Adaptive Signaling under Statistical Measurement Uncertainty in Wireless Communications

Honglei Zhang[1], Shuangqing Wei[2],
Ganesh Ananthaswamy[3] and Dennis L. Goeckel[1]
[1]University of Massachusetts at Amherst, [2]Louisiana State University,
[3]Analog Devices, Inc.
Email: [1]{hlzhang, goeckel}@ecs.umass.edu, [2]swei@ece.lsu.edu,
[3]ganesh.ananthaswamy@analog.com

Abstract

Wireless links form a critical component of communication systems that aim to provide ubiquitous access to information. However, the time-varying characteristics (or "state") of the wireless channel caused by the mobility of transmitters, receivers, and objects in the environment make it difficult to achieve reliable communication. Adaptive signaling exploits any channel state information (CSI) available at the transmitter to provide the potential to significantly increase the throughput of wireless links and/or greatly reduce the receiver complexity. As such, adaptive signaling has been a topic of significant research interest in the last decade and has found application in numerous commercial wireless systems, ranging from cellular data systems to wireless local area networks (WLANs). However, one of the great challenges of wireless communications is that it is difficult to obtain perfect CSI, since the CSI that the transmitter employs is inherently noisy and outdated. In response to this challenge, we have championed the idea of choosing the appropriate transmitted signal based on statistical models for the current channel state conditioned on the channel measurements. In this semi-tutorial paper, we overview how this class of methods has been developed over the last decade in design for single-antenna systems, and then present novel recent designs for multiple-antenna systems. Numerical results will demonstrate that such an approach provides a robust method for improving system data rate versus the commonly practice of employing link margin to compensate for such uncertainties.

[1]This work was supported by the National Science Foundation under grants NCR-9714597, ECS-0300130, and CCF-0430892, and a grant from Analog Devices, Inc.

[2]This work is based, in part, on work that appeared at IEEE GLOBECOM 2006.

1 Introduction

One of the foremost goals of modern technology is providing ubiquitous access to information. Under normal conditions in locations with a developed infrastructure, wireless communications is an attractive method of untethered access that extends the wired infrastructure to conveniently complete this ubiquity. However, in developing locations or in disaster situations, the wired infrastructure may not exist or be inadequate, and thus wireless communications may be the most feasible option for rapid and/or cost-effective connectivity. Hence, wireless communication links are an integral and important part of modern communications. However, although clearly vital, wireless links present one of the most challenging of communication channels. In particular, the transmitted radio signal reflects off the numerous objects commonly encountered between the transmitter and receiver, and, hence, the receiver sees the superposition of numerous copies of the transmitted signal. Due to the different path-lengths for the various reflections, these signals experience small relative delays, thus causing the multiple copies to arrive with different phases. In locations where these phases add constructively, signal enhancement takes place; in locations where the phases add destructively, signal strength is lost. Hence, one way to picture the result is as an interference pattern in space set up by the superposition of the transmitted and reflected waves. The resulting “multipath fading” is one of the distinguishing features of wireless system design and provides some of its greatest challenges.

Another characteristic of the wireless channel is that the channel (and thus its quality) can change at a rapid rate relative to traditional wired or satellite links. In fact, this can happen at a number of different time scales. At the longest time scale, changes in the distance between the transmitter and receiver (path-loss) and the existence of large objects between the transmitter and receiver (shadowing) can greatly alter signal strength. At short time scales, the multipath fading described above can significantly change. For example, consider a fixed transmitter and set of objects that reflect the signal. Per the discussion above, the transmitted signal reflects off these objects to form a spatial interference pattern. Since the receiver sees different channel characteristics as it moves across the interference pattern, any receiver mobility results in a communication channel with time-varying characteristics. Of course, the interference pattern is not static, either, and movement of either the transmitter or objects in the environment will cause it to change. Thus, the proper view is of a constantly shifting interference pattern with a receiver moving through it.

