

How Much Training is Optimal in Adaptive PSAM Over Markov Rayleigh Fading Channels

Khalid Zeineddine, Ibrahim Abou-Faycal

Dept. of Elec. and Comp. Eng., American University of Beirut
Beirut, Lebanon

kaz06@aub.edu.lb, iaf@alum.mit.edu

Abstract—To mitigate the effects of time-varying fading, Pilot Symbol Assisted Modulation (PSAM) has been introduced in [1], through which the transmitter periodically inserts known symbols or pilots in the data frame. The channel estimates obtained from these pilots are then employed in the coherent demodulation of data symbols. A novel PSAM scheme was introduced in [2] that adapts the transmitter’s coded modulation strategy to the quality of the channel estimate at the receiver without requiring any channel-feedback from the latter. In our work, we study the performance of this non-feedback adaptive PSAM scheme using clustering techniques. More precisely, instead of using one pilot per data frame, we send a cluster of pilots consecutively in the beginning of every data transmission interval. The performance of this scheme is measured in terms of achievable rates using binary signaling and modeling the fading process as a Gauss-Markov process. We show through numerical computations that our new strategy provides higher achievable rates at certain levels of SNR and fading correlation, and we provide numerical answers to how much training is “optimal” for the strategy.

Finally, we show how the methodology may be extended to high-order Gauss-Markov fading models, with which pilot clustering is expected to yield even higher benefits.

Keywords—Adaptive modulation, fading channels, Gauss-Markov, Kalman filtering, pilot symbol assisted modulation, Rayleigh fading, time-varying channels.

I. INTRODUCTION

Rapid fading is a central problem in digital mobile communications that degrades the Bit Error Rate (BER) and frequently introduces an error floor. To mitigate the effects of fading, Pilot Symbol Assisted Modulation (PSAM) has been introduced in [1], through which the transmitter periodically inserts known symbols or pilots in the data sequence or data frame. Channel estimates of the Channel State Information (CSI) obtained from these pilots are then employed in the coherent demodulation¹ of the data symbols. The pilots are multiplexed with data symbols at a constant ratio which is considered as a design parameter. Guidelines for the choice of this ratio was studied in [3].

Many modulation and coding techniques do not adapt to fading conditions. These nonadaptive methods require a fixed link margin to maintain acceptable performance when the channel quality is poor. Therefore these systems are effectively designed for the worst case channel conditions, resulting in insufficient utilization of the full channel capacity.

Channel-adaptive PSAM is a promising technique to enhance spectral efficiency of wireless transmissions over fading channels. In adaptive systems, a certain transmission parameter such as constellation size, transmitted power, or code rate is dynamically adjusted according to the channel quality. This increases the average spectral efficiency without wasting power or sacrificing error probability performance. It can provide higher bit rates relative to conventional signaling by transmitting at high rate under favorable channel conditions, and reducing throughput as the channel condition degrades. It is worth mentioning that such adaptive modulation and coding schemes are used in wireless standards such as the Enhanced Data GSM Evolution (EDGE).

Most of the adaptive schemes assume a parametrized channel-feedback link (parameters such as delay) that supplies the transmitter with the estimated channel side information [4], [5]. One disadvantage a feedback channel might induce in addition to the difficulty of realistic mathematical modeling, is a practical constraint when the channel is changing faster than it can be estimated and then fed back to the transmitter. In [2], the authors improved PSAM schemes over Rayleigh fading channels by adapting the coded modulation strategy at the sender to the quality of the channel estimation (estimation error variance) at the receiver, without requiring any channel feedback. In our work, we study the performance of this non-feedback adaptive PSAM scheme using clustering techniques. Instead of the regularly spaced single pilot scheme adopted in [2], we send a cluster of k pilots consecutively per data transmission interval as shown in Figure 1. The performance of our scheme is measured in terms of achievable rates using binary signaling and modeling the fading process as a Gauss-Markov process. Through numerical computations we find the “optimal” duration of the training period.

