Adaptive Coding for Pilot Symbol Assisted Modulation Without Feedback*

Muriel Médard[1], Ibrahim Abou-Faycal[1], Upamanyu Madhow[2]
[1] Laboratory for Information and Decision Systems, Massachusetts Institute of Technology
[2] Electrical and Computer Engineering Department, University of California Santa Barbara

Abstract

We show that Pilot symbol assisted modulation (PSAM) schemes can be improved by adapting the coded modulation strategy at the sender to the quality of the channel measurement at the receiver, without requiring any channel feedback from the receiver. The transmitter employs interleaved codes, with data symbols coded according to their distance from the nearest pilot symbols. The performance benefits from this approach are quantified in the context of binary signaling over time-varying Rayleigh fading channels described by a Gauss-Markov model. The spacing of the pilot symbols is optimized to maximize the mutual information between input and output in the preceding setting. Causal and non-causal channel estimators of varying complexity and delay are considered. It is shown that, by appropriate optimization for the spacing between consecutive pilot symbols, the adaptive coding techniques proposed can yield up to 30% improvement in achievable rate, without any feedback from the receiver to the sender.

1: Introduction

Adaptive schemes seek to modify the transmission scheme used by the sender according to the state of the channel seen by the receiver. Generally, such schemes involve feedback, concerning the state of the channel, from the receiver to the sender. In an information-theoretic context, adaptive signaling is used for Markov channels with perfect sender and receiver channel side information [3] or imperfect channel side information [2, 5, 7]. Power control is a commonly used type of adaptive transmission. Many practical schemes consider modifying the modulation used in order to combat fading. A common means of adapting transmission is to use different types of modulation, for instance different levels of QAM constellations [6], according to the state of the channel.

In this paper, we consider a different type of adaptive scheme. While the scheme still adapts the transmission to the channel seen by the receiver, the transmitter does not use feedback to determine its policy. Instead, the transmitter takes into account the time-varying quality of the channel measurement available at the receiver in order to modify its transmission policy. Thus, the transmitter adapts its signaling and coding to the quality of the channel measurement, rather than to the quality of the channel (in terms of carrier to noise ratio or other metric.) The channel measurement is obtained through regularly spaced pilot symbols.

Schemes studied in the literature generally consider the channel to be perfectly known at the receiver through the use of the pilot symbols. In the case where channel estimation at the receiver is provided by pilot symbols and where there is no feedback to the sender, the problem becomes, as we discuss below, that of transmitting over Ricean channels without channel side information at the receiver. The issue of capacity of Ricean channels is in general not known. In the special case of Rayleigh distribution of the channel with IID fading statistics among symbols, the capacity has been shown to be achieved using discrete inputs [1].

Our model is the following. We consider a single sender and receiver, connected by a continuously time-varying Rayleigh fading channel. The Rayleigh fading channel is modeled as a Gauss-Markov process. The sender transmits coded and modulated data. The sender has no information regarding the state of the channel and, therefore, does not adapt its transmission scheme in response to fades. Moreover, at regular intervals, the sender transmits a constant and known pilot symbol, whose purpose is to enable measurement of the

*This paper was partially presented at the 38th Annual Allerton Conference on Communication, Control and Computing, October, 2000

channel at the receiver. The pilot symbols have energy equal to the average energy constraint. The only channel estimate available at the receiver comes from the pilot symbols, hence there is no data-directed estimation of the channel. The channel estimate at a time sample k depends on the received pilot symbols through its position with respect to the pilot symbols. Given the estimate of the channel obtained through the pilot symbols, the channel is no longer Rayleigh but Ricean with a known specular component and an error component. Thus, we consider the channel estimation error explicitly rather than assume that the pilot symbols afford perfect channel side information at the receiver. We consider binary signaling, since such signaling performs well at low SNRs and achieves capacity for low SNRs for Rayleigh channels. The codes we consider are capacity-achieving binary random codes. Thus, our optimization is done directly in terms of mutual information. The maximization of mutual information subsumes the optimization of both the modulation and the coding. We select the binary signaling scheme that optimizes mutual information for our channel condition and energy constraints.