Because of this highly variable nature of the wireless link, it is difficult to prescribe a communication system that works well across all potential conditions. In particular, provisioning such that the system works with extremely high reliability can lead to a very conservative design with a significant reduction in the allowable data rate. One method of avoiding such a conservative approach is to adapt the transmitter as the propagation environment evolves. In particular, if one knows the characteristics of the current communication channel, one can match the current transmission to such. As a simple rough example, one can think of maintaining constant reliability by using a high data rate when the channel is good and a lower data rate when the channel is bad. Hence, rather than having to take a conservative position so as to make sure communication is reliable when the channel is bad, one envisions achieving whatever the channel will currently allow. Such techniques have been employed in wired channels to great success; for example, the discrete multitone approach (DMT) to digital subscriber lines (DSL) adapts its rate to the quality of the channel on each tone[1, 2].

Clearly, due to its variability, the wireless channel suggests an adaptive approach as well, and the idea was first investigated nearly four decades ago [3, 4]. Due to significant hardware advances, the last decade has seen rapid penetration of this adaptive approach into wireless applications and major industrial standards. Slow adaptations exploit long-term effects (path-loss, shadowing), and have been employed in numerous recent standards. For example, the second generation (2G) cellular communication systems of the late 1990's employed power control and rate adaptation techniques to solve the near-far problem and increase spectral efficiency [5, 6, 7]. In third generation (3G) cellular standards, faster (up to 1500 updates per second) and more sophisticated adaptive techniques are adopted to fight path-loss, interference and even multipath fading[8, 9, 10]. In addition, there has been recent penetration into the numerous WLAN air interfaces(e.g. IEEE 802.11), where multiple data rates are supported for the frame transmission to adapt to underlying channel qualities[11, 12, 13, 14]. In the IEEE broadband wireless access (BWA) standard 802.16, adaptive burst profiling is employed to adjust the modulation and coding schemes on a user-by-user and frame-by-frame basis[15]. In a promising mobile data system, Qualcomm's Flash-OFDM, coding and modulation are adapted quickly on a per-segment basis to each of the underlying channels[16].

Per above, it is the variation over time of the wireless channel that makes adaptation so attractive for improving performance. However, it is this same variation that makes adaptation in the wireless environment difficult. In particular, for the transmitter to adapt to the channel, it must obtain knowledge of the current state of such. This can be obtained via measurements fed back from the receiver, as in frequency division duplex (FDD) systems, or it might exploit reciprocity of the channel to obtain such by measuring the signal characteristics of a signal fed back from the receiver, as in time division duplex (TDD) systems. However, these estimates are perturbed by noise, of course, and, more importantly, by the time they are used these estimates have become outdated. As will be reviewed and further established below, such errors in measurements of the current state can have a significant impact on adaptation. In fact, we assert that dealing with this difficulty of outdated and/or noisy channel estimates is one of the key defining features of adaptation in the wireless environment, and this paper will be devoted to various methods of coping with such.

This paper is organized as follows. Section 2 briefly reviews various methods developed to perform adaptive coding with outdated and/or noisy measurements in the single-antenna case, and provides a basic tutorial treatment of the approach taken in [17]. Section 3 reviews analogous work for multiple-input multiple-output (MIMO) systems and then presents a novel technique for performing adaptation with full outdated CSI in such systems. Section 4 provides conclusions and directions for future work.

2 Adaptive Modulation in Single-Antenna Systems

2.1 System model and background

The basic system model to be considered is shown in Figure 1. The goal is to transmit the stream of information bits $\{b_i\}$, assumed to be an independent and identically (IID) sequence, over the wireless channel in such a way to achieve high reliability in the sequence $\{\hat{b}_i\}$, the corresponding stream of information bit estimates at the output of the receiver. A standard digital communication transmitted signal is given by $s(t) = \sum_{k=-\infty}^{\infty} z_k p(t - kT_s)$, where z_k is the k^{th} (complex) data symbol, $p(t)$ is a unit energy pulse that satisfies the Nyquist condition and thus results in no intersymbol interference in the samples (spaced at T_s) of the output of the matched filter at the receiver [18, pg. 543], and $\frac{1}{T_s}$ is the symbol rate. The additive noise $n(t)$ is a zero-mean white Gaussian random process with two-sided power spectral density $\frac{N_0}{2}$.