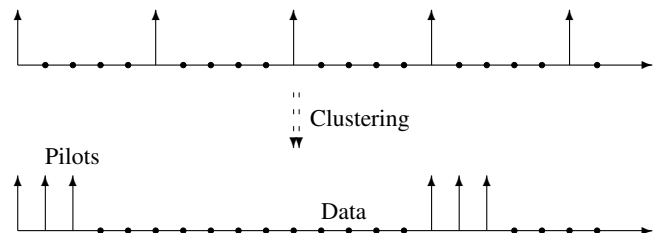


Fig. 1. Clustering Technique in PSAM

¹Coherent demodulation requires the extraction of a reliable phase reference from the received signal.

In scenarios where the fading process may be accurately modeled through a high-order Gauss-Markov process, we naturally expect the clustering of pilots to be even more beneficial. While we do not present specific results in this paper, we briefly present how the technique may be adjusted to these higher order models.

Finally, while there has been significant progress on understanding the theoretical aspects of time-adaptation in the previously-mentioned schemes, new challenges surface when dynamic transmission techniques are employed in broadband wireless networks with multiple signaling dimensions. Those additional dimensions are mainly frequency in multicarrier systems like Orthogonal Frequency-Division Multiplexing (OFDM), and space in Multiple-Input Multiple-Output communication systems (MIMO). Therefore we need adaptive schemes that integrate temporal, spatial, and spectral components together, i.e., exploit the variation of the wireless channel over time, frequency, and/or space [6], [7], [8]. In future work, we will try to extend our adaptive scheme to MIMO scenarios [9].

The organization of this paper is as follows: In Section II, we present the fading channel model. In Section III we explain the adaptive technique we use to transmit over the channel as well as the receiver details. The measure of performance is discussed in Section IV, and the numerical results are presented in Section V. In Section VI we show how the numerical computations may be improved and how the methodology may be extended to higher-order Gauss-Markov models. We conclude our paper in Section VII and motivate future work on the subject.

II. THE CHANNEL MODEL

Consider the single-user discrete-time model for the Rayleigh fading channel,

$$Y_i = R_i X_i + N_i,$$

where i is the time index, $X_i \in \mathbb{C}$ is the channel input at time i , $Y_i \in \mathbb{C}$ is its output, and R_i and N_i are independent complex circular Gaussian² random variables with mean zero and variance σ_R^2 and σ_N^2 respectively. The amplitude of the fading coefficient R_i is hence Rayleigh distributed and its phase is uniform over $[-\pi, \pi)$. To account for power constraints, the input is subject to

$$\mathbb{E} \left[|X_i|^2 \right] \leq P.$$

We assume that the fading process is a stationary first order Gauss-Markov process, i.e.,

$$R_i = \alpha R_{i-1} + Z_i, \quad (1)$$

where the samples $\{Z_i\}$ are Independent and Identically-Distributed (IID) complex circular Gaussian with mean zero

²A complex Gaussian random variable is circular if and only if it is zero-mean and its real and imaginary parts are independent with equal variances.

and variance equal to $\sigma_Z^2 = (1 - \alpha^2) \sigma_R^2$ such that $\alpha \in [0, 1)$ in order to guarantee stationarity.

This channel model was adopted in [10], where a relation between the parameter α and the coherence time was established as

$$\alpha^{T_c W} = \phi,$$

where T_c is the coherence time, W is the transmission bandwidth, and ϕ is the level of decorrelation. For bandwidths in the 10 kHz range, and Doppler spreads in the order of 100 Hz, α will typically range between 0.9 and 0.99. For instance, for a Doppler spread of 100 Hz, $W = 10^4$, and $\phi = 0.1$, we have $\alpha = 0.977$.

III. THE ADAPTIVE TRANSMISSION SCHEME

Our transmitter employs multiple codebooks that are interleaved with the data symbols being coded according to their distance from the pilot symbols. Symbols that are far away from the pilot symbols encounter poorer channel measurements at the receiver and are therefore coded with lower rate codes, while symbols that are close to the pilot symbols benefit from recent channel measurements and are coded with higher rate codes.

