lake City, UT, May 2001.
[8] ——, "Optimal placement of training for unknown channels," in *Proc. Conf. Information Sciences and Systems*, May 2001.
[9] ——, "Optimal placement of training for frequency-selective block-fading channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2338–2353, Aug. 2002.
[10] P. A. Bello and B. D. Nelin, "The effect of frequency selective fading on the binary error probabilities of incoherent and differentially coherent matched filter receivers," *IEEE Trans. Commun. Syst.*, vol. CS-11, no. 2, pp. 170–186, Jun. 1963.
[11] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
[12] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 11, pp. 686–693, Nov. 1991.
[13] T. H. Chan and F. R. Kschischang, "Capacity-achieving probability measure of an input-bounded optical channel with signal-dependent noise," presented at the 41st Annu. Allerton Conf. Communication, Control and Computing, Monticello, IL, Oct. 2003.
[14] D. C. Cox and R. P. Leck, "Correlation bandwidth and delay spread multipath propagation statistics for 910 MHz urban mobile radio," *IEEE Trans. Commun.*, vol. COM-23, no. 11, pp. 1271–1280, Nov. 1975.
[15] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, Sep. 1999.
[16] B. Forney and R. Koetter, *Codes, Graphs, and Systems*. Boston, MA: Kluwer, 2002.
[17] A. J. Goldsmith and M.-S. Alouini, "Area spectral efficiency of cellular mobile radio systems," *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp. 1047–1066, Jul. 1999.
[18] A. Goldsmith and S. G. Chua, "Variable rate variable power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.
[19] D. L. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 844–855, Oct. 1999.
[20] A. J. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.
[21] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
[22] P. Ho and J. H. Kim, "Pilot symbol-assisted detection of CPM schemes operating in fast fading channels," *IEEE Trans. Commun.*, vol. 44, no. 3, pp. 337–347, Mar. 1996.
[23] J. Huang and S. P. Meyn, "Characterization and computation of optimal distributions for channel coding," presented at the 37th Annu. Conf. Information Sciences and Systems, Mar. 2003.
[24] M. M. Khairy and E. Geraniotis, "BER evaluation of symbol-aided coherent demodulation for Rician and Rayleigh fading channels," in *Proc. IEEE Symp. Computers and Communications*, vol. 1, 1999, pp. 105–109.
[25] Y. S. Kim, C. J. Kim, G. Y. Jeong, Y. J. Bang, H. K. Park, and S. S. Choi, "New Rayleigh fading channel estimator based on PSAM channel sounding technique," in *Proc. IEEE ICC*, Jun. 1997, pp. 1518–1520.
[26] T. E. Klein, "Capacity of Gaussian Noise Channels with Side Information and Feedback," Ph.D. thesis, MIT, Feb. 2001.
[27] C. H. Lim and J. K. Jeong, "Adaptive modulation using the multipath fading compensation," *IEEE Trans. Veh. Technol.*, vol. 34, pp. 940–942, 1998.
[28] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.
[29] A. Lapidoth and S. Shamai, "Fading channels: How perfect need "perfect side information" be," in *Proc. IEEE Information Theory Workshop*, Jun. 1999, pp. 36–39.
[30] M. Médard, "The effect upon channel capacity in wireless communicationsof perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 935–946, Mar. 2000.
[31] J. P. McGeehan and A. J. Bateman, "Phase locked transparent tone-in-band (TTIB): A new spectrum configuration particularly suited to the transmission of data over SSB mobile radio network," *IEEE Trans. Commun.*, vol. COM-32, no. 1, pp. 81–87, Jan. 1984.
[32] X. Ma, G. B. Giannakis, and S. Barbarossa, "Non-data-aided frequency-offset and channel estimation in OFDM and related block transmissions," in *IEEE Int. Conf. Communications*, Jun. 2001, pp. 1866–1870.
[33] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly-selective fading channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351–1366, May 2003.
[34] ——, "Optimal training for block transmissions over doubly selective wireless fading channels," in *Proc. Int. Conf. ASSP*, May 2002.
[35] M. Médard and R. Srikant, "Capacity of nearly-decomposable Markovian fading channels under asymmetric receiver-sender side information," in *Proc. Int. Symp. Information Theory*, 2000, p. 413.
[36] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Cons. Electron.*, vol. 44, no. 3, pp. 1122–1128, Aug. 1998.
[37] S. Ohno and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2138–2145, Sep. 2004.
[38] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 4th ed. New York: McGraw-Hill, 2002.
[39] X. Qiu and K. Chawla, "On the performance of adaptive modulation in cellular systems," *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp. 1237–1251, Jul. 1999.
[40] J. S. Richters, "Communication over Fading Dispersive Channels,", MIT RLE Tech. Rep. 464, Nov. 1964.
[41] S. Sampei, S. Komaki, and N. Morinaga, "Adaptive modulation TDMA scheme for large capapcity personal multimedia communications systems," *IEICE Trans. Commun.*, vol. E77B, pp. 1096–1103, Sep. 1994.
[42] S. Sampei and T. Sunaga, "Rayleigh fading compensation for QAM in land mobile radio communications," *IEEE Trans. Veh. Technol.*, vol. 42, no. 5, pp. 137–147, May 1993.
[43] X. Tang, M.-S. Alouini, and A. J. Goldsmith, "Effect of channel estimation error on M-QAM BER performance in Rayleigh fading," *IEEE Trans. Commun.*, vol. 47, no. 12, pp. 1856–1864, Dec. 1999.
[44] J. M. Torrance and L. Hanzo, "Comparative study of pilot symbol assisted modem schemes," in *Proc. Radio Receivers and Associated Systems Conf.*, Sep. 1995, pp. 36–41.
[45] ——, "On the upper bound peformance of adaptive QAM in a slow Rayleigh fading," *Electron. Lett.*, vol. 32, pp. 1218–1230, Oct. 1997.
[46] ——, "Adaptive modems over slow indoor Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp. 1237–1251, Jul. 1999.
[47] J. M. Torrance, L. Hanzo, and T. Keller, "Interference aspects of adaptive modems over slow Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 48, no. 5, pp. 1527–1535, Sep. 1999.
[48] B. Hassibi, T. Kailath, and A. H. Sayed, *Linear Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 2000.
[49] S. Verdu, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.
[50] H. Viswanathan, "Capacity of time-varying channels with delayed feedback," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 761–771, Mar. 1999.
[51] L. Welburn and J. K. Cavers, "Pilot signals improve the performance of Reed–Solomon errors and erasures decoder in Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 5, pp. 1527–1535, May 1999.
[52] C. H. Wong and L. Hanzo, "Upper-bound performance of a wide-band adaptive modem," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 367–369, Mar. 2000.
[53] J. Wolfowitz, *Coding Theorems of Information Theory*. New York: Springer-Verlag, 1978.
[54] W. T. Webb and R. Steele, "Variable rate QAM for mobile radio," *IEEE Trans. Commun.*, vol. 43, no. 7, pp. 2223–2230, Jul. 1995.
[55] C. H. Wong, B. L. Yeap, and L. Hanzo, "Wideband burst-by-burst adaptive modulation with turbo equalization and iterative channel estimation," in *Proc. Vehicular Technology Conf.*, vol. 3, Spring 2000, pp. 2044–2048.
[56] Y. Xu, H.-J. Su, and E. Geraniotis, "Pilot symbol assisted QAM with iterative filtering and turbo decoding over Rayleigh flat-fading channels," in *Military Communications Conf. Proc.*, vol. 1, 1999, pp. 86–91.
[57] R. Gallager, "Power limited channels: Coding, multiaccess, and spread spectrum," MIT LIDS, Cambridge, MA, Nov. 1987.
[58] A. Bdeir, I. Abou-Fayçal, and M. Médard, "Power allocation schemes for pilot symbol assisted modulation over Rayleigh channels with no feedback," in *Communications Theory Symp. of the IEEE Int. Conf. Communications (ICC 2004)*, vol. 2, Jun. 2004, pp. 224–227.