The multipath fading is modeled as a complex multiplier $X(t)$, thus implying that the multipath fading causes the signal to experience a power gain or loss, but no pulse distortion. This model, termed frequency-nonselective fading, is appropriate for a narrowband wireless channel [18, pg. 816] or a single subchannel of a wireless multicarrier system [19]. Following the standard practice of assuming that there are a large number of paths, the Gaussian wide-sense stationary uncorrelated scattering (GWSSUS) fading model [18] is assumed, where the independent component Gaussian processes are zero mean with autocorrelation function $R_X(\tau)$; that is, $X(t) = X_R(t) + jX_I(t)$ is complex Gaussian, where $X_R(t)$ and $X_I(t)$ are the respective real and imaginary parts of $X(t)$, $j = \sqrt{-1}$, $E[X_R(t)] = E[X_I(t)] = 0$, and $E[X_R(t)X_R(t+\tau)] = E[X_I(t)X_I(t+\tau)] = R_X(\tau)$. Such a zero-mean complex Gaussian channel is termed a Rayleigh fading channel [18]. It is also assumed that $X(t)$ varies slowly enough that analysis can be performed by assuming it is constant over the support of the pulse $p(t)$ of a single data or pilot symbol.

The key difference between the system model of Figure 1 and that of a standard communication system is the availability at the transmitter of the vector $\underline{\hat{X}}$ of channel fading estimates, where $\tau_{i+1} > \tau_i, \forall i$, and N is the number of outdated channel estimates employed. These estimates can be obtained via literal feedback of measured fading values from the receiver, or the estimates can be measured by the transmitter from a signal sent from the current receiver to the current transmitter. Per the Introduction, using such estimates to adapt the transmitted signal to the current channel conditions can greatly improve performance of wireless systems, and many authors have considered the system design and the resulting potential gains.

Most previous authors have assumed that that current channel state can be estimated without error from the channel measurements. First, consider the fundamental limits of system performance improvement from the perspective of information theory. The seminal work by Goldsmith and Varaiya [20] indicates that the Shannon capacity of fading channels with perfect CSI at both transmitter (CSIT) and receiver (CSIR) can be achieved by varying both the transmission rate and power, although adapting the rate alone loses little optimality. Results in [21, 22] show that the variable-rate variable-power coding scheme in [20] is not mandatory to achieve the ergodic capacity of fading channels with perfect CSIT and CSIR. In particular, the channel capacity can be reached by using a single codebook with a waterfilling-type [23] power allocation.

If the channel correlation model is exactly known to the transmitter, the system model in Figure 1 can be treated as a special case depicted in [21], where the fading process has memory as characterized by the correlation function. When channel memory is modeled by a Markovian process and is assumed perfectly known to the receiver, [24] obtains the capacity of such channels with delayed feedback. In particular, shadowing channels with delayed feedback are investigated where the correlation structure is captured using the first order autoregressive (AR) model. Again, exact knowledge on such fading correlations is required to compute the capacity. However, as pointed out in [25], there in general does not exist analytical solutions to the capacity of Markovian channels with delayed feedback, not to mention the case where only imperfect CSIR is available.

As demonstrated in [20], even the improvement on ergodic capacity with *perfect* CSIT and CSIR over the case with only perfect CSIR is not significant. We therefore do not expect significant capacity improvement in the presence of imperfect CSIT and feedback delay. Other than the imperfect CSI and delay introduced in providing CSIT, an additional challenge we are going to confront is the uncertainty about the exact fading correlation function in our system model, which will be discussed further below.

Information theory results assume an infinite code length and random coding. Although such codes are impractical in practice, recent advances in turbo and low-density parity-check (LDPC) codes are approaching the Shannon limit; more importantly, information theory often provides guidance to practical communication system design [25]. But such guidance must be interpreted carefully in the presence of practical constraints.

In the case of adaptive signaling design for frequency-nonselective Rayleigh fading channel with perfect CSIT, Shannon results suggest that the adaptive approach only slightly improves the capacity over the non-adaptive one, which implies that adaptation is not attractive in this scenario. However, in contrast to what Shannon theory suggests, adaptive coding has proven to be effective for wireless channels with perfect CSIT in the context of communication theory. In particular, it has been demonstrated in [26] that for adaptive uncoded M-ary quadrature amplitude modulation (M-QAM), the gains through adaptation over the frequency-nonselective Rayleigh fading channel are enormous.

