Adaptive Modulation with Adaptive Pilot Symbol Assisted Estimation and Prediction of Rapidly Fading Channels

Xiaodong Cai and Georgios B. Giannakis[1]
Dept. of Electrical and Computer Engineering
University of Minnesota
200 Union Street SE
Minneapolis, MN 55455
e-mail:{caixd,georgios}@ece.umn.edu

Abstract — **Adaptive modulation requires channel state information (CSI) at both transmitter and receiver. In practice, CSI can be acquired at the receiver by inserting pilot symbols in the transmitted signal. In this paper, we first analyze the effect 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 the 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.**

I. Introduction

Channel-adaptive modulation is a promising technique to enhance spectral efficiency of wireless transmissions over fading channels [3, 8, 11]. 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 non-adaptive transmissions [5, 8]. 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 [1, 8]. For example, an adaptive transmitter relying on an error-free channel estimate to predict future channel values requires feedback delay $\tau < 0.01/f_d$, where f_d is the Doppler frequency [1]. The impact of channel prediction error on the performance of adaptive coded modulation was studied in [10]. While it is necessary to employ reliable channel estimators and predictors to minimize the effect of *imperfect* CSI, adaptive transmitters that account for CSI errors explicitly may have better performance and flexibility when operating in changing environments. Adaptive modulation for multi-antenna systems with channel mean feedback was designed in [12], where it was shown that multi-antennae transmissions increase spectral efficiency, as well as relax the requirement on the channel prediction quality. While the effect of channel prediction error has been considered in [10] and [12], perfect CSI is 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 [4, 9], pilot symbols are periodically inserted in the transmitted signal to facilitate channel estimation and prediction at the receiver. Different from these *non-adaptive* PSAM schemes, we will adjust transmission parameters to maximize spectral efficiency while adhering to a prescribed (target) bit error rate (BER). Our goals are: i) to design *adaptive* PSAM schemes that take into account both channel estimation and prediction errors to meet the target BER; and ii) to optimize the spacing and power of pilot symbols to maximize spectral efficiency.

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, Hermitian, respectively; $E[\cdot]$ denotes expectation with the random variable within the brackets; and $\lfloor x \rfloor$ 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 $\boldsymbol{\mathcal{D}}(\mathbf{x})$ stands for the diagonal matrix with $\mathbf{x}$ on its diagonal.

II. System Model and Bit Error Rate

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

[1]This work was supported by the ARL/CTA grant no. DAAD19-01-2-011.

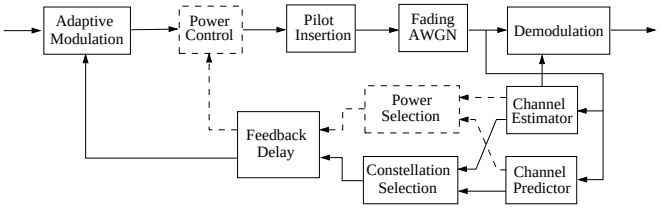

Figure 1: Adaptive PSAM System Model

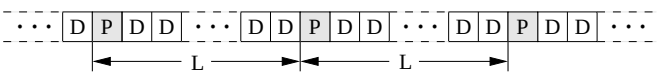

Figure 2: Transmitted frame structure

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:

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

where $\mathcal{E}_p$ and $\mathcal{E}_d$ represent, respectively, the power per pilot and data symbol; $s_p(n)$ is the pilot symbol and $\{s(n;l)\}_{l=1}^{L-1}$ are the data symbols 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 random process with zero-mean, and variance $\sigma_h^2 = 1$. Hence, the average signal-to-noise ratio (SNR) per pilot (data) symbol is $\mathcal{E}_p/N_0$ ($\mathcal{E}_d/N_0$). The autocorrelation function of the channel is determined by the Doppler spectrum. Unlike the PSAM in [4], where pilot and data symbols have the same transmit power, and similar to [9], we allow the pilot and data symbols to be transmitted with different power. The PSAM in [4] uses pilot symbols to enable coherent demodulation; and the ergodic capacity of time-selective fading channels with PSAM is maximized in [9], 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 transmitted 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 $\sigma_\epsilon^2(l) := E[|\epsilon(n;l)|^2]$. Given $\sigma_\epsilon^2(l)$ and a realization of the channel estimator $\hat{h}(n;l) = \hat{h}_0(n;l)$, our goal in this subsection is to derive the conditional BER $P(e|\hat{h}_0(n;l))$ for 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 [4] for the detailed derivation. This estimator uses 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 $\mathbf{s} := [s_p(n - \lfloor K_e/2 \rfloor), \ldots, s_p(n + \lfloor (K_e-1)/2 \rfloor)]^T$, $\mathbf{h} := [h[(n - \lfloor K_e/2 \rfloor); 0], \ldots, h[(n + \lfloor (K_e-1)/2 \rfloor); 0]]^T$, $\mathbf{R} := E[\mathbf{h}\mathbf{h}^{\mathcal{H}}]$, $\mathbf{r}_l := E[\mathbf{h}h^*(n;l)]$, the linear MMSE channel estimator for $h(n;l)$ is given by $\mathbf{w}_l = \sqrt{\mathcal{E}_p}\left(\mathcal{E}_p \boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{R}\boldsymbol{\mathcal{D}}^*(\mathbf{s}) + N_0 \mathbf{I}_{K_e}\right)^{-1} \boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{r}_l$ [4], which does not depend on n. The estimated channel is obtained as $\hat{h}(n;l) = \mathbf{w}_l^{\mathcal{H}}\mathbf{y}(n)$, 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 [4]

$$\sigma_\epsilon^2(l) = 1 - \sqrt{\mathcal{E}_p}\mathbf{r}_l^{\mathcal{H}} \boldsymbol{\mathcal{D}}^*(\mathbf{s})\mathbf{w}_l, \qquad (3)$$

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

We are interested in the BER of square M-QAM and BPSK, given $\hat{h}(n;l) = \hat{h}_0(n;l)$. 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 $\hat{h}(n;l)$, given $\hat{h}(n;l) = \hat{h}_0(n;l)$, the actual channel $\breve{h}(n;l) = \hat{h}_0(n;l) + \epsilon(n;l)$, 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

$$z(n;l) = \frac{\sqrt{\mathcal{E}_d}\breve{h}(n;l)s + \eta}{\sqrt{\mathcal{E}_d}\hat{h}_0} = s + \frac{s\epsilon}{\hat{h}_0} + \frac{\eta}{\sqrt{\mathcal{E}_d}\,\hat{h}_0}, \qquad (4)$$

where we omitted the time indices 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 of (4) can be expressed as

$$\gamma_1(n;l) = \frac{|s|^2\mathcal{E}_d|\hat{h}_0(n;l)|^2}{N_0 + \mathcal{E}_d\sigma_\epsilon^2(l)|s|^2}. \qquad (5)$$

For BPSK and 4-QAM, we have $|s| = 1$; and thus, BER can be calculated using the SNR in (5). For M-QAM with $M > 4$, we have $|s| \neq 1$; and the SNR in (5) can not be directly used to calculate BER. However, since $E[|s|^2] = 1$ and $|s|^2 < 2.65$ for

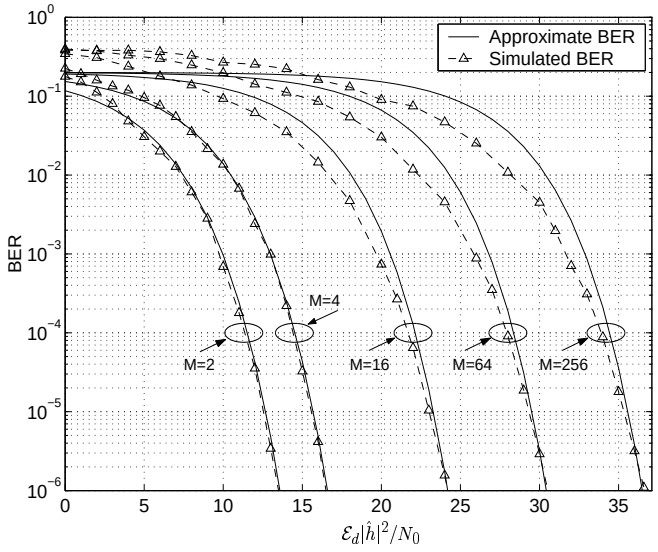

Figure 3: Conditional BER, $\sigma_\epsilon^2(l) = N_0/\mathcal{E}_d = 0.01$

$M \leq 256$, 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

$$\gamma_2(n;l) = \frac{\mathcal{E}_d|\hat{h}_0(n;l)|^2}{N_0 + 1.3\,\mathcal{E}_d\sigma_\epsilon^2(l)}. \tag{6}$$

