Adaptive PSAM Accounting for Channel Estimation and Prediction Errors

Xiaodong Cai, *Member, IEEE,* and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—Adaptive modulation requires channel state information (CSI), which can be acquired at the receiver by inserting pilot symbols in the transmitted signal. In this paper, we first analyze the effect linear minimum mean square error (MMSE) channel estimation and prediction errors have on bit-error rate (BER). Based on this analysis, we develop adaptive pilot symbol assisted modulation (PSAM) schemes that account for both channel estimation and prediction errors to meet a target BER. While pilot symbols facilitate channel acquisition, they consume part of transmitted power and bandwidth, which in turn, reduces spectral efficiency. With imperfect (and thus, partial) CSI available at the transmitter and receiver, two questions arise naturally: how often should pilot symbols be transmitted, and how much power should be allocated to pilot symbols. We address these two questions by optimizing pilot parameters to maximize spectral efficiency.**

***Index Terms*—Adaptive modulation, channel estimation, channel prediction, Rayleigh fading.**

I. Introduction

CHANNEL-ADAPTIVE modulation is a promising technique to enhance spectral efficiency of wireless transmissions over fading channels [4], [5], [11], [29]. In adaptive systems, certain transmission parameters such as constellation size, transmitted power, and code rate are dynamically adjusted according to the channel quality, which increases the average spectral efficiency without wasting power or sacrificing error probability performance.

Channel state information (CSI) is required for the transmitter to adapt its parameters, and for the receiver to perform coherent demodulation. It has been shown that adaptive modulation with perfect CSI offers performance gains relative to nonadaptive transmissions [3], [7], [11], [12]. However, adaptive modulation schemes designed based on *perfect* CSI work well only when CSI imperfections induced by channel estimation error and/or feedback delays are sufficiently small [2], [11]. For example, an adaptive transmitter relying on one channel sample to predict future channel values requires feedback delay $\tau < 0.01/f_d$, where f_d is the Doppler frequency [2]. The performance of adaptive coded modulation based on perfect CSI at the transmitter was analyzed in the presence of channel estimation errors in [18], while channel prediction and the effect of channel prediction errors were investigated for adaptive coded modulation in [25] and [23], respectively. While it is necessary to employ reliable channel estimators and predictors to minimize the effects of *imperfect* CSI, adaptive transmitters that account for CSI errors explicitly may have better performance. Adaptive modulation incorporating channel prediction errors were considered in [10], [16], and [30] for single antenna systems, and in [24], [31] for multiantenna systems. It was shown that multiantenna transmissions increase spectral efficiency, as well as relax the requirement on the channel prediction quality [23], [31]. While the effect of channel prediction error was considered in the adaptive transmitter designs of [10], [16], [24], [30], and [31], perfect CSI was assumed at the receiver.

In this paper, we deal with adaptive pilot symbol assisted modulation (PSAM) that accounts for both channel estimation and prediction errors. As advocated in [6] and [22], pilot symbols are periodically inserted in the transmitted signal to facilitate channel estimation and prediction at the receiver. Different from these *nonadaptive* PSAM schemes, we will adjust transmission parameters to maximize spectral efficiency while adhering to a prescribed (target) bit-error rate (BER). Our goals are 1) to design *adaptive* PSAM schemes that take into account both channel estimation and prediction errors to meet the target BER and 2) to optimize the spacing and power of pilot symbols to maximize spectral efficiency. We will focus on adaptive modulation without employing error control coding. Although it is possible to extend our adaptive PSAM to adaptive coded modulation (following the lines of [1], [10], [12], [14], [15], and [28]), we will not pursue this direction here.

The rest of this paper is organized as follows. Section II describes the system model, and studies the effect channel estimation and prediction errors have on BER. In Section III, adaptive PSAM schemes are developed, and their average BER performance is analyzed. Numerical results are presented in Section IV, and conclusions are drawn in Section V.

Notation: Superscripts ^{T}, *, and $^{\mathcal{H}}$ stand for transpose, conjugate, and Hermitian transpose, respectively; $E[\cdot]$ denotes expectation of the random variable within the brackets; and $\lfloor x \rfloor$

Manuscript received February 17, 2003; revised July 28, 2003, October 26, 2003; accepted October 29, 2003. The editor coordinating the review of this paper and approving it for publication is D. Gesbert. This work was prepared through collaborative participation in the Communications and Networks Consortium, sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. This paper was presented in part at the 37th Conference on Information Science and Systems, Baltimore, MD, March 12–14, 2003.

X. Cai was with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA. He is now with the Department of Electrical Engineering and Computer Engineering, University of Miami, Coral Gables, FL 33124 USA (e-mail: x.cai@miami.edu).

G. B. Giannakis is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: georgios@ece.umn.edu).

Digital Object Identifier 10.1109/TWC.2004.840200

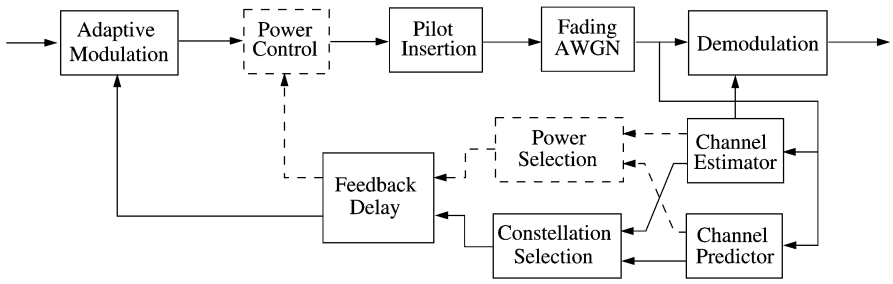

Fig. 1. Adaptive PSAM system model.

represents the smallest integer less than x. Column vectors (matrices) are denoted by boldface lower (upper) case letters; $\mathbf{I}_N$ represents the $N \times N$ identity matrix; and $\mathcal{D}(\mathbf{x})$ stands for the diagonal matrix with $\mathbf{x}$ on its diagonal.

II. System Model and BER Analysis

A. System Model

The adaptive system under consideration is outlined in the block diagram of Fig. 1. A pilot symbol is inserted every $L-1$ information bearing symbols, which results in the transmitted frame structure shown in Fig. 2, where P and D denote pilot and data (information) symbols, respectively. The discrete-time equivalent baseband channel includes transmitter and receiver filters, time-selective frequency-flat fading effects, and additive white Gaussian noise (AWGN). At the receiver, a channel estimator extracts the pilot signal and estimates the channel periodically. Using the estimated channel supplied by the channel estimator, the demodulator performs coherent detection of the data symbols. The pilot symbols are also used by the channel predictor to estimate the channel τ seconds ahead, where τ is the feedback delay that accounts for both actual transmission delay, and processing time at the receiver and transmitter. Based on the predicted channel and the quality of channel estimation and prediction, a constellation size is selected and fed back to the transmitter. There, the adaptive modulator maps incoming binary symbols to the selected constellation. If the instantaneous transmit power is allowed to vary, the transmit power level is determined adaptively along with the constellation size, and is also fed back to the transmitter. If the transmit power is constant, the system does not include the dashed line blocks of power selection and power control depicted in Fig. 1.

Let $r(n;l)$ denote the received signal sampled in the lth symbol period of the nth frame, or equivalently at $t = (nL+l)T$, where T is the symbol (equal to the sampling) period. The received samples corresponding to the pilot symbols can be written as

$$y_p(n) := r(n;0) = \sqrt{\mathcal{E}_p}h(n;0)s_p(n) + \eta(n;0) \qquad (1)$$

and similarly, those corresponding to the data symbols are:

$$\begin{aligned} y_d(n;l) &:= r(n;l) \\ &= \sqrt{\mathcal{E}_d}h(n;l)s(n;l) + \eta(n;l), \quad l \in [1, L-1] \end{aligned} \qquad (2)$$

where $\mathcal{E}_p$ and $\mathcal{E}_d$ represent, respectively, the power per pilot and data symbol, which will be specified later in Section III; $s_p(n)$ is the pilot symbol and $\{s(n;l)\}_{l=1}^{L-1}$ are the data symbols

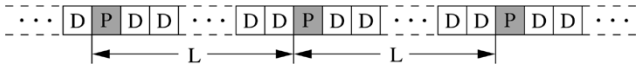

Fig. 2. Transmitted frame structure.

of the nth frame. We assume that $E[|s(n;l)|^2] = |s_p(n)|^2 = 1$; $\eta(\cdot)$ is complex AWGN with zero mean, and variance $N_0/2$ per dimension; and the channel $h(n;l)$ is a stationary complex Gaussian random process with zero mean, and variance $\sigma_h^2 = 1$. The channel statistics including the distribution and the autocorrelation function are assumed to be known at the receiver. Unlike the PSAM in [6], where pilot and data symbols have the same transmit power, and similar to [22], we allow the pilot and data symbols to be transmitted with different power. The PSAM in [6] uses pilot symbols to enable coherent demodulation and the ergodic capacity of time-selective fading channels with PSAM is maximized in [22] without CSI available at the transmitter.

The *adaptive* PSAM we will introduce here uses the estimated channel to perform coherent demodulation, and the predicted channel to adjust the constellation size and transmitted power, i.e., adaptive PSAM exploits partial CSI at the transmitter. Given a total transmit power budget, the power allocation between pilot and data symbols, as well as the pilot spacing L, will be optimized to maximize spectral efficiency with a target BER. Since both channel estimation and prediction errors affect BER, to design the adaptive PSAM, we first need to analyze the BER in the presence of channel estimation and prediction errors.