However, in systems with significant error control coding and interleaving [27, 17], the gains, although still significant, are much smaller. In particular, non-adaptive coded and interleaved systems are able to exploit diversity to significantly shrink the gap between their performance and the channel capacity. Since the capacity of adaptive systems is only slightly larger than non-adaptive systems, the potential gains are much smaller in coded systems, and, as noted above, almost disappear altogether in the limit of long channel codes.

Thus, it is reasonable to conclude that the gains of adaptive coding in terms of average rate are roughly inversely proportional to the amount of coding/decoding complexity in the system. Viewed differently, it could be stated that adaptive coded systems allow identical performance to non-adaptive coded systems at a significantly reduced receiver complexity.

2.2 Distribution of Fading Variables Conditioned on Measurements

The key to the design of an adaptive scheme is the selection of the transmission parameters based on the measured estimate. Since data rate is the most common and powerful adaptation in practical systems, the focus here is on such. We also consider only uncoded quadrature amplitude modulation (QAM) for simplicity of exposition, although the extension to coded systems is easily made through the techniques of [17, 28]. The data rate in bits for a given symbol is given by $\log_2 M$, where M is the number of signals in the signal set employed. Thus, the key is to select a signal set M to be employed when $\underline{\hat{X}} = \underline{x}$. Suppose the potential signal sets consist of 0-QAM (no transmission), 2-QAM (BPSK), 4-QAM (QPSK), 16-QAM, $\ldots$, M_{max}. Specifying an adaptive scheme is then equivalent to partitioning the space of all vectors $\underline{\hat{X}}$ into disjoint regions where each region corresponds to the set of measured vectors under which M-QAM will be employed.

Following [29], let the symbol of interest be the k^{th} symbol, whose pulse starts at time kT_s, and assume temporarily that $R_X(\tau)$ is known at the transmitter. Also assume that the i^{th} most recent channel estimate was made using a pilot symbol with average received energy E_p over the support of its pulse $p(t)$ that starts at $kT_s - \tau_i$, and let $\hat{X}(kT_s - \tau_i)$ be the product of the sample of the matched filter output for the pilot symbol and $\sqrt{\frac{2}{E_p}}$. Denoting Y as the magnitude of the fading that multiplies z_k in the matched filter output for the k^{th} symbol and using the fact that linear functionals of a Gaussian random process are jointly Gaussian, Y is Rician when conditioned on the vector $\underline{\hat{X}}$, with probability density function:

$$p_{Y|\underline{\hat{X}}}(y|\underline{x}) = \frac{y}{\sigma^2} e^{-\frac{y^2+s^2}{2\sigma^2}} I_0\left(\frac{ys}{\sigma^2}\right), \qquad y \geq 0, \tag{1}$$

where $I_0(\cdot)$ is the 0th-order modified Bessel function. Using the assumption that $X(t)$ can be assumed constant over the support of $p(t)$ and normalizing the fading such that $E[(X_R(kT_s))^2] = E[(X_I(kT_s))^2] = 1$ (Hence, for exposition purposes, making the average received energy twice the average transmitted energy, which will be accounted for below.), the noncentrality parameter in (1) is given by:

$$s^2 = (\underline{\rho}^T(\Sigma_{\underline{X}} + \sigma_\epsilon^2 I_N)^{-1}\underline{x}_R)^2 + (\underline{\rho}^T(\Sigma_{\underline{X}} + \sigma_\epsilon^2 I_N)^{-1}\underline{x}_I)^2$$

where I_N is an N by N identity matrix, and $\sigma_\epsilon^2 = \frac{1}{\frac{E_p}{N_0}}$ is the variance of the noise on the in-phase (or quadrature) component of a channel estimate. The $(m,n)^{th}$ element of $\Sigma_{\underline{X}}$, the N by N autocorrelation matrix of the in-phase component of $\hat{\underline{X}}$ when the channel estimates are noiseless, is given by $R_X(\tau_{N-m+1} - \tau_{N-n+1})$, and the correlation vector of the in-phase component of $\hat{\underline{X}}$ with the in-phase component of the fading of interest is given by $\underline{\rho}$, where $\rho_i = R_X(\tau_{N-i+1})$. The parameter σ^2 in (1) is the mean squared error of a minimum mean squared error (MMSE) prediction of the in-phase (or quadrature) fading of interest, and is given by $\sigma^2 = 1 - \underline{\rho}^T(\Sigma_{\underline{X}} + \sigma_\epsilon^2 I_N)^{-1}\underline{\rho}$.