Considering both BPSK and M-QAM, we will henceforth use the SNR

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

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 [5]

$$P(e_i|\hat{h}_0(n;l)) \approx 0.2\exp\Big(-\frac{c\gamma(n;l)}{M_i - 1}\Big), \tag{8}$$

where $\gamma(n;l)$ is given in (7), $c = 3.2/3$ for BPSK, and $c = 1.6$ for square M-QAM. Fig. 3 compares the simulated BER with the approximate BER calculated from (8), when $\sigma_\epsilon^2(l) = N_0/\mathcal{E}_d = 0.01$. The approximation is within 2dB for M-QAM, when the target BER $< 10^{-2}$.

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, and we assume that the delay is a multiple of LT for notational simplicity. 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 $\mathbf{s}_p := [s_p(n-D),\ldots,s_p(n-D-K_p+1)]^T$, $\mathbf{h}_p := [h(n-D;0),\ldots,h(n-D-K_p+1;0)]^T$, $\mathbf{R}_{h_p} := E[\mathbf{h}_p\mathbf{h}_p^{\mathcal{H}}]$, $\mathbf{r}_p(l) = E[\mathbf{h}_p h^*(n;l)]$, the MMSE of the predicted channel is written as $\sigma_\varepsilon^2(l) = 1 - \mathcal{E}_p\mathbf{r}_p^{\mathcal{H}}(l)\boldsymbol{\mathcal{D}}^*(\mathbf{s}_p)\big(\mathcal{E}_p\boldsymbol{\mathcal{D}}(\mathbf{s}_p)\mathbf{R}_{h_p}\boldsymbol{\mathcal{D}}^*(\mathbf{s}_p) + N_0\mathbf{I}_{K_p}\big)^{-1}\boldsymbol{\mathcal{D}}(\mathbf{s}_p)\mathbf{r}_p(l)$, where similar to $\sigma_\epsilon^2(l)$, $\sigma_\varepsilon^2(l)$ does not depend on n. As shown in [10], the length of the channel predictor, K_p, should be chosen large enough to achieve a sufficiently small MSE with a moderate delay. Similar to the estimated channel $\hat{h}(n;l)$, the predicted channel $\check{h}(n;l)$ is Gaussian distributed with zero-mean, and variance $\sigma_{\check{h}}^2(l) = \sigma_h^2 - \sigma_\varepsilon^2(l) = 1 - \sigma_\varepsilon^2(l)$.

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

$$\begin{aligned}\hat{h}(n;l) &= h(n;l) - \epsilon(n;l)\\ &= \check{h}(n;l) + \varepsilon(n;l) - \epsilon(n;l).\end{aligned} \tag{9}$$

By the orthogonality principle, $\varepsilon(n;l)$ and $\check{h}(n;l)$ are uncorrelated. The correlation between $\epsilon(n;l)$ and $\check{h}(n;l)$ is $E[\epsilon(n;l)\check{h}^*(n;l)] = E[\epsilon(n;l)(h^*(n;l) - \varepsilon^*(n;l))] = \sigma_\epsilon^2(l) - E[\epsilon(n;l)\varepsilon^*(n;l)]$. Since $|E[\epsilon(n;l)\varepsilon^*(n;l)]| \leq \sigma_\epsilon(l)\sigma_\varepsilon(l)$, we have

$$\frac{\sigma_\epsilon(l) - \sigma_\varepsilon(l)}{\sigma_{\check{h}}(l)} \leq \frac{E[\epsilon(n;l)\check{h}^*(n;l)]}{\sigma_\epsilon(l)\sigma_{\check{h}}(l)} \leq \frac{\sigma_\epsilon(l) + \sigma_\varepsilon(l)}{\sigma_{\check{h}}(l)}. \tag{10}$$