B. BER in the Presence of Channel Estimation Errors

Let $\hat{h}(n;l)$ be the estimator of $h(n;l)$, and $\epsilon(n;l) := h(n;l) - \hat{h}(n;l)$ denote the channel estimation error. The quality of channel estimation is measured by the channel mean square error (MSE) which is defined as[1] $\sigma_\epsilon^2(l) := E[|\epsilon(n;l)|^2]$. Given $\sigma_\epsilon^2(l)$ and a realization of the channel estimator[2] $\hat{h}(n;l) = \hat{h}_0(n;l)$, our goal in this subsection is to derive the conditional BER $P(e|\hat{h}_0(n;l))$ for binary phase shift keying (BPSK) and square M-QAM. In Section II.C, based on $P(e|\hat{h}_0(n;l))$, we will derive the BER in the presence of both channel estimation and prediction errors, which will be used later in Section III in adapting PSAM to meet the target BER.

We consider the linear MMSE channel estimator, and refer the reader to [6] for the detailed derivation. This estimator uses

[1]For the linear minimum mean square error (MMSE) channel estimator under consideration, it will soon be clear that $\sigma_\epsilon^2(l)$ does not depend on n.

[2]We will use $\hat{h}_0(n;l)$ to denote a realization of the random variable $\hat{h}(n;l)$; similar notation will be used for the predicted channel $\check{h}(n;l)$ in Section II-C.

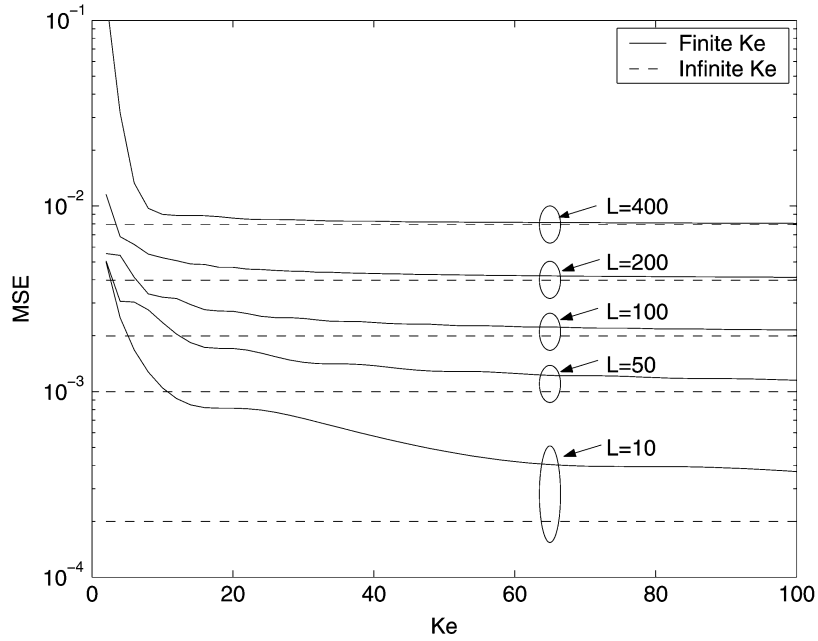

Fig. 3. MSE of channel estimate, classical Doppler spectrum, $f_dT = 0.001$.

K_e pilot samples, $y_p(n - \lfloor K_e/2 \rfloor), \ldots, y_p(n + \lfloor (K_e - 1)/2 \rfloor$, to estimate $\{h(n;l)\}_{l=1}^{L-1}$. Defining $\mathbf{s} := [s_p(n - \lfloor K_e/2 \rfloor), \ldots, s_p(n + \lfloor (K_e - 1)/2 \rfloor)]^T$, $\mathbf{h} := [h[(n - \lfloor K_e/2 \rfloor); 0], \ldots, h[(n + \lfloor (K_e - 1)/2 \rfloor); 0]]^T$, $\mathbf{R} := E[\mathbf{h}\mathbf{h}^{\mathcal{H}}]$, and $\mathbf{r}_l := E[\mathbf{h}h^*(n;l)]$, the linear MMSE channel estimator for $h(n;l)$ is given by $\mathbf{w}_l = \sqrt{\mathcal{E}_p}[\mathcal{E}_p \mathcal{D}(\mathbf{s}) \mathbf{R} \mathcal{D}^*(\mathbf{s}) + N_0 \mathbf{I}_{K_e}]^{-1} \mathcal{D}(\mathbf{s}) \mathbf{r}_l$ [6], which does not depend on n. The estimated channel is obtained as

$$\hat{h}(n;l) = \mathbf{w}_l^{\mathcal{H}} \mathbf{y}(n) \tag{3}$$

where $\mathbf{y}(n) := [y_p(n - \lfloor K_e/2 \rfloor), \ldots, y_p(n + \lfloor (K_e - 1)/2 \rfloor]^T$; and the channel MMSE can be written as [6]

$$\sigma_\epsilon^2(l) = 1 - \sqrt{\mathcal{E}_p} \mathbf{r}_l^{\mathcal{H}} \mathcal{D}^*(\mathbf{s}) \mathbf{w}_l \tag{4}$$

which confirms that indeed $\sigma_\epsilon^2(l)$ does not depend on n.

Since $\mathbf{r}_l$ and $\mathbf{w}_l$ depend on L, K_e, and the Doppler spectrum, we infer from (4) that the channel MMSE is determined by $\mathcal{E}_p/N_0, L, K_e$, and the Doppler spectrum. It will be shown later that the channel mean square error (MSE) has direct impact on BER, and thus, it is desirable to see how these parameters affect $\sigma_\epsilon^2(l)$. In the Appendix, we derive the channel MMSE in the limiting case where $K_e = \infty$, both for the classic Doppler spectrum based on Jakes' channel model, and for the ideal low-pass Doppler spectrum. It is shown that $\sigma_\epsilon^2(l)$ decreases as L decreases, and $\mathcal{E}_p/N_0$ increases. Fig. 3 illustrates $\sigma_\epsilon^2(\lfloor L/2 \rfloor)$ as a function of K_e and L, for the Jakes' spectrum. When L is large, a linear MMSE estimator with about 20 taps has almost the same MSE as the MMSE estimator with infinite number of taps. When L is small, $\sigma_\epsilon^2(l)$ decreases very slowly as K_e increases beyond 20. In practice, the length of the MMSE channel estimator, K_e, is chosen around 20. When $K_e \geq 20$, our calculations show that $\sigma_\epsilon^2(l)$ is almost identical $\forall l \in [1, L-1]$. When the channel Doppler spectrum is modeled as an ideal low-pass spectrum, it turns out that the channel estimator exhibits almost the same MMSE as when the channel is modeled to vary according to Jakes' spectrum, which is also observed in [21, p.651].

We are interested in the BER of square M-QAM and BPSK, given $\hat{h}(n;l) = \hat{h}_0(n;l)$. The optimal receiver can be derived from the maximum likelihood detection rule based on all received data ([21, p. 686]); however, it has prohibitive complexity. Hence, we will pursue a low-complexity symbol-by-symbol detector. If the channel were known perfectly at the receiver, the decision variable for $s(n;l)$ with symbol-by-symbol maximum-likelihood detection would be $z(n;l) = y_d(n;l)/(\sqrt{\mathcal{E}_d} h(n;l))$. With the estimated channel, we can replace $h(n;l)$ with $\hat{h}(n;l)$ in the decision variable, although this detection rule is no longer optimum. Since the channel is a Gaussian random process with zero mean, $\hat{h}(n;l)$ and $\epsilon(n;l)$ are zero mean Gaussian random variables. Because the orthogonality principle renders $\epsilon(n;l)$ uncorrelated with $\hat{h}(n;l)$, the actual channel[3], $\breve{h}(n;l) = \hat{h}_0(n;l) + \epsilon(n;l)$, when $\hat{h}(n;l) = \hat{h}_0(n;l)$ is given, is Gaussian distributed with mean $\hat{h}_0(n;l)$, and variance $\sigma_\epsilon^2(l)$. Hence, given the realization $\hat{h}_0(n;l)$ of the channel estimator, the decision variable for $s(n;l)$ can be written as

$$\begin{aligned} z(n;l) &= \frac{\sqrt{\mathcal{E}_d} \breve{h}(n;l) s + \eta}{\sqrt{\mathcal{E}_d} \hat{h}_0} \\ &= s + \frac{s\epsilon}{\hat{h}_0} + \frac{\eta}{\sqrt{\mathcal{E}_d} \hat{h}_0} \end{aligned} \tag{5}$$

where we omitted the time indexes of $s(n;l), \epsilon(n;l)$, $\hat{h}_0(n;l)$, and $\eta(n;l)$, for notational brevity. Since ϵ and η are independent, the SNR in (5) can be expressed as

$$\gamma_1(n;l) = \frac{|s|^2 \mathcal{E}_d |\hat{h}_0(n;l)|^2}{N_0 + \mathcal{E}_d \sigma_\epsilon^2(l) |s|^2}. \tag{6}$$

For BPSK and 4-QAM, we have $|s| = 1$; and thus, BER can be calculated using the SNR in (6). For M-QAM with $M > 4$,

[3]We use the notation $\breve{h}(n;l)$ to differentiate the actual channel after observing a channel estimate from the actual channel $h(n;l)$ without any CSI.