To interpret (1), consider the case of a single outdated noiseless estimate ($N = 1$). Let $h \triangleq |\hat{X}(kT_s - \tau_1)|$. For $N = 1$, the parameters for (1) simplify to $s^2 = \left(\frac{\rho_1}{\sqrt{1+\sigma_\epsilon^2}}\right)^2 \left(\frac{h}{\sqrt{1+\sigma_\epsilon^2}}\right)^2$ and $\sigma^2 = 1 - \left(\frac{\rho_1}{\sqrt{1+\sigma_\epsilon^2}}\right)^2$. There is an important observation to be made: delay between channel measurement and data transmission or noise in the estimates transforms the adaptive signaling problem from signaling for an additive white Gaussian noise (AWGN) channel with a modified received signal-to-noise ratio as in [26], to signaling for a Rician channel on which not only does the average received signal-to-noise ratio vary with the estimate but the Rician parameter varies as well.

Hence, (1) gives a nice characterization of the problem. If the predictions are accurate (i.e. σ^2 is very small), one can simply predict the channel and assume that that prediction is perfect [30, 31]. If the predictions are inaccurate (i.e. σ^2 is not very small), then one needs to take this inaccuracy into account, and it is the latter case that is considered here.

2.3 Potential designs

There are a number of methods for dealing with the case of outdated or noisy estimates in the wireless channel. In this section, we briefly review the various types before focusing on a specific type of adaptation.

The first method of dealing with inaccuracy in the channel measurements is to simply account for it in the link budget; in other words, a few dB of energy is expended to bring the performance of the adaptive system with inaccurate estimates to that of the adaptive system with perfect estimates. Such an approach has the advantage of not only being simple but also robust, since it needs not assume anything about the characteristics of the inaccuracy besides a rough estimate of the amount of degradation incurred. Its disadvantage, as we will see very clearly below, is that such an approach also limits the performance of the system.

A second approach is to develop a system architecture where you devote a certain amount of resources (e.g. code rate) to correcting errors in the state measurements.

A third approach to dealing with the outdated nature of the channel estimates is to employ prediction[30, 31]. In particular, outdated estimates of the channel are used to predict the current state of the channel, and then that estimate is used to perform adaptation of the transmitter. Clearly, such an approach is effective as long as the predictions are accurate. Potential limitations include the lack of knowledge of various environmental parameters (e.g. the autocorrelation function of the multipath fading), which are notoriously difficult to estimate. Also, in extreme mobility, the prediction becomes noisy, and then system performance degrades.

The last approach, the one considered extensively here, can be used on the raw channel measurements or in conjunction with a noisy predictor. Rather than trying to make a prediction of the channel, the uncertainty in the current status of the channel state is taken into account in much the same manner as system noise or fading is treated. In particular, the optimal adaptation scheme is to determine the current transmission parameters by simply finding the scheme that, when conditioned on the raw channel measurements, provides the highest rate with acceptable reliability over all possible correlation models in a uncertainty set [17].

2.4 Model-based Approach

2.4.1 Robustness

Prediction-based methods are effective if there is only moderate mobility and $R_X(\tau)$ is known exactly at the transmitter. However, spectral estimation (i.e. estimation of $R_X(\tau)$) is notoriously hard, and thus we seek methods that do not require knowledge of such. In particular, we seek design methods that will guarantee performance for each observation value $\underline{\hat{X}} = \underline{x}$

and any particular autocorrelation function in a given class $\mathcal{R}$. As in [17], we will consider schemes that only employ a single outdated channel estimate ($N = 1$). The resulting methods will be applicable to two cases: Robust adaptive signaling with $N = 1$, and adaptive signaling with arbitrary N when $R_X(\tau)$ is accurately estimated. Design for the second case is a simplification of design for the first case; thus, robust adaptive signaling with a single outdated fading estimate will be considered. This consists of guaranteeing results for all possible correlation coefficients $\rho_1 \in [\rho_{min}, 1]$, where ρ_{min} is the minimum correlation coefficient over $R_X(\tau) \in \mathcal{R}$ between $\hat{X}_R(kT_s - \tau_1)$ and $X_R(kT_s)$.