While the scheme is adaptive, note that the transmitter *does not use feedback* to determine its policy but instead takes into account the time-varying quality of the channel measurement available at the receiver to modify this policy. The key idea is that the transmitter adapts to the quality of channel estimation (the resulting mean-square error specifically) rather than to the quality of the channel. The need for information at the transmitter, which is supplied by perfect feedback in some schemes in the literature, is compensated by the information acquired through the knowledge of the estimation strategy adopted by the receiver. The need to know the channel state is compensated by the knowledge of the statistics of the estimation error, which can be computed offline.

The transmission model is as follows: at regular intervals, the transmitter sends k successive, constant and known pilot symbols whose purpose is to enable the estimation of the channel at the receiver. The pilot symbols have energy equal to the average energy constraint. This model is an extension to the one adopted in [2], where a single pilot is transmitted for estimation followed by m data symbols. In our modified scheme (shown in Figure 2), k consecutive pilots are transmitted in the beginning of every data frame.

At the receiver, the channel estimation is assumed to be based only on the pilot symbols and no data-directed estimation is used. For each time sample i , the receiver computes the Minimum Mean-Square Estimate (MMSE) of the channel, the quality of which depends on its position with respect to the pilot symbols. After estimation, the channel, as seen by the receiver, is composed of a specular part (the estimate) and a zero-mean Gaussian distributed error. Therefore, the channel seen by the receiver is a Rician channel in which the specular

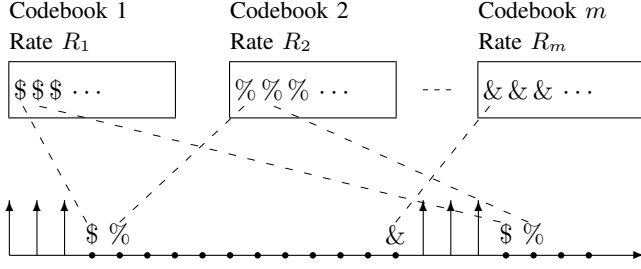


Fig. 2. Multiple Codebook Interleaving

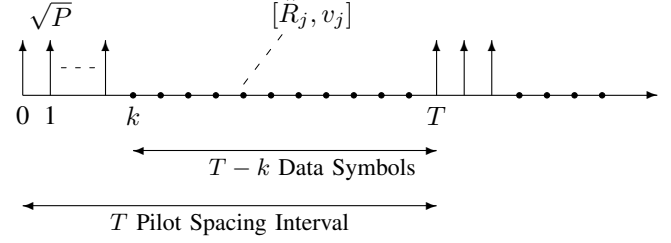


Fig. 3. Pilot Symbols and Channel Estimation

part is given by the estimate and the Rayleigh component by the estimation error.

Since we are primarily interested in the low Signal-to-Noise Ratio (SNR) regime, we only consider binary signaling. The motivation for this choice is multiple folds. First, in [11] the authors prove that for discrete-time memoryless Rayleigh fading channels subject to average power constraints, the capacity achieving distribution is discrete with a finite number of mass points. Moreover, a binary distribution was found to be optimal at low and moderate values of SNR [11], [12], [13]. Second, for a memoryless Rician fading channel, Luo [14] established a similar result that, combined with Gallager's in [12], implies that the binary input distribution is asymptotically optimal at low SNR [14]. Consequently, we choose the alphabet of every codebook to consist in general of two symbols:

$$\begin{cases} m_1 = a_1 + jb_1 & \text{with probability } p_1 \\ m_2 = a_2 + jb_2 & \text{with probability } p_2 = (1 - p_1). \end{cases} \quad (2)$$

We change the rate of the codebooks by modifying the probability distribution of the mass points instead of increasing the power of the codeword, as we adopt a flat power allocation.