Overall, we consider three cases. First, we consider the case where no pilot symbols are transmitted and the channel is not estimated at the receiver. The purpose of considering the Rayleigh channel with no pilot symbols is to establish a basis of comparison for the other two cases, which do employ pilot symbols. Indeed, the first question we pose is whether it is preferable to forego pilot symbols and channel estimation altogether and devote to coded data the time allocated to pilot symbols. Second, we consider the case where we use pilot symbols to aid in the detection and decoding at the receiver but do not modify the distribution of the transmitted signal. We term this scheme the non-adaptive scheme with pilot tones. Third, we consider the case where we use pilot symbols as well as adaptive signaling and coding, but maintain the average per symbol power constant in the coded data. Note here that, since we are studying information rates, we are considering codes with block lengths increasing to infinity, and any estimation done in a non-causal fashion is non-restrictive. For instance, a reasonable non-causal estimation procedure would be to use all the pilot tones within a codeword.

In Section 2, we present our channel model and the principles of non-causal and causal estimation. In Section 3, we discuss in detail the different receiver estimation procedures that are considered. In Section 4 we discuss the non-adaptive and adaptive schemes used, and in Section 5, we present our numerical results. Finally, conclusions are presented in Section 6.

2: Channel Model

We consider the following discrete-time model for the Rayleigh fading channel

$$Y = RX + N,$$

where X is the channel input, Y the output, and R and N are independent complex circular Gaussian random variables with mean zero and variance σ_R^2 and σ_N^2 respectively. Equivalently, the amplitude of the fading coefficient R is Rayleigh distributed and its phase is uniform. The input X is average power limited: $E\left[|X|^2\right] \leq P$.

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

$$R_k = \alpha R_{k-1} + Z_k,$$

where k denotes the time index and the $\{Z_k\}$'s are IID circular Gaussian distributed with zero mean and variance equal to $(1 - |\alpha|^2)\sigma_R^2$. Depending on the scale of the time variations of the channel, α[1] may vary between one (corresponding to no channel variations) and zero (where the fading is fast, changing independently from one symbol time to the next.) A similar channel model was introduced in [4], where the relation between this channel and the coherence time is established.

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

For an adaptively coded system, we consider the case where a pilot signal of power P is transmitted once at the beginning of each interval of length T, enabling the receiver to estimate the fading coefficients governing the channel statistics. The resulting outputs are

$$Y_{lT} = R_{lT}\sqrt{P} + N_{lT}, \quad l \in \mathbb{N}.$$

Based on the observation of $\{y_{lT}\}_l$, estimates of the fading coefficients $\{R_k\}_{lT}^{(l+1)T-1}$ are obtained at each time step $k, (lT < k < (l+1)T)$. Using its estimate the receiver experiences a Ricean channel where the fading coefficient has mean $\hat{R}_k$ and variance v_k, where $\hat{R}_k$ is the estimate of the value of R_k and v_k is the variance of the estimation error:

$$p_{Y_k|X_k}(y_k|x_k) = \frac{1}{\pi(v_k|x_k|^2 + \sigma_N^2)} \exp\left\{-\frac{|y_k - \hat{R}_k x_k|^2}{v_k|x_k|^2 + \sigma_N^2}\right\}$$

[1] Since the fading is Rayleigh distributed, it is sufficient to consider only real positive values for α.

Therefore[2], for a given received sequence $\{y_{lT}\}_l$ the mutual information $I_k(X_k;Y_k|\{Y_{lT}\}=\{y_{lT}\})$ at time $k\in\{lT+1,\cdots,(l+1)T-1\}$ is given by

$$I_k\big(X_k;Y_k|\{Y_{lT}\}=\{y_{lT}\}\big)=$$
$$p_k(1)\int p_{Y_k|X_k}(y|x_k(1))\ln\frac{p_{Y_k|X_k}(y|x_k(1))}{p_{Y_k}(y)}dy+$$
$$(1-p_k(1))\int p_{Y_k|X_k}(y|x_k(2))\ln\frac{p_{Y_k|X_k}(y|x_k(2))}{p_{Y_k}(y)}dy,$$

where the probability mass function of the binary input $\{x_k(1),x_k(2)\}$ is $\{p_k(1),(1-p_k(1))\}$.