2.4.2 Design Rules

The design rules for uncoded systems have been well-established by a number of authors [32, 17, 26]. The signal sets considered in this section will be 0-QAM (no data transmitted), BPSK, QPSK, 16-QAM, 64-QAM with two-dimensional Gray mapping, although the extension to any set of signal sets is immediate.

Following [17], let P_b be the target bit error probability for the system, which operates at the average received signal-to-noise ratio $\frac{E_s}{N_0}$, where E_s is the average received energy per QAM symbol. For now, assume that the average energy E_s is constant across symbols. Energy adaptation is an important topic that will be discussed in detail below. Per above, specification of the adaptive transmitter requires determining the signal set to employ for each potential estimate; simplified to the $N = 1$ case, this requires finding $\tilde{M}(h), \forall h$, where $\tilde{M}(h)$ is the number of signals in the QAM signal set employed when $|\hat{X}(kT_s - \tau_1)| = h$. Since we want to maintain P_b for each h, $\tilde{M}(h)$ is chosen such that:

$$\tilde{M}(h) = \max\{M : \sup_{\rho_{min} \leq \rho \leq 1} \tilde{P}_M\left(\frac{E_s}{N_0}, h, \rho\right) \leq P_b\}, \tag{2}$$

where $\tilde{P}_M(\frac{E_s}{N_0}, h, \rho)$ is defined as the bit error probability of the M-QAM signal set at average received SNR $\frac{E_s}{N_0}$ when $R_X(\tau_1) = \rho$ and $|\hat{X}(kT_s - \tau_1)| = h$. Assume that optimal symbol detection given the current channel fading amplitude is employed on the samples of the matched filter output at the receiver. An approximation to the bit error rate(BER) of M-QAM modulations is given by [26]:

$$P_M\left(y^2\frac{E_s}{N_0}\right) \approx 0.2\exp\left(-\frac{3}{4(M-1)}\frac{E_s}{N_0}y^2\right), \tag{3}$$

which will be employed for all M here. If errors in channel estimation *at the receiver* are considered, the right side of (3) will increase, of course, but it will often fit into the same functional form [32], which is convenient, since the same optimization will apply. Using (3) yields [17]:

$$\begin{aligned}\tilde{P}_M\left(\frac{E_s}{N_0}, h, \rho\right) &= E\left[P_M\left(Y^2\frac{E_s}{N_0}\right)\Big||\hat{X}(kT_s-\tau_1)| = h\right] \\ &\approx \begin{cases} \dfrac{0.2\exp\left[-\frac{h^2\rho^2}{2(1-\rho^2)}\left(1-\frac{1}{1+\frac{3}{2}\frac{E_s}{N_0}\frac{(1-\rho^2)}{(M-1)}}\right)\right]}{1+\frac{3}{2}\frac{E_s}{N_0}\frac{(1-\rho^2)}{(M-1)}} & \rho < 1 \\ 0.2\exp\left(-\frac{3}{4}\frac{E_s}{N_0}\frac{h^2}{(M-1)}\right) & \rho = 1 \end{cases}\end{aligned} \tag{4}$$

where the second line is obtained from [33, 6.614.3].