In practice, $\sigma_\epsilon(l)$ and $\sigma_\varepsilon(l)$ are very small. And because $\sigma_{\check{h}}(l)$ is close to one, (10) implies that the correlation coefficient between $\epsilon(n;l)$ and $\check{h}(n;l)$ is very small. Hence, we can assume that $\epsilon(n;l)$ and $\check{h}(n;l)$ are uncorrelated. Given $\check{h}(n;l) = \check{h}_0(n;l)$, $\hat{h}(n;l)$ is Gaussian distributed with mean $\check{h}_0(n;l)$. Since $\check{h}(n;l)$ is uncorrelated with $\varepsilon(n;l)$ and $\epsilon(n;l)$, the conditional variance of $\hat{h}(n;l)$ is $\tilde{\sigma}_{\check{h}}^2 := E[|\varepsilon(n;l) - \epsilon(n;l)|^2]$. Based on the definition of $\varepsilon(n;l)$ and $\epsilon(n;l)$, $\tilde{\sigma}_{\check{h}}^2$ can be calculated using the channel autocorrelation function. However, since $(\sigma_\epsilon(l) - \sigma_\varepsilon(l))^2 \leq \tilde{\sigma}_{\check{h}}^2 \leq (\sigma_\epsilon(l) + \sigma_\varepsilon(l))^2$, we will use $\tilde{\sigma}_{\check{h}}^2 = \sigma_\epsilon^2(l) + \sigma_\varepsilon^2(l)$ in calculating the BER, which amounts to assuming that $\epsilon(n;l)$ and $\varepsilon(n;l)$ are uncorrelated.

Given $\check{h}(n;l) = \check{h}_0(n;l)$, the amplitude of $\hat{h}(n;l)$ follows a Rice distribution with Rician factor $K = |\check{h}_0(n;l)|^2/\tilde{\sigma}_{\check{h}}^2$. 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

$$P(e_i|\check{h}_0) = \int_0^\infty P(e_i||\hat{h}_0|)p(|\hat{h}_0||\check{h}_0)d|\hat{h}_0|, \tag{11}$$

where we omitted the time indices of $\hat{h}_0(n;l)$ and $\check{h}_0(n;l)$. From (8), it is seen that $p(e_i\big||\hat{h}_0(n;l)|) = p(e_i|\hat{h}_0(n;l))$; thus, we substitute (8) into (11), and obtain [6]

$$P(e_i|\check{h}_0(n;l)) \approx \frac{0.2\exp\big(-\frac{a_i(l)\mathcal{E}_d|\check{h}_0(n;l)|^2}{a_i(l)\mathcal{E}_d\tilde{\sigma}_{\check{h}}^2+1}\big)}{a_i(l)\mathcal{E}_d\tilde{\sigma}_{\check{h}}^2 + 1}, \tag{12}$$

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 it is seen from (8) and (12) that $P(e_i|\check{h}_0(n;l)) = P(e_i|h_0(n;l))$, as expected. The BER $P(e_i|\check{h}_0(n;l))$ is mainly determined by the exponential term in (12). To see how the channel prediction and estimation errors affect the BER, we express the exponent in (12)

as

$$\frac{a_i(l)\mathcal{E}_d|\check{h}_0(n;l)|^2}{a_i(l)\mathcal{E}_d\tilde{\sigma}^2_{\hat{h}}+1}=\frac{c\mathcal{E}_d|\check{h}_0(n;l)|^2}{(M_i-1)N_0}\times G, \tag{13}$$

where

$$G=\frac{N_0}{N_0+\mathcal{E}_d\sigma^2_\epsilon(l)(g+c/(M_i-1))+c\mathcal{E}_d\sigma^2_\varepsilon(l)/(M_i-1)} \tag{14}$$

reflects the performance loss caused by channel estimation and prediction. From (14), we observe that the same channel prediction error incurs less performance loss for large constellation sizes, because the term $c\mathcal{E}_d\sigma^2_\varepsilon(l)/(M_i-1)$ reduces as M_i increases. On the other hand, the same channel estimation error causes almost the same performance loss for all M-QAM and BPSK constellations, since the term $\mathcal{E}_d\sigma^2_\epsilon(l)(g+c/(M_i-1))$ changes very slowly when M_i increases beyond a moderately large value.