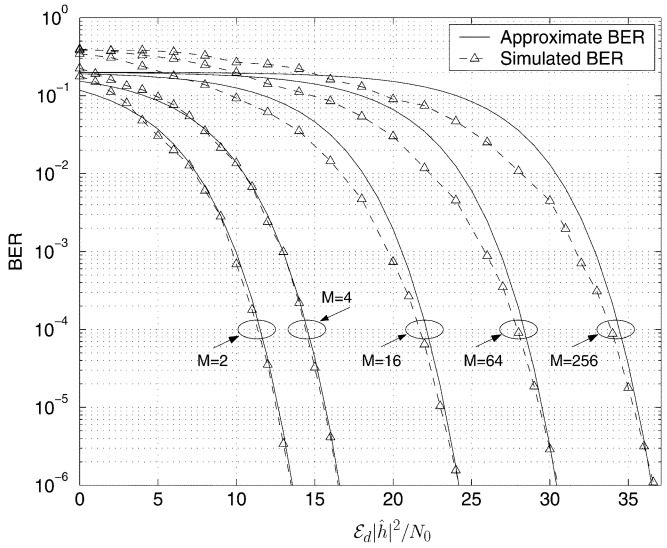

Fig. 4. Conditional BER, $\sigma_\epsilon^2(l) = N_0/\mathcal{E}_d = 0.01$.

we have $|s| \neq 1$; and the SNR in (6) can not be directly used to calculate BER. However, our simulations show that the BER of M-QAM $(4 < M \leq 256)$ with channel estimation error is well approximated by the BER formula using an average SNR given by

$$\gamma_2(n;l) = \frac{\mathcal{E}_d|\hat{h}_0(n;l)|^2}{N_0 + 1.3\,\mathcal{E}_d\sigma_\epsilon^2(l)}. \tag{7}$$

Since $|s|^2 < 2.65$ for $M \leq 256$, the factor 1.3 in the denominator of (7) is found via simulation by picking a number between 0 and 2.65. Considering both BPSK and M-QAM, we will henceforth use the SNR

$$\gamma(n;l) := \frac{\mathcal{E}_d|\hat{h}_0(n;l)|^2}{N_0 + g\mathcal{E}_d\sigma_\epsilon^2(l)} \tag{8}$$

where $g = 1$ for BPSK and 4-QAM, and $g = 1.3$ for *M*-QAM with $M > 4$.

Suppose that N different constellations of size $\{M_i\}_{i=1}^N$ are available, and Gray coding is used to map information bits to QAM symbols. Then, given $\hat{h}(n;l) = \hat{h}_0(n;l)$, the BER of the ith constellation can be approximately calculated by [7]

$$P(e_i|\hat{h}_0(n;l)) \approx 0.2\exp\left(-\frac{c\gamma(n;l)}{M_i - 1}\right) \tag{9}$$

where $\gamma(n;l)$ is given in (8), $c = 3.2/3$ for BPSK, and $c = 1.6$ for square M-QAM. Fig. 4 compares the simulated BER with the approximate BER calculated from (9), when $\sigma_\epsilon^2(l) = N_0/\mathcal{E}_d = 0.01$. The approximation is within 2 dB for *M*-QAM, when the target BER $< 10^{-2}$, or when the SNR is relatively high.

C. BER in the Presence of Channel Estimation and Prediction Errors

Suppose that the length of the linear MMSE channel prediction filter is K_p, and the feedback delay is $\tau = DLT$, where D is a positive integer. For notational simplicity, we assume that the delay is a multiple of LT. The channel predictor estimates $\{h(n;l)\}_{l=1}^{L-1}$, using K_p past pilot samples $\{y_p(n-D-k)\}_{k=0}^{K_p-1}$. Letting $\check{h}(n;l)$ be the predicted channel, the prediction error is given by $\varepsilon(n;l) = h(n;l) - \check{h}(n;l)$, and the MSE of the predicted channel is $\sigma_\varepsilon^2(l) := E[|\varepsilon(n;l)|^2]$. Given the predicted channel $\check{h}(n;l) = \check{h}_0(n;l)$, our adaptive PSAM will adjust the constellation size of $s(n;l)$ based on $\check{h}_0(n;l)$, and the quality (MSE) of channel estimation and prediction. To this end, in this subsection, we will express the BER for BPSK and *M*-QAM as a function of $\check{h}_0(n;l)$, $\sigma_\epsilon^2(l)$, and $\sigma_\varepsilon^2(l)$. Using this BER expression, the adaptive PSAM in Section III will choose transmission parameters to maximize spectral efficiency, while satisfying the target BER.

Defining $\mathbf{s}_p := [s_p(n-D),\ldots,s_p(n-D-K_p+1)]^T$, $\mathbf{h}_p := [h(n-D;0),\ldots,h(n-D-K_p+1;0)]^T$, $\mathbf{R}_{h_p} := E[\mathbf{h}_p\mathbf{h}_p^{\mathcal{H}}]$, and $\mathbf{r}_p(l) = E[\mathbf{h}_p h^*(n;l)]$, the MMSE of the predicted channel is written as

$$\sigma_\varepsilon^2(l) = 1 - \mathcal{E}_p\mathbf{r}_p^{\mathcal{H}}(l)\mathcal{D}^*(\mathbf{s}_p)(\mathcal{E}_p\mathcal{D}(\mathbf{s}_p)\mathbf{R}_{h_p}\mathcal{D}^*(\mathbf{s}_p) + N_0\mathbf{I}_{K_p})^{-1}\mathcal{D}(\mathbf{s}_p)\mathbf{r}_p(l) \tag{10}$$

where similar to $\sigma_\epsilon^2(l)$, $\sigma_\varepsilon^2(l)$ does not depend on n. As shown in [23], the length of the channel predictor, K_p, should be chosen large enough to achieve a sufficiently small MSE with a moderate delay. For a detailed discussion on how K_p affects the performance of the channel predictor, we refer the reader to [23]. Similar to the estimated channel $\hat{h}(n;l)$, the predicted channel $\check{h}(n;l)$ is Gaussian distributed with zero mean, and variance $\sigma_{\check{h}}^2(l) = \sigma_{\hat{h}}^2 - \sigma_\varepsilon^2(l) = 1 - \sigma_\varepsilon^2(l)$.

Considering the relationship between the estimated and predicted channels, we have

$$\begin{aligned}\hat{h}(n;l) &= h(n;l) - \epsilon(n;l)\\ &= \check{h}(n;l) + \varepsilon(n;l) - \epsilon(n;l).\end{aligned} \tag{11}$$

By the orthogonality principle, $\varepsilon(n;l)$ and $\check{h}(n;l)$ are uncorrelated. Based on the MMSE channel estimator and predictor, it is ready to find $\rho := E[\epsilon(n;l)\check{h}^*(n;l)]/\sigma_{\check{h}}^2(l)$; and thus, given $\check{h}(n;l) = \check{h}_0(n;l)$, $\hat{h}(n;l)$ is Gaussian distributed with mean $E[\hat{h}(n;l)|\check{h}_0(n;l)] = (1-\rho)\check{h}_0(n;l)$ and variance $\tilde{\sigma}_{\hat{h}}^2 := E[|\varepsilon(n;l) - \tilde{\epsilon}(n;l)|^2]$, where $\tilde{\epsilon}(n;l) := \epsilon(n;l) - \rho\check{h}_0(n;l)$. Hence, given $\check{h}(n;l) = \check{h}_0(n;l)$, the amplitude of $\hat{h}(n;l)$ follows a Rice distribution with Rician factor $K = |(1-\rho)\check{h}_0(n;l)|^2/\tilde{\sigma}_{\hat{h}}^2$. Letting $p(|\hat{h}_0|\,|\check{h}_0)$ denote this conditional probability density function (pdf), the conditional BER of $s(n;l)$ can be expressed as

$$P(e_i\,|\,\check{h}_0) = \int_0^\infty P(e_i\,||\hat{h}_0|)p(|\hat{h}_0|\,|\check{h}_0)d|\hat{h}_0| \tag{12}$$

where we omitted the time indices of $\hat{h}_0(n;l)$ and $\check{h}_0(n;l)$. For the coherent detection specified by (5), since the SNR in (8) depends on the estimated channel amplitude $|\hat{h}_0(n;l)|$, it is seen from (9) that $P(e_i\,||\hat{h}_0(n;l)|) = P(e_i|\hat{h}_0(n;l))$; thus, we substitute (9) into (12), and obtain [8]

$$P(e_i\,|\,\check{h}_0(n;l)) \approx \frac{0.2\exp\left(-\frac{a_i(l)\mathcal{E}_d|(1+\rho)\check{h}_0(n;l)|^2}{a_i(l)\mathcal{E}_d\tilde{\sigma}_{\hat{h}}^2+1}\right)}{a_i(l)\mathcal{E}_d\tilde{\sigma}_{\hat{h}}^2 + 1} \tag{13}$$

where $a_i(l) := c/[(M_i-1)(N_0+g\mathcal{E}_d\sigma_\epsilon^2(l))]$. If $\sigma_\varepsilon^2(l) = \sigma_\epsilon^2(l) = 0$, then $\check{h}_0(n;l) = \hat{h}_0(n;l) = h_0(n;l)$, and $\rho = 0$; and it is seen from (9) and (13) that $P(e_i|\check{h}_0(n;l)) = P(e_i|h_0(n;l))$, as expected. While it is straightforward to calculate $\tilde{\sigma}_{\hat{h}}^2$ from the MMSE channel estimator and predictor, for simplicity, we will henceforth use $\tilde{\sigma}_{\hat{h}}^2 = \sigma_\epsilon^2(l) + \sigma_\varepsilon^2(l)$, which amounts to assuming that $\epsilon(n;l)$ is uncorrelated with $\varepsilon(n;l)$. The BER calculated under this simplification is almost identical to that computed using the exact $\tilde{\sigma}_{\hat{h}}^2$ due to the following reason. Since the channel estimator is noncausal and the channel predictor is strictly causal, we have $\sigma_\varepsilon^2(l) \gg \sigma_\epsilon^2(l)$; and we also have $\tilde{\sigma}_\epsilon^2(l) < \sigma_\epsilon^2(l)$, where $\tilde{\sigma}_\epsilon^2(l)$ is the variance of $\tilde{\epsilon}(n;l)$. Thus, based on the definition of $\tilde{\sigma}_{\hat{h}}^2$, we deduce that $\tilde{\sigma}_{\hat{h}}^2 \approx \sigma_\epsilon^2(l) + \sigma_\varepsilon^2(l)$.