From (2), (4) must be maximized over $\rho \in [\rho_{min}, 1]$. Per [17], this maximization is easily performed, as follows. Define

$$\tilde{\rho} = \begin{cases} 0 & h \geq \sqrt{2} \\ \sqrt{\left(1+\frac{2(M-1)}{3}\frac{N_0}{E_s}\right)\frac{(2-h^2)}{2}} & 0 \leq h \leq \sqrt{2} \end{cases}$$

The worst case ρ is then given by

$$\rho^* = \begin{cases} \rho_{min} & \tilde{\rho} \leq \rho_{min} \\ \tilde{\rho} & \rho_{min} < \tilde{\rho} < 1 \\ 1 & 1 \leq \tilde{\rho} \end{cases}. \tag{5}$$

The signal set is specified using (4) and (5) in $\tilde{M}(h) = \max\{M : \tilde{P}_M\left(\frac{E_s}{N_0}, h, \rho^*\right) \leq P_b\}$. Note that, since $\tilde{M}(h)$ is nondecreasing in h, the adaptive scheme can be specified by finding the values $h_m, m = 2, 4, 16, 64$, where h_m is the threshold above which m-QAM can be successfully employed.

The discrete nature of the set of rates for any finite collection of signal sets hurts the performance of the system. In particular, when m-QAM is employed, h will fall in the region $h_m < h < h_{2m}$, and thus bit error performance will be better than anticipated. Although it is tempting to view this positively, this is really a negative feature, since, in reality, data rate has been lost. Stated differently, an optimal adaptive system under a given bit error constraint should meet that constraint with equality. Although the proposed approach is not optimal, meeting the constraint with equality through the following technique improves the data rate while still meeting the bit error constraint. Rather than employing the energy adaptation method of [26] to solve this problem, we instead employ a method analogous to the power-pruning of [34]. The advantage of this method is that, with very little loss of optimality, it is easily extended to coded modulation structures, where the overall optimization problem of [26] is not easily framed when channel prediction is not perfect [17]. Once a signal set has been chosen, the system is essentially a fixed rate system; thus, the goal changes from maximizing average rate, to attempting to allow communication at this fixed rate with the least amount of power. Thus, after the signal set is chosen, (4) and (5) are used to decide the minimum energy required to maintain P_b given the channel estimate h, and this energy is employed rather than the average energy. Any excess energy is put into a "bank" on which successive symbols can draw.

2.4.3 Numerical Results/Analysis

Per Section 2.1, systems with a significant amount of decoding complexity and allowable latency only have the potential for a small amount of improvement when CSI is provided to the transmitter, and uncoded systems, which have the least decoder complexity and essentially no latency, benefit the most when transmitter CSI is available. In particular, uncoded systems operating over frequency-nonselective Rayleigh fading channels perform very poorly, because they do not achieve diversity. Because of this, coherently decoded quadrature phase-shift keying (QPSK) with only receiver CSI requires an SNR of 25 dB to achieve a bit error rate of 10^{-3} on a frequency-nonselective Rayleigh fading channel [18, pg. 829], whereas the same technique requires an SNR of well under 10 dB to achieve the same bit error rate on an AWGN channel. The reason for this discrepancy is that the QPSK system operating over the Rayleigh fading channel is extremely susceptible to deep signal fades. Although the occurrence of such is relatively uncommon, the error rate during a bad fade can be orders of magnitude above that occurring when the average received SNR is observed, and thus these bad fades dominate the error rate.

In adaptive signaling, CSI is available at the transmitter. Arguably, the greatest utility of such information is that signaling can be avoided when bad fades are present. Figure 4 demonstrate the spectral efficiencies of several designs with perfect CSIT in Rayleigh fading channels[26, 35]. The target BER is 10^{-3}, and the constellation set includes 0-QAM, BPSK, QPSK, 16-QAM, 64-QAM and 256-QAM. The variable power is continuous. As shown in the figure, using discrete constellations loses little spectral efficiency, and power adaptation provides almost 1 bit date rate gain when discrete constellations are used. In particular, average rates in excess of 2 bits per symbol are possible for average received SNRs under 13 dB; even the simple scheme that applies constant power and truncates the continuous constellation to the closest MQAM provides 2 bits per symbol under 15 dB. Thus, there is a significant gain in system performance when transmitter CSI is available in uncoded systems.