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 $d_i(l)=(a_i(l)\mathcal{E}_dN_0)/(\bar{\mathcal{E}}_db_i(l))$, we can also write $P(e_i|\check{h}_0(n;l))$ in (12) as

$$P(e_i|\check{\gamma}(l))\approx\frac{0.2\exp(-d_i(l)\check{\gamma}(l))}{b_i(l)}, \tag{15}$$

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

$$p(\check{\gamma}(l))=\frac{\exp(-\check{\gamma}(l)/\bar{\check{\gamma}}(l))}{\bar{\check{\gamma}}(l)}. \tag{16}$$

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 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 $\sigma^2_\epsilon(l)$ in (3) as

$$\sigma^2_\epsilon(l)=1-\sum_{i=1}^{K_e}\frac{\mathcal{E}_p|\mathbf{u}_i^{\mathcal{H}}\mathbf{r}_l|^2}{\mathcal{E}_p\lambda_i+N_0}=1-\sum_{i=1}^{K_e}\frac{|\mathbf{u}_i^{\mathcal{H}}\mathbf{r}_l|^2(1-\alpha)L\bar{\gamma}}{(1-\alpha)L\bar{\gamma}\lambda_i+1}, \tag{17}$$

where $\bar{\gamma}:=\mathcal{E}/N_0$ is the average SNR. Since $\mathbf{R}$ depends on L, so does $\sigma^2_\epsilon(l)$. The MSE of the predicted channel, $\sigma^2_\varepsilon(l)$, 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 transmitted power 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 (15) 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 showed in [2], when $K_e\geq 20$, 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 $l=L-1$. Hence, the receiver can use $\sigma_\epsilon(L-1)$ and $\sigma_\varepsilon(L-1)$ in calculating the switching thresholds 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 indices 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 switching threshold for the ith constellation, which 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 [5]

$$\mathcal{E}_d=\frac{\bar{\mathcal{E}}_d}{\int_{\check{\gamma}_1}^{\infty}p(\check{\gamma})d\check{\gamma}}=\bar{\mathcal{E}}_d\exp(\check{\gamma}_1/\bar{\check{\gamma}}). \tag{18}$$

Suppose that the target BER is equal to B. Using (15) and letting $P(e|\check{\gamma}_i)=B$, we can find the switching thresholds

$$\check{\gamma}_i=-\frac{\ln(5b_iB)}{d_i},\quad i=1,\ldots,N. \tag{19}$$

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

$$\begin{aligned}S&=(1-\frac{1}{L})\sum_{i=1}^{N}k_iP(k_i)\\&=(1-\frac{1}{L})[k_1\exp(-\frac{\check{\gamma}_1}{\bar{\check{\gamma}}})+\sum_{i=2}^{N}(k_i-k_{i-1})\exp(-\frac{\check{\gamma}_i}{\bar{\check{\gamma}}})].\end{aligned} \tag{20}$$

It is evident from (17) 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 $L_{max}=\lfloor 1/(2f_dT)\rfloor$ [4, 9]. For each $L\in[2,L_{max}]$, we have the following optimization problem to be solved for the optimum α and $\mathcal{E}_d$:

$$\begin{aligned}\max_{\alpha,\mathcal{E}_d}\quad & S(\alpha,\mathcal{E}_d)\\ \text{subject to}\quad & \begin{cases}0<\alpha<1\\ (L-1)\mathcal{E}_d=\alpha L\mathcal{E}\exp(-\frac{\ln(5b_1B)}{d_1\bar{\check{\gamma}}}),\end{cases}\end{aligned} \tag{21}$$

where the last constraint follows from (18), and we explicitly write S as a function of α, and $\mathcal{E}_d$. Numerical search based on e.g., sequential quadratic programming (SQP) [7], can be used to solve this optimization problem with starting point $\alpha=(L-1)/L$, $\mathcal{E}_d=\bar{\mathcal{E}}_d$. After solving (21) 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 2dB of those with continuous power adaptation [8], and is higher than that of constant power transmission [5]. Here, the advantage of discrete power adaptation relative to constant power transmission may be even larger, because discrete power adaptation