Since $E[\epsilon(n;l)\check{h}^*(n;l)] = E[\epsilon(n;l)(h^*(n;l) - \varepsilon^*(n;l))] = \sigma_\epsilon^2(l) - E[\epsilon(n;l)\varepsilon^*(n;l)]$, and $|E[\epsilon(n;l)\ \varepsilon^*(n;l)]| \le \sigma_\epsilon(l)\sigma_\varepsilon(l)$, we have

$$\frac{[\sigma_\epsilon(l)-\sigma_\varepsilon(l)]\sigma_\epsilon(l)}{\sigma_{\check{h}}^2(l)} \le \rho \le \frac{[\sigma_\epsilon(l)+\sigma_\varepsilon(l)]\sigma_\epsilon(l)}{\sigma_{\check{h}}^2(l)}. \tag{14}$$

In the practical SNR region, $\sigma_\epsilon(l)/\sigma_{\check{h}}(l)$ and $\sigma_\varepsilon(l)/\sigma_{\check{h}}(l)$ are very small and thus, ρ is also very small. For example, in the case we consider in Section III-C, where $f_dT = 10^{-3}, \tau = 0.2/f_d, K_e = 20$, and $K_p = 250$, we can find that $-0.14 < 10\log_{10}(1+\rho)^2 < 0.20$ for $L = 10$ and $\mathcal{E}_p/N_0 = 10$ dB. Hence, in this case, the BER calculated using $\rho = 0$ will be within 0.2 dB relative to that computed using the exact ρ. In order to gain insight on the effects channel estimation and prediction errors have on BER, let us set $\rho = 0$. The BER $P(e_i\,|\,\check{h}_0(n;l))$ is mainly determined by the exponential term in (13). With $\rho = 0$, we can express this exponent as

$$\frac{a_i(l)\mathcal{E}_d|\check{h}_0(n;l)|^2}{a_i(l)\mathcal{E}_d\tilde{\sigma}_{\hat{h}}^2+1} = \frac{c\mathcal{E}_d|\check{h}_0(n;l)|^2}{(M_i-1)N_0} \times G \tag{15}$$

where

$$G = \frac{N_0}{N_0 + \mathcal{E}_d\sigma_\epsilon^2(l)(g + c/(M_i-1)) + c\mathcal{E}_d\sigma_\varepsilon^2(l)/(M_i-1)} \tag{16}$$

reflects the performance loss caused by channel estimation and prediction. From (16), we observe that the same channel estimation error causes almost the same performance loss for all M-QAM and BPSK constellations, since the term $\mathcal{E}_d\sigma_\epsilon^2(l)(g + c/(M_i-1))$ changes very slowly when M_i increases beyond a moderately large value. On the other hand, the same channel prediction error incurs less performance loss for large constellation sizes, because the term $c\mathcal{E}_d\sigma_\varepsilon^2(l)/(M_i-1)$ decreases as M_i increases. The BER in (13) is obtained by averaging (9) over the Rician pdf of the channel amplitude. Due to the term $M-1$ in the denominator of the exponent in (9), when the constellation size increases, the variance of the channel amplitude effectively decreases, which reduces the performance loss incurred by channel prediction errors.

Let $\bar{\mathcal{E}}_d$ denote the average transmit power of the data symbols. Defining $\check{\gamma}(l) := \bar{\mathcal{E}}_d|\check{h}_0(n;l)|^2/N_0$, $b_i(l) := a_i(l)\mathcal{E}_d\tilde{\sigma}_{\hat{h}}(l)^2+1$, and $d_i(l) := a_i(l)\mathcal{E}_dN_0(1+\rho)^2/(\bar{\mathcal{E}}_d b_i(l))$, we can also write $P(e_i|\check{h}_0(n;l))$ in (13) as

$$P(e_i\,|\,\check{\gamma}(l)) \approx \frac{0.2\exp(-d_i(l)\check{\gamma}(l))}{b_i(l)} \tag{17}$$

where we omitted the frame index n for notational brevity. Letting $\bar{\check{\gamma}}(l) := \bar{\mathcal{E}}_d\sigma_{\hat{h}}^2(l)/N_0 = \bar{\mathcal{E}}_d(1-\sigma_\varepsilon^2(l))/N_0$, the pdf of $\check{\gamma}(l)$ can be expressed as

$$p(\check{\gamma}(l)) = \frac{\exp(-\check{\gamma}(l)/\bar{\check{\gamma}}(l))}{\bar{\check{\gamma}}(l)}. \tag{18}$$

This pdf will be used in the next section to develop the adaptive PSAM, as well as to evaluate the average BER.

III. Adaptive PSAM

Let $\mathcal{E}$ be the total average power including both pilot and data power. Clearly, $\bar{\mathcal{E}}_d = \alpha L\mathcal{E}/(L-1)$, and $\mathcal{E}_p = (1-\alpha)L\mathcal{E}$, where α $(0<\alpha<1)$ determines the power allocation between data and pilot symbols. Note that $\alpha = (L-1)/L$ corresponds to equal power allocation. The channel MSE can be explicitly expressed as a function of α. Let $\mathbf{u}_i$ be the ith eigenvector of $\mathbf{R}$, and λ_i be the corresponding eigenvalue. Using the constant modulus property of pilot symbols, we can write $\sigma_\epsilon^2(l)$ in (4) as

$$\begin{aligned}\sigma_\epsilon^2(l) &= 1 - \sum_{i=1}^{K_e} \frac{\mathcal{E}_p\left|\mathbf{u}_i^{\mathcal{H}}\mathbf{r}_l\right|^2}{\mathcal{E}_p\lambda_i + N_0} \\ &= 1 - \sum_{i=1}^{K_e} \frac{\left|\mathbf{u}_i^{\mathcal{H}}\mathbf{r}_l\right|^2(1-\alpha)L\bar{\gamma}}{(1-\alpha)L\bar{\gamma}\lambda_i + 1}\end{aligned} \tag{19}$$

where $\bar{\gamma} := \mathcal{E}/N_0$ is the average transmit SNR. Since $\mathbf{R}$ depends on L, so does $\sigma_\epsilon^2(l)$. The MSE of the predicted channel, $\sigma_\varepsilon^2(l)$, can also be written in a similar formula.

Based on the feedback, our goal in this section is to adapt the constellation size M_i, and possibly the power transmitted per constellation, as well as the spacing L, and the power level (α) of PSAM, so as to maximize spectral efficiency, while adhering to the target BER. The BER in (17) depends on the predicted channel value $\check{h}_0(n;l)$, as well as the channel MSE $\sigma_\epsilon(l)$ and $\sigma_\varepsilon(l)$. As we claimed in Section II.B, when $K_e \ge 20$, the MSE of the estimated channel is almost identical for any l. The MSE of the predicted channel $\check{h}(n;l)$ is largest when $l = L-1$, since the prediction range for $h(n;l)$ is $(DL+l/L)T$, which is maximized at $l = L-1$. Hence, the receiver can use $\sigma_\epsilon(L-1)$ and $\sigma_\varepsilon(L-1)$ in calculating the constellation switching thresholds (which will be specified later) for $\{s(n;l)\}_{l=1}^{L-1}$. With a small penalty in spectral efficiency, this reduces computational complexity to $1/L$ relative to the case where the switching thresholds of $\{s(n;l)\}_{l=1}^{L-1}$ are, respectively, calculated using $\{\sigma_\epsilon(l-1)\}_{l=1}^{L-1}$ and $\{\sigma_\varepsilon(l-1)\}_{l=1}^{L-1}$. When the channel changes very slowly in a frame, this also reduces the feedback rate, since $\check{h}_0(n;l)$ is almost identical $\forall l$, and the same constellation can be used for $\{s(n;l)\}_{l=1}^{L-1}$. We will henceforth omit the time indexes in $\check{\gamma}_i(l), \bar{\check{\gamma}}(l), a_i(l), b_i(l)$, and $d_i(l)$, for notational simplicity.