Furthermore, since we are eliminating the possibility of data-directed estimation, the achievable rates[3] are

$$\mathsf{E}\left[\frac{1}{T}\sum_{k=lT+1}^{(l+1)T-1} I_k\big(X_k;Y_k|\{Y_{lT}\}=\{y_{lT}\}\big)\right]$$
$$=\frac{1}{T}\sum_{k=lT+1}^{(l+1)T-1}\mathsf{E}\left[I_k\big(X_k;Y_k|\{Y_{lT}\}=\{y_{lT}\}\big)\right],\qquad(1)$$

where the expectation is over the (correlated) random variables $\{Y_{lT}\}$ which are circular Gaussian distributed with mean zero and variance $\sigma_R^2P+\sigma_N^2$.

3: Estimation at the Receiver

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

1. E(1,0) The channel parameters estimation is causal and based on the most recently sent pilot tone. For each interval, based on the observation of y_{lT}, estimates of the fading coefficients $\{R_k\}_{lT}^{(l+1)T-1}$ are obtained by a standard Bayesian least square estimation procedure.

2. E(∞,0) The receiver performs a maximum-likelihood estimation of the channel parameters using all the previous data available from the pilot tones. Practically speaking, the receiver is designed to continuously run (in a recursive manner) a Kalman filter on the data received from the pilots transmissions.

3. E(1,1) The estimation is non-causal, based on the most recently transmitted pilot, as well as on the pilot transmitted next, i.e. $\{R_k\}_{lT}^{(l+1)T-1}$ are based on the values of y_{lT} and $y_{(l+1)T}$.

4. E(∞,∞) Finally, the channel parameters are estimated non-causally using a fixed-time Kalman filter smoother. In essence, the receiver is assumed to perform a smoothing operation where the channel parameters are estimated given data that extends to times beyond those at which these estimates are sought. On a practical level, the analysis corresponds to the limiting behavior of a receiver that is designed to run a finite delay fixed point Kalman smoother on the data received from the pilots transmissions. The computations can be done recursively according to the well-known Kalman filter equations.

4: Non-Adaptive and Adaptive Coding

First we consider the scheme where the transmitter does not adapt its transmission strategy to the statistics of the channel estimates at the receiver. Namely, we compute the achievable rates when the transmitter is using a single fixed input distribution at all times.

We look mainly at the case where the transmitter considers the channel to be block-faded, i.e. the fading coefficient is assumed constant over intervals of length T and changing independently from one interval to the next. Consequently, we find first the optimal input distribution for the block-faded system, and then compute the average mutual information under this distribution for different values of α. While when $\alpha=1$, the model based on procedure E(1,0) would correspond to block-fading, as α decreases we expect the performance of this non-adaptive system to deteriorate, and we quantify this performance loss.

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

Equivalently, one can think of the problem as that of finding the best input strategy that maximizes the expected mutual information $\mathsf{E}\left[I_k\big(X_k;Y_k|\{Y_{lT}\}=\{y_{lT}\}\big)\right]$ for each time step between $lT+1$ and $(l+1)T-1$. For these computations, we considered the estimation methods described in section 3, i.e. for the pair $\left(\{\hat{R}_k\},\{v_k\}\right)$.

Since no closed form expression can be obtained for the optimal input distribution, we use standard *Matlab* tools to optimize, for each time period $k\in$

[2] Under the assumption of binary signaling.

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

$\{1, \cdots, T-1\}$, the expected mutual information over the input probability distribution. The corresponding optimal distribution yields of course the highest achievable rates depending on how far the transmission is occurring with respect to the pilot signals. Sending pilot tones frequently clearly reduces the rates as a significant portion of the time and power is used to estimate the channel and no information is conveyed from the transmitter to the receiver. On the other hand, when the pilots are used very infrequently, the channel estimates at the receiver are of poor quality and the information rates are low. One of the questions that we will try to answer in the following section is: what is the optimal value of T that would yield the best compromise?

5: Numerical Results

In the following plots, the line with the long dashes corresponds to the case where we send no pilot tones. The dashed curves correspond to the case where we do not vary the input distribution, but rather use a single distribution for all symbols other than pilot symbols. The full curves consider signaling and codes that adapt to the location from a coded symbol to the pilot symbols. Thus, symbols closer to the pilot symbols are coded with higher rate codes, while codes for symbols further away are coded with lower rate codes. This scheme is readily implemented by using interleaved codes where the interleaving period is the period of the sounding interval. The curves are parameterized by values of α, for $\alpha = 0.9, 0.95, 0.97, 0.99$, with higher curves corresponding to higher values of α. The optimal value for a particular scheme and SNR occurs at the maximum of the curve.