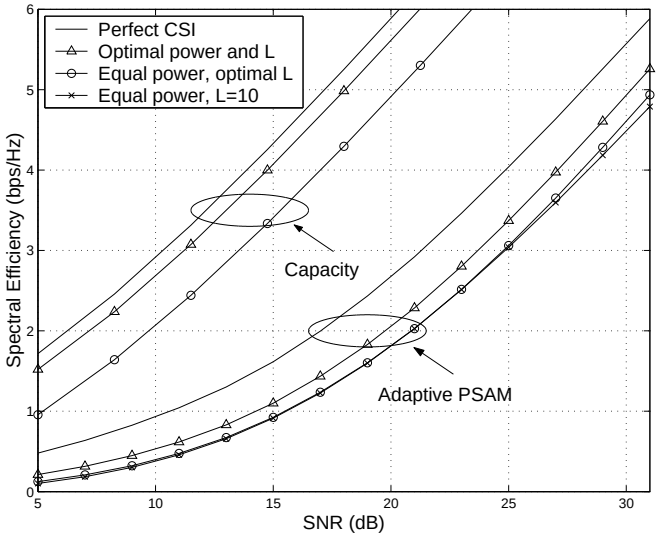

Figure 4: Spectral efficiency of constant power adaptive modulation

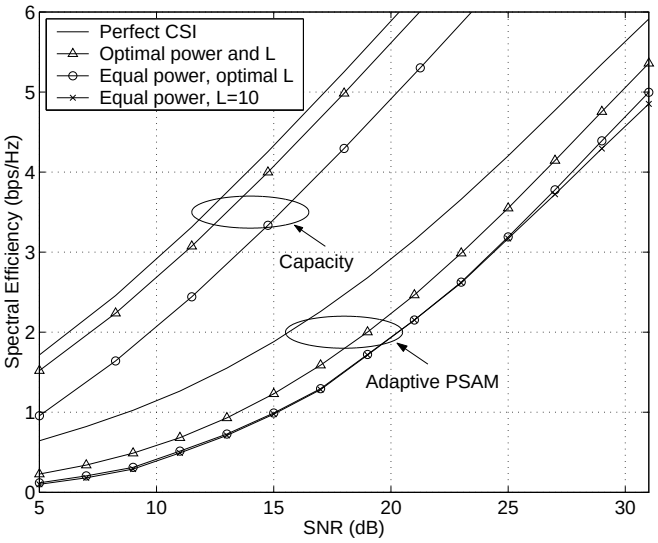

Figure 5: Spectral efficiency of discrete power adaptive modulation

increases the probability of using large constellations, and as corroborated by (14), 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 (19) 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

$$\sum_{i=1}^{N} \mathcal{E}_{di} P(k_i) = \bar{\mathcal{E}}_d. \tag{22}$$

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

$$\begin{aligned} &\max_{\alpha,\{\mathcal{E}_{di}\}_{i=1}^{N}} \quad S(\alpha, \mathcal{E}_{d1}, \ldots \mathcal{E}_{dN}) \\ &\text{subject to} \begin{cases} 0 < \alpha < 1 \\ \mathcal{E}_{di} > 0, \ i = 1, \ldots, N \\ \sum_{i=1}^{N} \mathcal{E}_{di}[\exp(-\check{\gamma}_i/\bar{\bar{\gamma}}) - \exp(-\check{\gamma}_{i+1}/\bar{\bar{\gamma}})] \\ = \alpha L \mathcal{E}/(L-1), \end{cases} \end{aligned} \tag{23}$$

where the last equation in the constraints follows from (22), 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., sequential quadratic programming (SQP) [7], 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 (23).

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, 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

$$P(e_i) = \int_{\check{\gamma}_i}^{\check{\gamma}_{i+1}} P(e_i|\check{\gamma}) p(\check{\gamma}) d\check{\gamma}. \tag{24}$$

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 [1, 5]

$$P(e) = \frac{\sum_{i=1}^{N} k_i P(e_i)}{\sum_{i=1}^{N} k_i P(k_i)}. \tag{25}$$

Based on the conditional BER in (15) and the pdf of the predicted SNR in (16), we can approximately calculate $P(e_i)$ from (24) as

$$P(e_i) \approx \frac{0.2}{b_i \beta_i} \left[\exp\left(-\frac{\beta_i \check{\gamma}_i}{\bar{\bar{\gamma}}}\right) - \exp\left(-\frac{\beta_i \check{\gamma}_{i+1}}{\bar{\bar{\gamma}}}\right) \right], \tag{26}$$

where $\beta_i := \bar{\bar{\gamma}} d_i + 1$. Then, the average BER in (25) can be approximately calculated by substituting (26) into (25).