A. Constant Power

Let $\check{\gamma}_i$ denote the constellation switching threshold such that the ith constellation will be transmitted if $\check{\gamma} \in [\check{\gamma}_i\ \check{\gamma}_{i+1})$, where $\check{\gamma}_0 = 0, \check{\gamma}_{N+1} = \infty$, and $\{\check{\gamma}_i\}_{i=1}^N$ will be found later to meet a target BER. In this subsection, we assume that the transmit power is constant, and no data is transmitted when $\check{\gamma} < \check{\gamma}_1$. Then, the actual transmit power is [7]

$$\mathcal{E}_d = \frac{\bar{\mathcal{E}}_d}{\int_{\check{\gamma}_1}^{\infty} p(\check{\gamma})\, d\check{\gamma}} = \bar{\mathcal{E}}_d \exp(\check{\gamma}_1/\bar{\check{\gamma}}). \tag{20}$$

Suppose that the target BER is equal to B. Using (17) and letting $P(e\,|\,\check{\gamma}_i) = B$, we can find the switching thresholds

$$\check{\gamma}_i = -\frac{\ln(5b_iB)}{d_i}, \qquad i = 1, \ldots, N. \tag{21}$$

Letting $k_i := \log_2(M_i)$, and $P(k_i) := P(\check{\gamma}_i < \check{\gamma} < \check{\gamma}_{i+1})$ $= \exp(-\check{\gamma}_i/\bar{\check{\gamma}}) - \exp(-\check{\gamma}_{i+1}/\bar{\check{\gamma}})$, the spectral efficiency can be expressed as

$$\begin{aligned} S &= \left(1 - \frac{1}{L}\right) \sum_{i=1}^{N} k_i P(k_i) \\ &= \left(1 - \frac{1}{L}\right) \left[k_1 \exp(-\check{\gamma}_1/\bar{\check{\gamma}}) \right. \\ &\qquad \left. + \sum_{i=2}^{N} (k_i - k_{i-1}) \exp(-\check{\gamma}_i/\bar{\check{\gamma}}) \right]. \end{aligned} \tag{22}$$

It is evident from (19) that the MSE of the estimated channel depends on L and α. Similarly, the MSE of the predicted channel is also a function of L and α. Hence, the switching thresholds, as well as the spectral efficiency depend on L and α. Numerical techniques can be used to find the optimum L and α to maximize S. Specifically, the maximum value of L can be found as $L_{\max} = \lfloor 1/(2f_dT) \rfloor$ [6], [22]. For each $L \in [2, L_{\max}]$, we have the following optimization problem to be solved for the optimum α and $\mathcal{E}_d$:

$$\begin{aligned} &\max_{\alpha, \mathcal{E}_d} \quad S(\alpha, \mathcal{E}_d) \\ &\text{subject to} \begin{cases} 0 < \alpha < 1 \\ \mathcal{E}_d = \alpha L \mathcal{E} \exp(-\ln(5b_1B)/(d_1\bar{\check{\gamma}}))/(L-1) \end{cases} \end{aligned} \tag{23}$$

where the last constraint follows from (20), and we explicitly write S as a function of α, and $\mathcal{E}_d$. Numerical search based on e.g., sequential quadratic programming (SQP) [9], can be used to solve this optimization problem with starting point $\alpha = (L-1)/L$, $\mathcal{E}_d = \bar{\mathcal{E}}_d$. After solving (23) for each L, the optimum L and the maximum spectral efficiency can be found by searching over all possible values of L. If the pilot and data symbols have the same transmit power, we let $\alpha = (L-1)/L$ and the same procedure can be used to find the maximum spectral efficiency, as well as the optimum L.

B. Discrete Power

Here, we allow symbols chosen from different constellation sizes to be transmitted with different power levels. For the perfect CSI case, it has been shown that the spectral efficiency of discrete power transmissions is within 2 dB of those with continuous power adaptation [11], and is higher than that of constant power transmission [7]. Here, the advantage of discrete power adaptation relative to constant power transmission may be even larger because discrete power adaptation increases the probability of using large constellations, and as corroborated by (16), the same channel prediction error incurs smaller performance loss for larger constellations.

Letting $\mathcal{E}_{di}$ be the transmit power of the ith constellation, the switching threshold $\check{\gamma}_i$ can be calculated from (21) with $\mathcal{E}_d$ being replaced by $\mathcal{E}_{di}$. Since the average transmit power of the data symbols is $\bar{\mathcal{E}}_d$, we must have

$$\sum_{i=1}^{N} \mathcal{E}_{di} P(k_i) = \bar{\mathcal{E}}_d. \tag{24}$$

For each L, we can formulate the following optimization problem to be solved for the optimum α and $\{\mathcal{E}_{di}\}_{i=1}^N$:

$$\begin{aligned} &\max_{\alpha, \{\mathcal{E}_{di}\}_{i=1}^N} \quad S(\alpha, \mathcal{E}_{d1}, \ldots \mathcal{E}_{dN}) \\ &\text{subject to} \begin{cases} 0 < \alpha < 1 \\ \mathcal{E}_{di} > 0, \quad i = 1, \ldots, N \\ \sum_{i=1}^{N} \mathcal{E}_{di} [\exp(-\check{\gamma}_i/\bar{\check{\gamma}}) \\ \quad - \exp(-\check{\gamma}_{i+1}/\bar{\check{\gamma}})] = \alpha L \mathcal{E}/(L-1) \end{cases} \end{aligned} \tag{25}$$

where the last equation in the constraints follows from (24), and we explicitly write S as a function of α, and $\{\mathcal{E}_{di}\}_{i=1}^N$. This optimization problem can also be solved by numerical search based on e.g., SQP [9], using $\alpha = (L-1)/L$ and $\mathcal{E}_{di} = \mathcal{E}, \forall i$, as the starting point. Similar to the constant power case, the optimum L and the maximum spectral efficiency can be found by searching over all possible values of L, using the solution to the optimization problem (25).

C. Average BER Performance Analysis

Since the switching threshold $\check{\gamma}_i$ is chosen such that the BER of the ith constellation meets the target BER: $P(e|\check{\gamma}_i) = B$, and the same constellation is used in the interval $[\check{\gamma}_i, \check{\gamma}_{i+1})$, the actual average BER is expected to be less than the target BER. In this subsection, we will analyze the average BER performance of the adaptive PSAM schemes developed in Sections III-A and III-B.

The average BER for the ith constellation can be written as

$$P(e_i) = \int_{\check{\gamma}_i}^{\check{\gamma}_{i+1}} P(e_i\,|\,\check{\gamma}) p(\check{\gamma})\, d\check{\gamma}. \tag{26}$$

The overall average BER is then computed as the ratio of the average number of bits in error over the total average number of transmitted bits [2], [7]

$$P(e) = \frac{\sum_{i=1}^{N} k_i P(e_i)}{\sum_{i=1}^{N} k_i P(k_i)}. \tag{27}$$

Substituting the conditional BER in (17) and the pdf of the predicted SNR in (18) into (26), we obtain

$$P(e_i) \approx \frac{0.2}{b_i \beta_i} \left[\exp\left(-\frac{\beta_i \check{\gamma}_i}{\bar{\check{\gamma}}}\right) - \exp\left(-\frac{\beta_i \check{\gamma}_{i+1}}{\bar{\check{\gamma}}}\right) \right] \tag{28}$$

where $\beta_i := \bar{\check{\gamma}} d_i + 1$. Then, the average BER in (27) can be approximately calculated by substituting (28) into (27). The conditional BER $P(e_i \,|\, \check{\gamma})$ in (17) is derived from the BER approximation in (9) for AWGN channels, which is within 1 dB accurate when the true BER is less than 10^{-3} [7]. The accuracy of the average BER calculated from (28) is thus limited by the BER approximation $P(e_i \,|\, \check{\gamma})$. In the following, we will derive a more accurate approximation for the average BER.

An approximate expression for the BER of QAM in AWGN, which is very accurate across the SNR region, was derived in [20]

$$P_G(e_i \,|\, \gamma) \approx 4\frac{\sqrt{M_i}-1}{k_i\sqrt{M_i}} \sum_{j=1}^{\sqrt{M_i}/2} Q\left[(2j-1)\sqrt{3\gamma/(M_i-1)}\right] \tag{29}$$

where $Q(x) := (1/\sqrt{2\pi}) \int_x^\infty \exp(-x^2/2)dx$, and γ denotes the SNR. Based on the SNR in (8), we can use (29) instead of (9) to calculate the BER in the presence of channel estimation error. As we discussed in Section II-C, the amplitude of the estimated channel, $|\hat{h}(n;l)|$, given the predicted channel $\check{h}(n;l) = \check{h}_0(n;l)$, is Rice distributed. Hence, γ in (8), conditioned on the predicted SNR, $\check{\gamma}$, follows a noncentral chi-square distribution. Unfortunately, it is difficult to obtain a tractable expression for the BER, $P(e_i|\check{\gamma})$ by averaging $P_G(e_i|\gamma)$ in (29) over the noncentral chi-square distribution of γ. If we rely on numerical integration to calculate the average BER from (26) and (29), then a triple integral is required. Furthermore, since two of the three upper limits of the triple integral are infinity, their numerical evaluation is considerably complicated. We next derive a simpler method to compute the average BER.

It is well known that a Rice distribution with Rician factor K can be well approximated by a Nakagami-m distribution with parameter m satisfying $m = (1+K)^2/(1+2K)$ ([27], p. 23). Then, the Rice distribution of $|\hat{h}(n;l)|$ conditioned on $\check{h}(n;l) = \check{h}_0(n;l)$ can be closely approximated by a Nakagami-m distribution with parameter

$$\begin{aligned} m &= \frac{\left(1+|(1-\rho)\check{h}_0(n;l)|^2 \Big/ \tilde{\sigma}^2_{\hat{h}}\right)^2}{1+2|(1-\rho)\check{h}_0(n;l)|^2 \Big/ \tilde{\sigma}^2_{\hat{h}}} \\ &= \frac{\left(1+(1-\rho)^2\check{\gamma}N_0 \Big/ \left(\bar{\mathcal{E}}_d\tilde{\sigma}^2_{\hat{h}}\right)\right)^2}{1+2(1-\rho)^2\check{\gamma}N_0 \Big/ \left(\bar{\mathcal{E}}_d\tilde{\sigma}^2_{\hat{h}}\right)}. \end{aligned} \tag{30}$$

The distribution of γ conditioned on $\check{\gamma}$ is thus well approximated by the gamma distribution ([27, p.22])

$$p(\gamma \,|\, \check{\gamma}) = \frac{m^m \gamma^{m-1}}{\bar{\gamma}_c^m \Gamma(m)} \exp\left(-\frac{m\gamma}{\bar{\gamma}_c}\right) \tag{31}$$

where m is given by (30), $\Gamma(\,\cdot\,)$ is the gamma function, and $\bar{\gamma}_c$ is the expectation of γ conditioned on $\check{\gamma}$, which is given by $\bar{\gamma}_c = (\check{\gamma}N_0/\bar{\mathcal{E}}_d)(\mathcal{E}_{di}/(N_0 + g\mathcal{E}_{di}\sigma^2_\epsilon))$.