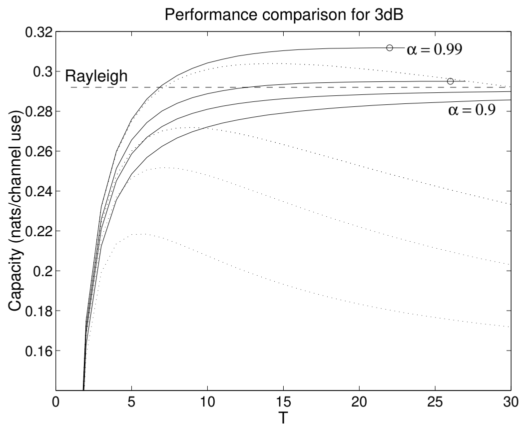
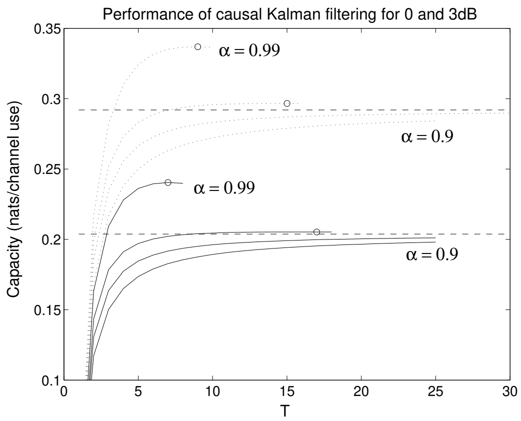
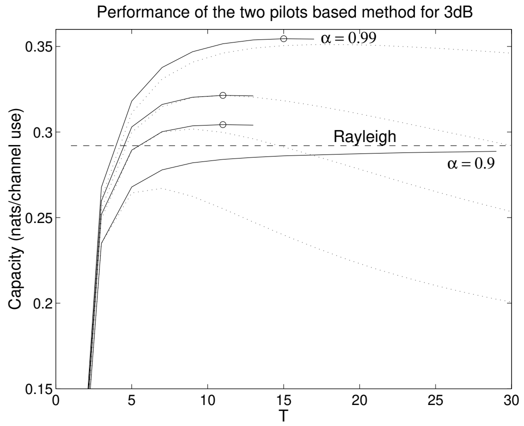


Figure 1 shows the results for applying E(1,0) for an SNR of 3 dB. Note that for the considered SNR, there is no benefit gained from using pilot symbol assisted modulation for $\alpha = 0.9, 0.95$. Indeed, for these low correlation models, not sending a pilot tone and coding instead to a memoryless Rayleigh fading channel outperforms the adaptive and non-adaptive causal scheme we have described. However, for $\alpha = 0.97, 0.99$, the figures show not only that an improvement is possible, but also that there is a trade-off between infrequent transmission of pilots and the channel estimate quality. Indeed, for every value of the SNR there is an optimal value for T where the rates are maximized, and these values are apparent in the figure. This holds for both the adaptive and non-adaptive schemes.



Similar results hold for E(∞,0) and Figure 2 is analogous to Figure 1. The same conclusions hold and the general behavior of the curves remains the same. While there is an advantage to considering all past pilot symbols, E(∞,0) does not provide significant improvement over E(1,0). The improvement is limited by the fact that pilot symbols before the last pilot symbol are fairly weakly correlated with the current channel realization. For a high correlation among channel realizations during different pilot tone transmissions, we would require a very low T, which would be inefficient. The optimal value of T, however, is reduced with respect to E(1,0). The benefit of sounding more frequently springs from the fact that pilot symbols provide more benefit than for E(1,0), since pilot symbols are used in many channel estimates rather than in a single one.