The numerical results in [2] indicate that the optimal mass points always lie between the extremes of on-off keying (optimal for the IID Rayleigh fading case) and the antipodal signaling (optimal for a perfectly known channel). Hence the optimal input distribution consists of two non zero masses, the first of which is negative, located between $-\sqrt{P}$ and zero, while the second is positive and greater than \sqrt{P} . Furthermore, since any rotational transformation of the two mass points will not affect the mutual information [11], [14], we restrict ourselves to input distributions with *real* mass points i.e. $m_1, m_2 \in \mathbb{R}$.

A. Channel Estimation at the Receiver

Given a pilot spacing interval of T and given that we will use k pilots ($1 \leq k \leq T - 1$) as shown in Figure 3, the output of the channel during this interval is,

$$Y_i = \sqrt{P}R_i + N_i \quad i = 0, \dots, k - 1.$$

We want to find the causal MMSE estimate $\hat{R}_j(Y_0, \dots, Y_{k-1})$ of R_j for $j = k, \dots, T - 1$. Since $\{R_j, Y_0, \dots, Y_{k-1}\}$ are jointly Gaussian, the MMSE

estimator is linear and is identical to the Linear Least-Square Estimator (LLSE). We shall use a causal Kalman filter to calculate the estimates \hat{R}_j and corresponding error variances v_j in a recursive and efficient manner.

Using the basic results for Kalman filters [15, sec. 13.4], we have

$$\begin{aligned} q_j &= \begin{cases} \frac{\sqrt{P}\sigma_Z^2 + \alpha^2\sqrt{P}v_{j-1}}{P\sigma_Z^2 + \alpha^2Pv_{j-1} + \sigma_N^2} & 0 \leq j \leq k - 1 \\ 0 & k \leq j \leq T - 1 \end{cases} \\ \hat{R}_j &= \alpha\hat{R}_{j-1} + q_j(Y_j - \alpha\sqrt{P}\hat{R}_{j-1}) \\ v_j &= (\alpha^2v_{j-1} + \sigma_Z^2)(1 - \sqrt{P}q_j), \end{aligned} \quad (3)$$

where q_j is the innovation factor after each new observation. We note that these factors and the estimation error variances (3) may be computed offline for all $0 \leq j \leq T - 1$. These computations may be performed by the encoder without requiring feedback, and more importantly, they are actually known at design time.

IV. ACHIEVABLE RATES

Since our system is coded, its performance may be measured through the rates that we can achieve. These rates are readily obtained using information-theoretic tools. More precisely, they are equal to the mutual information between the input and the output.

Let $S = \{0, \dots, k - 1\}$ and let $\{Y_s\}_{s \in S}$ be the received values during training. Due to interleaving, the mutual information can be written as

$$\begin{aligned} &I(X_k, \dots, X_{T-1}; Y_k, \dots, Y_{T-1} | \{Y_s\}_{s \in S}) \\ &= \frac{1}{T} \sum_{i=k}^{T-1} I(X_i; Y_i | \{Y_s\}_{s \in S}) \\ &= \mathbb{E}_{\{Y_s\}_{s \in S}} \left[\frac{1}{T} \sum_{i=k}^{T-1} I(X_i; Y_i | \{y_s\}_{s \in S}) \right]. \end{aligned}$$

Given a sample $\{y_s\}_{s \in S}$ we have,

$$Y_i = R_i X_i + N_i = (\hat{R}_i + \Gamma_i) X_i + N_i, \text{ for } k \leq i \leq T - 1,$$

where Γ_i is a zero-mean complex Gaussian error term that has a variance v_i . Therefore,

$$\begin{aligned} p(y_i|x_i, \{y_s\}_{s \in S}) &= \mathcal{N}_C(\hat{R}_i x_i, v_i |x_i|^2 + \sigma_N^2) \\ &= \frac{1}{\pi (v_i |x_i|^2 + \sigma_N^2)} \exp\left(-\frac{|y_i - \hat{R}_i x_i|^2}{v_i |x_i|^2 + \sigma_N^2}\right). \end{aligned}$$