Let $p_m(x)$ denote a Nakagami-m distribution with parameter m, and define $\bar{x} := \int_0^\infty x p_m(x)\,dx$. It has been shown in ([27, p.103]) that

$$\begin{aligned} f(a,m,\bar{x}) &:= \int_0^\infty Q(a\sqrt{x})p_m(x)\,dx \\ &= \frac{1}{2\sqrt{\pi}} \frac{\sqrt{a^2\bar{x}/(2m)}}{[1+a^2\bar{x}/(2m)]^{m+1/2}} \frac{\Gamma(m+1/2)}{\Gamma(m+1)} \\ &\quad \times {}_2F_1\left(1, m+1/2; m+1; \frac{m}{m+a^2\bar{x}/2}\right) \end{aligned} \tag{32}$$

where ${}_2F_1(\cdot,\cdot;\cdot;\cdot)$ is the Gauss hypergeometric function ([19, p. 238], and m is not an integer. Using (29), (31), and (32), $P(e_i \,|\, \check{\gamma})$ can be found as

$$\begin{aligned} P(e_i \,|\, \check{\gamma}) &\approx \int_0^\infty P_G(e_i \,|\, \gamma) p(\gamma \,|\, \check{\gamma})\,d\gamma \\ &= 4\frac{\sqrt{M_i}-1}{k_i\sqrt{M_i}} \\ &\quad \times \sum_{j=1}^{\sqrt{M_i}/2} f((2j-1)\sqrt{3/(M_i-1)}, m, \bar{\gamma}_c) \end{aligned} \tag{33}$$

where m is given by (30). Substituting (33) into (26), we can obtain the average BER $P(e_i)$ for the ith constellation by using numerical integration. The overall average BER $P(e)$ in (27) can then be calculated from $P(e_i)$. Since $P(e_i \,|\, \check{\gamma})$ in (33) offers a tighter approximation than (17), the overall average BER calculated from (33) is more accurate than that calculated from (28).

IV. Numerical Results

We consider an adaptive modulation system which employs 4 different square *M*-QAM constellations corresponding to 2 (4-QAM), 4 (16-QAM), 6 (64-QAM), and 8 (256-QAM) bits/symbol, in addition to BPSK. The carrier frequency is $f_c = 2$ GHz; and the symbol rate is 200 ksps, which corresponds to a symbol period of 5 μs. We consider a delay $\tau = 1$ ms, a mobile velocity $v = 108$ km/h or $v = 7$ km/h. Hence, the Doppler spread is $f_d = 200$ Hz for $v = 108$ km/h, or $f_d = 13$ Hz for $v = 7$ km/h; the normalized Doppler spread is $f_dT = 10^{-3}$ for $v = 108$ km/h, or, $f_dT = 6.5 \times 10^{-5}$ for $v = 7$ km/h. The delay can also be written as $\tau = 0.2/f_d$ for $v = 108$ km/h, or $\tau = 0.013/f_d$ for $v = 7$ km/h. We assume that the channel has a Jakes' Doppler spectrum. We choose a target BER equal to 10^{-5}, and the length of the channel estimation (prediction) filter to be $K_e = 20$ ($K_p = 250$). The spectral efficiency of constant power adaptive PSAM is shown in Fig. 5. The channel capacity with perfect CSI, as well as lower bounds on channel capacity with estimated channels, which were derived in [22], are also depicted in Fig. 5. With a small Doppler spread corresponding to $v = 7$ km/h, the spectral efficiency of our adaptive PSAM is very close to that with perfect CSI. When the mobile speed is $v = 108$ km/h, the Doppler spread is relatively large; thus, more bandwidth and power are allocated to pilot symbols, which in turn decreases the spectral efficiency. It is seen that the scheme with optimum

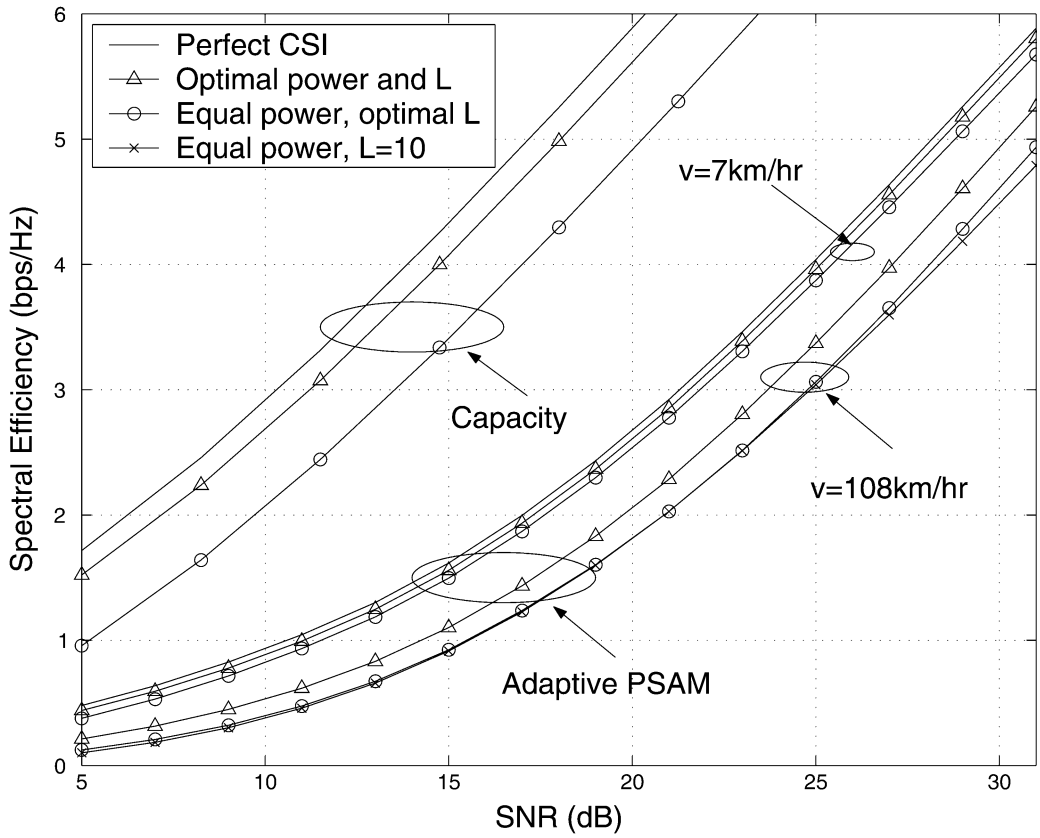

Fig. 5. Spectral efficiency of constant power adaptive modulation.

power allocation has about 1 dB advantage over the equal power scheme at $\mathrm{SNR} = 15$ dB for $v = 108$ km/h. With optimum training, channel capacity is very close to that with perfect CSI, because only channel estimation error affects capacity. On the other hand, both channel prediction and estimation errors degrade the performance of adaptive PSAM. Hence, adaptive PSAM with optimally estimated channels loses considerably in spectral efficiency compared to that with perfect CSI, when the Doppler frequency is relatively large. Fig. 6 compares spectral efficiency of constant power adaptive PSAM with and without channel estimation errors. We here use the same setting of pilot symbols as in Fig. 5. It is observed from Fig. 6 that with optimal power allocation, channel estimation errors are so small that they do not affect spectral efficiency and the estimated channel can be treated as perfect. On the other hand, with equal power allocation, channel estimation errors cause relatively large spectral efficiency loss. Note that we here consider frequency-flat fading channels. If the fading channel is frequency-selective, or if multiple transmit antennas are deployed, then channel estimation errors will be even larger because the number of unknowns to be estimated increases. Hence, channel estimation errors should be taken into account in designing adaptive PSAM. Fig. 7 shows the spectral efficiency of discrete power adaptive PSAM. Comparing Fig. 5 with Fig. 7, we infer that the discrete power adaptive PSAM has larger spectral efficiency than the constant power scheme, as expected. Notice that the discrete power scheme has higher complexity and requires larger feedback bandwidth. From Fig. 7, we see that the optimum power allocation has a higher spectral efficiency than the equal power allocation scheme. The optimum L and α are depicted in Figs. 8 and 9, respectively. Figs. 8 and 9 reveal that the optimum training allocates more power to pilot symbols, while increasing the pilot spacing L, to maintain the quality of the channel prediction and estimation. The net effect is that the spectral efficiency is maximized. Fig. 10 shows the average BER of both constant and discrete power adaptive PSAM for $v = 108$ km/h. We observe that the average BER is much lower than the target BER, even though the feedback delay, $\tau = 0.2/f_d$, is relatively large. On the other hand, our calculations show that the MSE of the predicted channel in this case is larger than the critical value derived in [32] across the SNR region; thus, the adaptive modulation schemes treating the predicted channel as perfect CSI are not able to meet the target BER with such a large feedback delay. This clearly shows the advantage of our adaptive PSAM that accounts for both channel estimation and prediction errors.