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 mobile velocity is $v = 108$km/hr. This results in a Doppler spread $f_d = 200$Hz. We assume that the channel has a Jakes' Doppler spectrum. The symbol rate is 200ksps, which corresponds to a symbol period of $5\mu s$. The normalized Doppler spread is thus $f_d T = 10^{-3}$. We consider a delay $\tau = 0.2/f_d$, and choose 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 depicted in Fig. 4. It is seen that the scheme with optimum power allocation has about 1dB advantage over the equal power scheme at SNR = 15dB. We also show the channel capacity with perfect CSI, as well as lower bounds on channel capacity with estimated channels, which were derived in [9]. With optimum training, channel capacity is very close to that with perfect

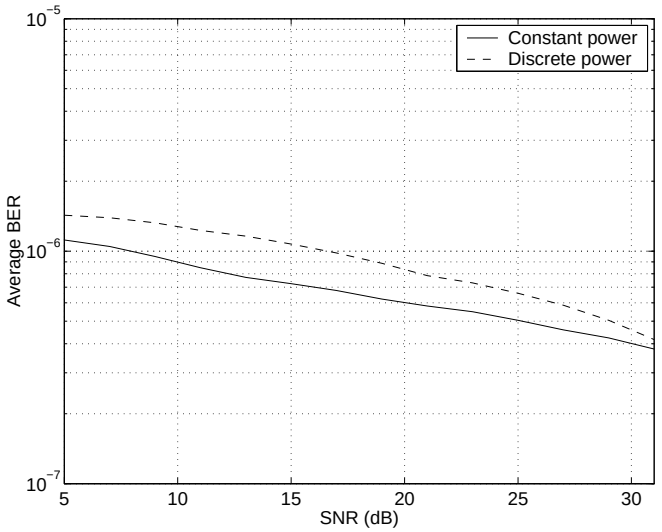

Figure 6: Average BER

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 spectral efficiency compared to that with perfect CSI. Fig. 5 shows the spectral efficiency of discrete power adaptive PSAM. Comparing Fig. 4 with Fig. 5, we infer that the discrete power adaptive PSAM has larger spectral efficiency than the constant power scheme, as expected. From Fig. 5, we see that the optimum power allocation has higher spectral efficiency than the equal power allocation scheme. Fig. 6 shows the average BER of both constant and discrete power adaptive PSAM. 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, the adaptive modulation schemes treating the predicted channel as perfect CSI may not be 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. Conclusions

We have studied adaptive transmission systems with pilot symbol assisted estimation and prediction of rapidly fading channels. The effect of 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.

References

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

[2] X. Cai and G. B. Giannakis, "Adaptive PSAM accounting for channel estimation and prediction errors," *IEEE Trans. Wireless Commun.*, submitted, Feb. 2003.

[3] J. K. Cavers, "Variable-rate transmission for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 20, pp. 15–22, Feb. 1972.

[4] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 686–693, Nov. 1991.

[5] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *IEEE Trans. Commun.*, vol. 49, pp. 1561–1571, Sept. 2001.

[6] D. Divsalar and M. K. Simon, "The design of trellis coded MPSK for fading channels: performance criteria," *IEEE Trans. Commun.*, vol. 36, pp. 1004–1012, Sept. 1988.

[7] R. Fletcher, *Practical Methods of Optimization, Vol. 2, Constrained Optimization*. John Wiley and Sons, 1980.

[8] A. J. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 1218–1230, Oct. 1997.

[9] S. Ohno and G. B. Giannakis, "Average-rate optimal PSAM transmissions over time-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 712–720, Oct. 2002.

[10] G. E. Øien, H. Holm, and K. J. Hole, "Impact of imperfect channel prediction on adaptive coded modulation performance," *IEEE Trans. Veh. Technol.*, submitted June 2002; see also *Proc. of Int. Symp. Advances in Wireless Commun.*, Victoria, BC Canada, Sept. 2002, pp. 19–20.

[11] W. T. Webb and R. Steele, "Variable rate QAM for mobile radio," *IEEE Trans. Commun.*, vol. 43, pp. 2223–2230, July 1995.

[12] S. Zhou and G. B. Giannakis, "Adaptive modulation for multi-antenna transmissions with channel mean feedback," in *Proc. of ICC*, Anchorage, Alaska, May 2003.