Using a binary input (2) with two mass points m_1 and m_2 ,

$$\begin{aligned} I(X_i; Y_i | \{y_s\}_{s \in S}) &= \int_{y_i} \left[p_1 p(y_i|m_1, \{y_s\}_{s \in S}) \ln \frac{p(y_i|m_1, \{y_s\}_{s \in S})}{p(y_i | \{y_s\}_{s \in S})} \right. \\ &\quad \left. + p_2 p(y_i|m_2, \{y_s\}_{s \in S}) \ln \frac{p(y_i|m_2, \{y_s\}_{s \in S})}{p(y_i | \{y_s\}_{s \in S})} \right] dy_i. \end{aligned}$$

Finally, the mutual information can be expressed as

$$\begin{aligned} I(X_k, \dots, X_{T-1}; Y_k, \dots, Y_{T-1} | \{Y_s\}_{s \in S}) &= \mathbb{E}_{\{Y_s\}_{s \in S}} \left[\frac{1}{T} \sum_{i=k}^{T-1} I(X_i; Y_i | \{y_s\}_{s \in S}) \right] \\ &= \frac{1}{T} \sum_{i=k}^{T-1} \mathbb{E}_{\hat{R}_i} \left[I(X_i; Y_i | \hat{R}_i) \right], \end{aligned} \quad (4)$$

where now the expectation is over the random variable \hat{R}_i . We note that \hat{R}_i is a linear combination of the observations

$$\hat{R}_i = \sum_{m=0}^{i-1} \beta_m Y_m = R_i - \Gamma_i \sim \mathcal{N}_C(0, \sigma_R^2 - v_i).$$

The i^{th} term, $\mathbb{E}_{\hat{R}_i} [I(X_i; Y_i | \hat{R}_i)]$, in (4) depends on the choice of corresponding binary probability distribution fully characterized by the three parameters $\{m_1, m_2, p_1\}_i$. Therefore the input distribution $\{m_1, m_2, p_1\}_i$ for all $k \leq i \leq T-1$ is determined by solving the following optimization problem

$$\frac{1}{T} \sum_{i=k}^{T-1} \max_{\{m_1, m_2, p_1\}_i} \mathbb{E}_{\hat{R}_i} [I(X_i; Y_i | \hat{R}_i)] \quad (5)$$

subject to $\mathbb{E}[|X_i|^2] \leq P$, for all $k \leq i \leq T-1$.

V. NUMERICAL RESULTS

The numerical results confirm the observation previously made. Namely, the achievable rate in (4) depends on the choice of $\{m_1, m_2, p_1\}_i$ which determines the input distribution of the i -th symbol. As the symbol gets further away from the pilots, the channel estimation quality is degraded and hence the amount of information that may be sent over the channel decreases. Therefore, as i increases, $\{m_1, m_2, p_1\}_i$ shifts from the antipodal distribution (optimal for a perfectly known channel) with $p_1 \simeq p_2$ and migrates toward the other extreme of on-off keying (optimal for the IID Rayleigh fading case) with p_1 much greater than p_2 .

Figures 4, 5, and 6 present the achievable rates (in nats per channel use) of the strategy we follow for multiple lengths of training periods. We study the performance of 1, 2, 3, 4, 5, and 6 clustered pilot schemes and the comparison is done for various values of SNR and α .

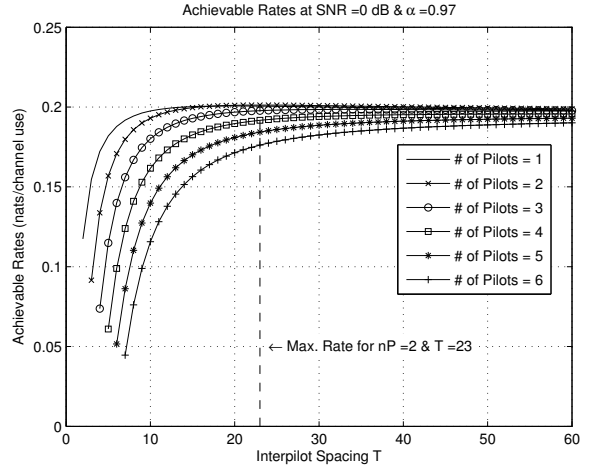


Fig. 4. Achievable Rates for SNR = 0dB and $\alpha = 0.97$.