**Ibrahim C. Abou-Fayçal** (S'96–M'00) received the Ingenieur des Arts et Manufactures diploma from Ecole Centrale Paris, Paris, France, and the S.M. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1996 and 2001, respectively.

From 2001 to 2003, he was with the Department of Electrical Engineering and Computer Science at MIT as a Postdoctoral Lecturer, and since September 2003, he has been with the Electrical and Computer Engineering Department at the American University of Beirut, Beirut, Lebanon. His research interests are in information theory, digital communications, stochastic systems and optimization.

**Muriel Médard** (S'91–M'95–SM'00) received the B.S. degrees in electrical engineering and computer science and in mathematics in 1989, a B.S. degree in humanities in 1990, a M.S. degree in electrical engineering in 1991, and a Sc.D. degree in electrical engineering in 1995, all from the Massachusetts Institute of Technology (MIT), Cambridge.

She is currently a Harold E. and Esther Edgerton Associate Professor in the Electrical Engineering and Computer Science Department at MIT and a member of the Laboratory for Information and Decision Systems (LIDS). She was previously an Assistant Professor in the Electrical and Computer Engineering Department and a member of the Coordinated Science Laboratory at the University of Illinois, Urbana-Champaign. From 1995 to 1998, she was a Staff Member at the MIT Lincoln Laboratory in the Optical Communications and the Advanced Networking Groups. Her research interests are in the areas of reliable communications, particularly for optical and wireless networks.

Dr. Médard serves as an Associate Editor for the Optical Communications and Networking Series of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and as an Associate Editor in Communications for the IEEE TRANSACTIONS ON INFORMATION THEORY. She has served as a Guest Editor for the IEEE JOURNAL OF LIGHTWAVE TECHNOLOGY. She received a National Science Foundation CAREER award in 2000. She was awarded the IEEE Leon K. Kirchmayer Prize Paper Award in 2002. She received the 2004 MIT Edgerton Junior Faculty Award. Her website is http://web.mit.edu/medard/www/medard.html.

**Upamanyu Madhow** (M'90–SM'96–F'05) received the Bachelor's degree in electrical engineering from the Indian Institute of Technology, Kanpur, in 1985 and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois, Urbana-Champaign, in 1987 and 1990, respectively.

From 1990 to 1991, he was a Visiting Assistant Professor at the University of Illinois. From 1991 to 1994, he was a Research Scientist with Bell Communications Research, Morristown, NJ. From 1994 to 1999, he was with the Department of Electrical and Computer Engineering, University of Illinois, first as an Assistant Professor, and since 1998, as an Associate Professor. Since December 1999, he has been an Associate Professor with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, where he is currently a Professor. His research interests are in communication systems and networking, with current emphasis on wireless communication.

Dr. Madhow is a recipient of the National Science Foundation CAREER Award. He has served as an Associate Editor for Spread Spectrum for the IEEE TRANSACTIONS ON COMMUNICATIONS, and as an Associate Editor for Detection and Estimation for the IEEE TRANSACTIONS ON INFORMATION THEORY. His website is http://www.ece.ucsb.edu/Faculty/Madhow.