V. Conclusion

We have studied adaptive transmission systems with pilot symbol assisted estimation and prediction of rapidly fading channels. The effect of linear MMSE channel estimation and prediction errors on BER was investigated for BPSK and square M-QAM. Adaptive pilot symbol assisted modulation schemes that account for both channel estimation and prediction errors were developed and their BER performance was analyzed. The spacing between two consecutive pilot symbols and the power allocation between pilot and data symbols were optimized to maximize spectral efficiency. Numerical results show that our adaptive PSAM schemes work well even when the feedback

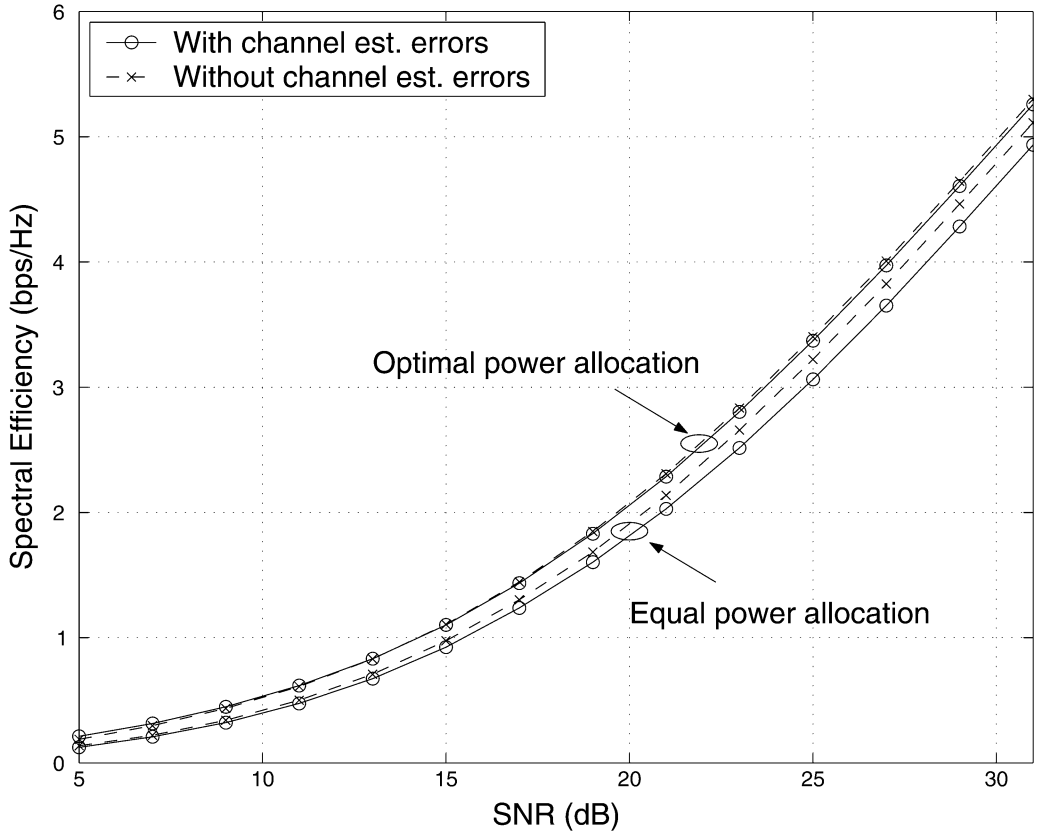

Fig. 6. Effects of channel estimation errors on spectral efficiency of constant power adaptive modulation, $v = 108$ km/h.

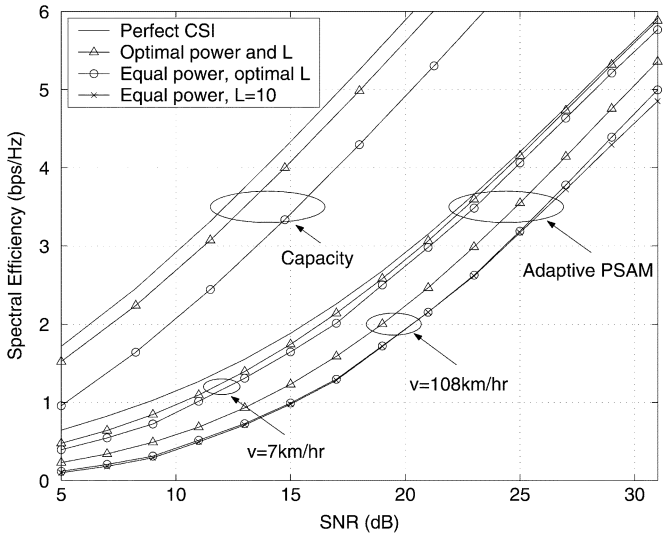

Fig. 7. Spectral efficiency of discrete power adaptive modulation.

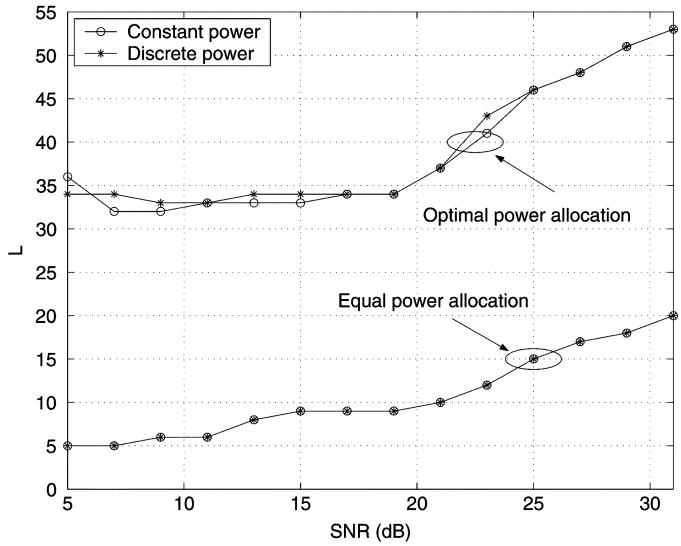

Fig. 8. Optimum pilot spacing, $v = 108$ km/h.

delay is relatively large, and that our optimized pilots improve spectral efficiency.

APPENDIX

A. Channel MSE When $K_e = \infty$

Supposing that we have $Lf_dT < 1/2$, the channel MSE in the limiting case where $K_e = \infty$, can be expressed as ([26, p. 487])

$$\tilde{\sigma}_\epsilon^2(l) = \int_{-1/2}^{1/2} \frac{S_h(f)}{1 + \mathcal{E}_p S_h(f)/(LN_0)}\, df \tag{34}$$

where $S_h(f)$ is the Doppler spectrum of the discrete channel. While $\tilde{\sigma}_\epsilon^2(l)$ is common $\forall l$, $\sigma_\epsilon^2(l)$ in (4) is generally a function of l. For the classic Doppler spectrum based on Jakes' model, $S_h(f) = (\pi\sqrt{(f_dT)^2 - f})^{-1}$, $|f| < f_dT$, and

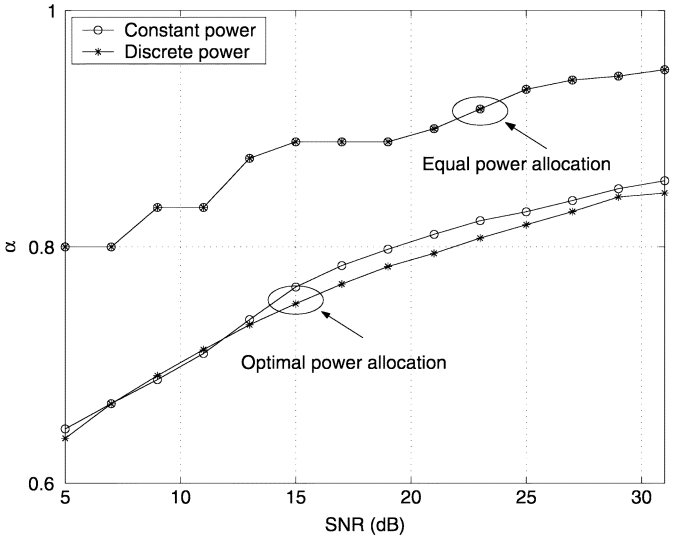

Fig. 9. Optimum power allocation, $v = 108$ km/h.

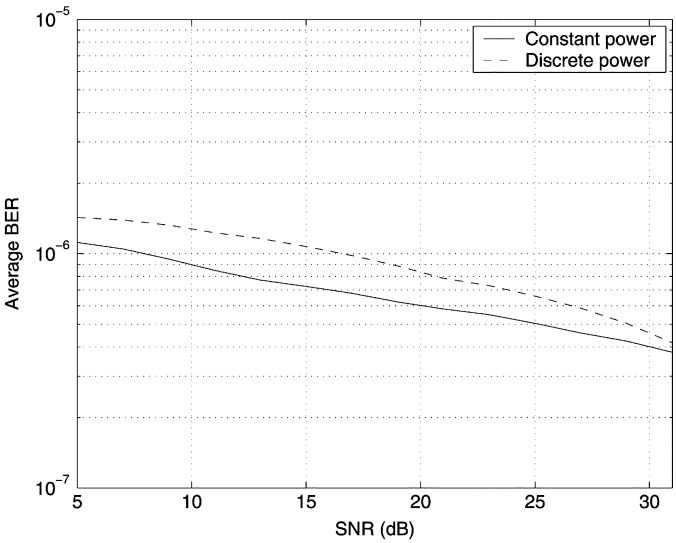

Fig. 10. Average BER, $v = 108$ km/h.