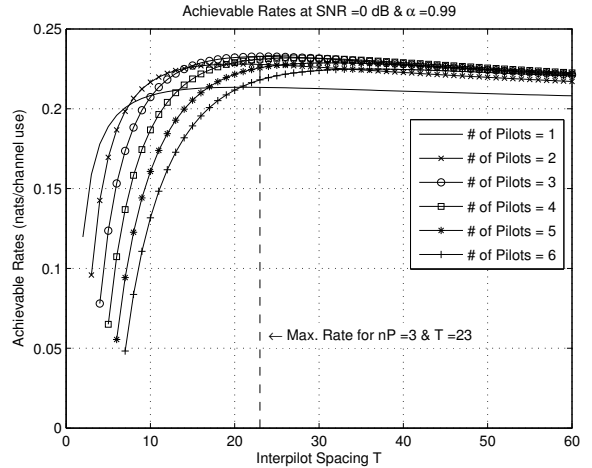


Fig. 5. Achievable Rates for SNR = 0dB and $\alpha = 0.99$.

For an SNR of 0dB (Figures 4 and 5), there is a clear benefit for pilot clustering. When $\alpha = 0.97$ the optimal clustering is that of two pilots, but the improvement over a single pilot is minimal. Indeed, the plots indicate that one should forgo training all together as the achievable rates are lower than those without training (equal to those of the IID Rayleigh fading channel). As α increases the optimal size of the training window increases as indicated in Figure 5 where $\alpha = 0.99$ and a cluster of three pilots is optimal. This can be attributed to the fact that at low SNR, when the correlation between fading

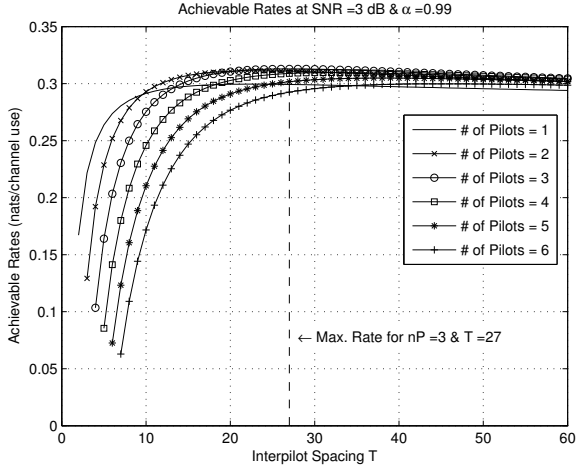


Fig. 6. Achievable Rates for SNR = 3dB and $\alpha = 0.99$.

coefficients is high, more training significantly improves the channel side information quality.

At a higher value of SNR of 3dB and for $\alpha = 0.99$, Figure 6 shows that the optimal number of pilots is three, indicating again that clustering is beneficial.

Albeit not clear from the few plots presented here, the authors note that as a general rule, longer training sequences have higher efficiency as the correlation factor increases and as the SNR decreases.

VI. IMPROVEMENTS & EXTENSIONS

A. Achievable Rates and MMSE Error Variance

The achievable rates' expression in (5) is the normalized sum of elementary "mutual information" quantities,

$$\max_{\{m_1, m_2, p\}_i} \mathbb{E}_{\hat{R}_i} \left[I \left(X_i; Y_i | \hat{R}_i \right) \right], \quad (6)$$

where $\hat{R}_i \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_R^2 - v_i)$. The quantity is a function of the estimation error variance v_i , and hence will be denoted by $I^k(v_i)$ for a k -pilot clustering strategy.

For given values of α , SNR, inter-training period T and k -pilot clustering strategy, we have $T - k$ elementary mutual information quantities each characterized by an estimation error variance v_j for $j = k, \dots, T - 1$. Let us denote this set of values of v_j for k -pilot clustering by $\mathcal{V}_T = \{v_j, j = k, \dots, T - 1\}$.