The above discussion for E(∞,0) motivates considering E(1,1), in which the two closest pilot symbols are used for estimation. Figure 3 shows the results for an SNR of 3 dB. Comparing for $\alpha = 0.95$ our results with those for the causal methods, E(1,0) and E(∞,0), we see that even for non-adaptive signaling and coding, sounding is preferable over no pilot symbols. Note also that the difference between the performance of the adaptive and non-adaptive schemes is lower for E(1,1) than for the causal methods, particularly when $\alpha = 0.99$. The reason for this reduced difference in performance is the following. In the case



of two-pilot non-causal estimation, the variance of the estimation error does not fluctuate throughout the interval between consecutive pilot symbols as much as for causal methods. For causal methods, the channel estimate is poor for symbols far from the last sounding symbol. Indeed, the worst estimation error variance is for the coded symbol immediately preceding the next pilot symbol. In the non-causal case, the proximity to the next pilot symbol is used to provide a good estimate.

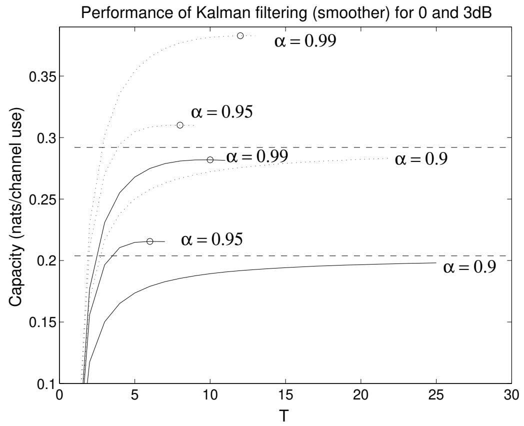
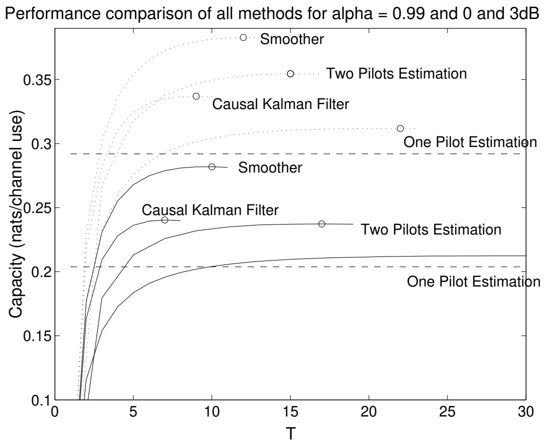


Figure 4 shows the performance of E(∞,∞) for SNRs of 0 and 3 dB. As expected, the method performs better than any of the previous methods. The behavior is generally comparable to that of E(1,1). A joint comparison of all methods for adaptive coding with $\alpha = 0.99$ and 3 dB SNR is shown in Figure 5.

Note that for an SNR of 3 dB, E(1,1) outperforms E(∞,0). Moreover, the difference in performance among the methods tends to be reduced as the SNR increases. While E(∞,∞) requires in theory an infinite number of past and future symbols, in practice it can be implemented with a finite window the performance of which would fall between that of E(1,1) and E(∞,∞).



6: Conclusion

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

References

[1] I.C. Abou-Faycal, M.D. Trott, and S. Shamai. The capacity of discrete-time memoryless Rayleigh-fading channels. *IEEE Transactions on Information Theory*, 47(4):1290–1301, May 2001.

[2] G. Caire and S. Shamai. On the capacity of some channels with channel state information. *IEEE Transactions on Information Theory*, 45(6):2007–2019, September 1999.

[3] A.J. Goldsmith and P. Varaiya. Capacity of fading channels with channel side information. *IEEE Transactions on Information Theory*, November 1997.

[4] M. Médard. The effect upon channel capacity in wireless communicationsof perfect and imperfect knowledge of the channel. *IEEE Transactions on Information Theory*, 46(3):935–946, 2000.

[5] M. Médard and R. Srikant. Capacity of nearly-decomposable markovian fading channels under asymmetric receiver-sender side information. In *International Symposium on Information Theory*, page 413, 2000.

[6] J.M. Torrance and L. Hanzo. On the upper bound peformance of adaptive QAM in a slow Rayleigh fading. *Electronics Letters*, 32:1218–1230, October 1997.

[7] H. Viswanathan. Capacity of time-varying channels with delayed feedback. *IEEE Transactions on Information Theory*, 45(2):761–771, March 1999.