$S_h(f) = 0, f_dT < |f| < 1/2$ [17]. Substituting the expression of Doppler spectrum into (34), the channel MSE $\tilde{\sigma}^2_\epsilon(l)$ becomes

$$\tilde{\sigma}^2_{\epsilon,J} = \int_{-f_dT}^{f_dT} \frac{1}{\sqrt{(f_dT)^2 - f^2} + \mathcal{E}_p/(LN_0)}\, df. \tag{35}$$

Letting $f = f_dT\sin(x)$, we can write (35) as

$$\tilde{\sigma}^2_{\epsilon,J} = 1 - \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} \frac{1}{1+\mu\cos(x)}\, dx. \tag{36}$$

Using the formula ([13, p. 168])

$$\int \frac{1}{b + a\cos(x)}\, dx = \frac{2}{\sqrt{b^2 - a^2}} \arctan\left[\sqrt{\frac{b-a}{b+a}}\tan(x/2)\right] \tag{37}$$

where b and a are two constants, and $b > a$, (36) becomes

$$\tilde{\sigma}^2_{\epsilon,J} = 1 - \frac{4}{\pi\sqrt{1-\mu^2}} \arctan\left(\sqrt{\frac{1-\mu}{1+\mu}}\right) \tag{38}$$

where $\mu := \pi L f_d T N_0/\mathcal{E}_p$. To gain more insight into the channel estimation error, let us specialize to the ideal lowpass Doppler spectrum, which is given by $S_h(f) = 1/(2f_dT), |f| < f_dT$, and $S_h(f) = 0, f_dT < |f| < 1/2$; the MSE $\tilde{\sigma}^2_\epsilon(l)$ in (34) then becomes

$$\tilde{\sigma}^2_{\epsilon,I} = \left(1 + \frac{\mathcal{E}_p}{N_0}\frac{1}{2Lf_dT}\right)^{-1}. \tag{39}$$

When $\mathcal{E}_p/N_0$ is sufficiently large, $\tilde{\sigma}^2_{\epsilon,I} \approx (2Lf_dT)N_0/\mathcal{E}_p$. Since $2Lf_dT < 1$, we have $\tilde{\sigma}^2_{\epsilon,I} < N_0/\mathcal{E}_p$. To lower $\tilde{\sigma}^2_{\epsilon,I}$ we can either reduce L and/or increase $\mathcal{E}_p/N_0$. □

ACKNOWLEDGMENT

The authors are grateful to Dr. S. Ohno at Hisroshima University, Japan, for providing the program to calculate the channel capacity with optimum training [22]. The authors would like to thank the anonymous reviewers for their careful reading and critique of the manuscript. Their comments and suggestions improved the presentation of this paper.

REFERENCES

[1] S. M. Alamouti and S. Kallel, "Adaptive trellis-coded multiple-phase-shift keying for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 42, no. 6, pp. 2305–2314, Jun. 1994.

[2] M. S. Alouini and A. J. Goldsmith, "Adaptive modulation over Nakagami fading channels," *Kluwer J. Wireless Commun.*, vol. 13, pp. 119–143, May 2000.

[3] M. S. Alouini, X. Tang, and A. J. Goldsmith, "An adaptive modulation scheme for simultaneous voice and data transmission over fading channels," *IEEE J. Select. Areas Commun.*, vol. 17, no. 5, pp. 837–850, May 1999.

[4] S. Catreux, V. Erceg, D. Gesbert, and R. W. Heath, "Adaptive modulation and MIMO coding for broadband wireless data networks," *IEEE Commn. Mag.*, vol. 40, pp. 108–115, Jun. 2002.

[5] J. K. Cavers, "Variable-rate transmission for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 20, no. 2, pp. 15–22, Feb. 1972.

[6] ——, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 6, pp. 686–693, Nov. 1991.

[7] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: A unified view," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1561–1571, Sep. 2001.

[8] D. Divsalar and M. K. Simon, "The design of trellis coded MPSK for fading channels: Performance criteria," *IEEE Trans. Commun.*, vol. 36, no. 9, pp. 1004–1012, Sept. 1988.

[9] R. Fletcher, *Practical Methods of Optimization, Constrained Optimization*. New York: Wiley, 1980, vol. 2.

[10] D. L. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 844–855, Jun. 1999.

[11] A. J. 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.

[12] ——, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 595–602, May 1998.

[13] I. S. Gradshtein and I. Ryzhik, *Table of Integrals, Series, and Products*. San Diego, CA: Academic, 2000.

[14] L. Hanzo, C. H. Wong, and M. S. Yee, *Adaptive Wireless Transceiver*. New York: Wiley, 2002.

[15] K. J. Hole, H. Holm, and G. E. Øien, "Adaptive multidimensional coded modulation over flat fading channels," *IEEE J. Select. Areas Commun.*, vol. 18, no. 7, pp. 1153–1158, Jul. 2000.

[16] S. Hu, A. Duel-Hallen, and H. Hallen, "Long range prediction makes adaptive modulation feasible for realistic mobile radio channels," in *Proc. 34rd Ann. Conf. Inform. Sci. Syst.*, Princeton, NJ, Mar. 2000, pp. WP4-7–WP4-13.
[17] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.
[18] V. K. N. Lau and M. D. Macleod, "Variable-rate trellis coded QAM for flat-fading channels," *IEEE Trans. Commun.*, vol. 49, pp. 1550–1560, Sep. 2001.
[19] N. N. Lebedev and R. A. Silverman, *Special Functions & Their Applications*. New York: Dover , 1972.
[20] J. Lu, K. B. Letaief, J. C.-I. Chung, and M. L. Liou, "M-PSK and M-QAM BER computation using signal-space concepts," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 181–184, Feb. 1999.
[21] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing*. New York: Wiley, 1998.
[22] S. Ohno and G. B. Giannakis, "Average-rate optimal PSAM transmissions over time-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 10, pp. 712–720, Oct. 2002.
[23] G. E. Øien, H. Holm, and K. J. Hole, "Impact of imperfect channel prediction on adaptive coded modulation performance," *IEEE Trans. Veh. Technol.*, no. 3, pp. 758–769, May, 2004.
[24] ——, "Adaptive coded modulation with imperfect channel state information: System design and performance analysis aspects," in *Proc. Int. Symp. Adv. Wireless Commun.*, Victoria, BC, Canada, Sep. 2002, pp. 19–20.
[25] ——, "Channel prediction for adaptive coded modulation in Rayleigh fading," in *Proc. 11th Euro. Signal Process. Conf.*, Toulouse, France, Sep. 3–6, 2002.
[26] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1991.
[27] M. K. Simon and M.-S. Alouini, *Digital Communication Over Fading Channels: A Unified Approach to Performance Analysis*. New York: Wiley, 2000.
[28] B. Vucetic, "An adaptive coding scheme for time-varying channels," *IEEE Trans. Commun.*, vol. 39, no. 5, pp. 653–663, May 1991.
[29] W. T. Webb and R. Steele, "Variable rate QAM for mobile radio," *IEEE Trans. Commun.*, vol. 43, no. 7, pp. 2223–2230, July 1995.
[30] T. S. Yang and A. Duel-Hallen, "Adaptive modulation using outdated samples of another fading channel," in *Proc. WCNC'02*, Orlando, FL, Mar. 2002, pp. 477–481.
[31] S. Zhou and G. B. Giannakis, "Adaptive modulation for multi-antenna transmissions with channel mean feedback," *IEEE ICC '03*, vol. 4, pp. 2281–2285, May 11–15, 2003.
[32] ——, "How accurate channel prediction needs to be for adaptive modulation in Rayleigh MIMO channels?," *IEEE Proc. ICASSP '03*, vol. 4, p. IV-81–4, Apr., 6–10.

Xiaodong Cai (M'01) received the B.S. degree from Zhejiang University, Zhejiang, China, the M.Eng. degree from the National University of Singapore, Singapore, and the Ph.D. degree from the New Jersey Institute of Technology, Newark, NJ, in 2001, all in electrical engineering.

From February 2001 to June 2001, he was a member of technical staff at Lucent Technologies, NJ, working on WCDMA project. From July 2001 to October 2001, he was a Senior System Engineer at Sony Technology Center in San Diego, CA, where he was involved in developing high data rate wireless modem. Since November 2001, he has been a Postdoctoral Research Associate in the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN. His research interests lie in the areas of communication theory, signal processing, and networking. Current research focuses on cooperative communications and collaborative signal processing for wireless ad hoc and sensor networks.

Georgios B. Giannakis (S'84–M'86–SM'91–F'97) received the Diploma degree in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1981. He received the M.Sc. degree in electrical engineering, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering from the University of Southern California (USC), Los Angeles, in 1983, 1986, and 1986, respectively.

After lecturing for one year at USC, he joined the University of Virginia, Charlottesville, in 1987, where he became a Professor of Electrical Engineering, in 1997. Since 1999, he has been a Professor with the Department of Electrical and Computer Engineering at the University of Minnesota, Minneapolis, where he now holds an ADC Chair in Wireless Telecommunications. His general interests span the areas of communications and signal processing, estimation and detection theory, time-series analysis, and system identification; subjects on which he has published more than 180 journal papers, 340 conference papers, and two edited books. His current research focuses on transmitter and receiver diversity techniques for single- and multiuser fading communication channels, complex-field and space–time coding, multicarrier, ultra-wide band wireless communication systems, cross-layer designs, and distributed sensor networks.

Dr. Giannakis is the corecipient of four best paper awards from the IEEE Signal Processing Society (1992, 1998, 2000, 2001). He also received the Society's Technical Achievement Award in 2000. He served as Editor in Chief for the IEEE SIGNAL PROCESSING LETTERS, as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS, as Secretary of the Signal Processing Conference Board, as Member of the Signal Processing Publications Board, as Member and Vice-Chair of the Statistical Signal and Array Processing Technical Committee, as Chair of the Signal Processing for Communications Technical Committee, and as a Member of the IEEE Fellows Election Committee. He is currently a Member of the IEEE Signal Processing Society Board of Governors, the Editorial Board for the PROCEEDINGS OF THE IEEE, and Chairs the steering committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.