If we plot the variation of $\{\mathcal{V}_{T_1}, I^{k_1}(\mathcal{V}_{T_1})\}$, $\{\mathcal{V}_{T_2}, I^{k_2}(\mathcal{V}_{T_2})\}$, and $\{\mathcal{V}_{T_3}, I^{k_3}(\mathcal{V}_{T_3})\}$ corresponding to k_1, k_2 , and k_3 -pilot clustering techniques with T_1, T_2 , and T_3 pilot spacing intervals respectively, then according to (6) the 3 curves should superimpose to form one global curve $\{\mathcal{V} = \mathcal{V}_{T_1} \cup \mathcal{V}_{T_2} \cup \mathcal{V}_{T_3}, I(\mathcal{V})\}$. This is illustrated in Fig. 7 for SNR = 0dB, $\alpha = 0.99$ with 1, 2, and 3 pilot clustering strategies and $T_1 = T_2 = T_3 = 60$.

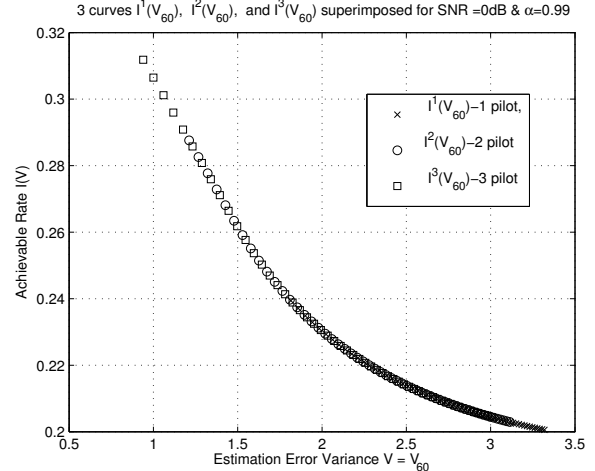


Fig. 7. The Global Curve $I(\mathcal{V})$ for SNR = 0dB and $\alpha = 0.99$.

For efficient computations of achievable rates in section (V), we actually compute $I(\mathcal{V})$ for a fine grid of values $\mathcal{V} = \{v < \sigma_R^2\}$ and every candidate level of α and SNR. Then, for given a specific k -pilot clustering strategy with $\mathcal{V}_T \subset \mathcal{V}$, then the curve $I^k(\mathcal{V}_T)$ can be easily extracted from the global curve $I(\mathcal{V})$.

Inspecting the global curve $\{\mathcal{V}, I(\mathcal{V})\}$ can give us insight about the efficiency of the clustering technique. For example, we notice that for a fixed value of SNR, the channels with higher degree of pilot clustering face low values of estimation error variance v in the beginning compared to channels with lower degree of clustering. However this gain margin in the values of v between high and low degrees of clustering shrinks significantly as SNR increases. This emphasizes the fact that training in high SNR regimes is not as beneficial and emphasizes our conclusion that clustering efficiency increases at low SNR levels.

B. High-Order Markov Rayleigh Fading Channels

The previous analysis may be readily generalized to high-order Markov Rayleigh channels. One may model the fading as a complex Auto-Regressive (AR) process of order p that can be generated via the time domain recursion,

$$R_l = - \sum_{m=1}^p \alpha_m R_{l-m} + Z_l,$$

where the $\{Z_l\}$ are IID complex circular Gaussian with mean zero and variance equal to σ_Z^2 . The AR model parameters consist of the filter coefficients $\{\alpha_1, \alpha_2, \dots, \alpha_p\}$ and σ_Z^2 .

Determining these parameters may be achieved through multiple techniques. For example, using Jakes' model [16], the normalized (unit variance) continuous-time auto-correlation function of the fading process is given by

$$\phi_{RR}(\tau) = J_0(2\pi f_d \tau),$$

where J_0 is the zeroth-order Bessel function of the first kind and f_d is the maximum Doppler frequency in Hertz. For the purposes of discrete-time simulation of this model [17], the auto-correlation sequence becomes

$$\phi_{RR}[l] = J_0(2\pi f_d T_s |l|),$$

where $1/T_s$ is the symbol rate. The relationship between Jakes' model ($\phi_{RR}[k]$) and the AR parameters may be determined by [18],

$$\begin{aligned} \phi_{RR}[l] &= -\sum_{m=1}^p \alpha_m \phi_{RR}[l-m], \quad l \geq 1 \\ \sigma_z^2 &= \phi_{RR}[0] + \sum_{m=1}^p \alpha_m \phi_{RR}[-m]. \end{aligned}$$

The authors are currently studying the performance of the proposed scheme for these higher-order Markov fading models. We expect that increasing the order of training through pilot clustering would be more suitable in these settings and achieve higher rates.

VII. CONCLUSION

In this work, we studied achievable rates over time-varying Rayleigh fading channels with memory and with imperfect side information acquired from channel estimation. The study of achievable rates of certain transmission schemes over these channels rather than capacity is common in the literature since tackling the capacity problem is infeasible.

We studied an adaptive non-feedback PSAM scheme where a cluster of pilots is used for the purpose of channel estimation at the receiver. Results show that a cluster (more than one) of pilots is optimal for correlation levels corresponding to Doppler frequency shifts of 100 Hz and bandwidths of 10 KHz. As the coherence time decreases, the optimal size of the cluster decreases. On the other hand, clustering is especially beneficial at low SNRs.

Extensions to this study are currently attempted on two fronts:

- 1) Determine optimal clustering sizes for channels whose Rayleigh fading is accurately modeled by an AR process. We expect the benefits of clustering to be more accentuated in these cases.

- 2) Adapt the scheme to the case of MIMO communications.

REFERENCES

- [1] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 11, pp. 686–693, Nov. 1991.
- [2] I. Abou-Faycal, M. Médard, and U. Madhow, "Binary Adaptive Coded Pilot Symbol Assisted Modulation over Rayleigh Fading Channels without Feedback," *IEEE Transactions on Communications*, vol. 53, no. 6, pp. 1036–1046, June 2005.
- [3] M. M. Khairy and E. Geraniotis, "BER evaluation of symbol-aided coherent demodulation for Rician and Rayleigh fading channels," *IEEE Symp. Computers and Communications*, vol. 1, pp. 105–109, 1999.
- [4] 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.
- [5] X. Cai and G. B. Giannakis, "Adaptive PSAM Accounting for Channel Estimation and Prediction Errors," *IEEE Trans. On Wireless Commun.*, vol. 4, no. 1, pp. 246–256, Jan. 2005.
- [6] S. Catreux, V. Erceg, D. Gesbert, and R. W. Heath, "Adaptive Modulation and MIMO coding for broadband wireless data networks," *IEEE Comm. Mag.*, vol. 40, pp. 108–115, June 2002.
- [7] G. E. Øien, H. Holm, and K. J. Hole, "Adaptive coded modulation with imperfect channel state information: System design and performance analysis aspects," *Proc. Int. Symp. Adv. Wireless Commun.*, pp. 19–20, Sep. 2002.
- [8] S. Zhou and G. B. Giannakis, "Adaptive modulation for multi-antenna transmissions with channel mean feedback," *IEEE Trans. on Wireless Commun.*, vol. 3, no. 5, pp. 1626–1636, Sep. 2004.
- [9] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, April 2003.
- [10] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 935–946, March 2000.
- [11] I. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.
- [12] R. Gallager, "Power limited channels: Coding, multiaccess, and spread spectrum," *MIT LIDS*, Nov. 1987.
- [13] S. Verdú, "On Channel Capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.
- [14] C. Luo, "Communication for wideband fading channels: on theory and practice," Ph.D. dissertation, Massachusetts Institute of Technology, Feb. 2006.
- [15] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993, vol. I.
- [16] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.
- [17] K. E. Baddour and N. C. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Trans. on Wireless Commun.*, vol. 4, no. 4, pp. 1650–1662, July 2005.
- [18] S. M. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.