However, as pointed out in [36, 17], the assumption of perfect CSI is dangerous when channel estimates are outdated and/or noisy, as would be the case with realistic delay in the feedback path from the receiver to the transmitter. In particular, the conditional density function given in (1) becomes Rician (rather than a delta function), and, hence, the conditional channel acts like a fading channel. For example, as displayed in Figure 5, adaptive signaling assuming perfect channel estimation can miss its target bit error rate by two orders of magnitude - even for the relatively high correlation coefficients of $\rho = 0.96$. In this case, bad predictions, which are relatively uncommon, lead to instantaneous error rates that are orders of magnitude above the target and thus dominate system performance.

Per Section 2.3, a common method of countering imperfect CSIT effect on BER is to add energy margin during the energy adaptation process. Figure 6 compares the average data rates of this energy margin scheme and the design method of (1). As indicated in the figure, there are still significant gains in adaptive signaling versus non-adaptive signaling when CSIT is not perfect - even when the correlation coefficient drops as low as $\rho = 0.85$. In Figure 7, the average rates versus SNR are plotted for a number of values of ρ_{min} with the design BER 10^{-3} . The close correspondence of the $\rho_{min} = 1.0$ curve to the results in [26] suggests that the flexible power allocation is effective. We conclude from this section that adaptive signaling is particularly effective for simple, low-latency systems such as adaptive uncoded QAM systems [36, 17].

3 Adaptive Modulation in Beamformed Multiple-antenna Systems

The union of wireless and wired data systems continues to drive wireless communication systems to support even higher data rates. For example, next generation WLAN will support up to 600Mbps in over-the-air data rate[14]. Due to stringent power and spectrum constraints, designing high-speed wireless systems with guaranteed quality-of-service(QoS) forms a tremendous challenge. An emerging technology, multiple-input multiple-output(MIMO), has been developed to meet this challenge[37, 38]. In MIMO communication systems, multiple antennas are deployed at the transmitter and receiver to reap the degree-of-freedom and diversity gain in wireless channels with rich scattering.

Single-user MIMO communications has been extensively studied and continues to be an active research area in communication theory. Without an increase of transmit power and bandwidth, the MIMO system with m transmit antennas and n receive antennas can achieve ergodic capacity approximately $\min(m, n)$ times that of the single-antenna system. In general, CSI is assumed to be available at the receiver, but inaccessible at the transmitter due to the time-varying and asymmetric nature of wireless channels. In this case, the optimal power allocation is equally splitting among the transmit spatial substreams. With CSI available at the transmitter, the capacity-achieving strategy is to decompose the channel into parallel non-interfering subchannels with singular value decomposition(SVD) and waterfill the power over these subchannels(called channel eigenmodes)[37]. Figure 2 shows the k-th eigenmode: the data signal is transmitted along the k-th right singular vector, and at the receiver the channel output vector is combined with the k-th left singular vector of the channel matrix. Mathematically, this is a process of coordinate transformations in the transmit and receive signal space; in the perspective of channel modeling, this is just to beamform the signal towards the scatters. The capacity improvement over the open-loop case is substantial. Moreover, the diagonal equivalent channel can significantly reduce the receiver complexity.

In reality, the transmitter cannot have perfect CSI due to a number of reasons, among which are the constantly varying nature of wireless channels, channel estimation error, quantization error, and limited bandwidth of the feedback channel etc. The channel capacity with partial CSI at the transmitter and perfect CSI at the receiver is well documented. Research has focused on four mathematic models of transmit CSI: zero-mean spatially white(ZMSW), channel mean information(CMI), channel covariance information(CCI)[39] and CSI with finite rate feedback [40].

CMI models the case that transmitter has an imperfect instantaneous measurement of the channel. The major results of the channel capacity of the CMI model are briefly introduced in the following. For multiple-input single-output(MISO) systems, the principal eigenvector of the optimal input covariance is along the channel mean vector and the eigenvalues corresponding to other eigenvectors are equal[41]. For MIMO systems, it does not lose any optimality to transmit along the eigenvectors of the channel matrix[42]. For both MIMO and MISO systems, there is no closed-form optimal power allocation. Beamforming is a method to transmit signals over the principal eigenmode. Under some conditions, beamforming can achieve the channel capacity. The optimality condition of beamforming is given in [43] and [44] for MISO and MIMO systems, respectively. In short, beamforming becomes optimal when the power decreases, or the feedback quality improves in MISO systems.

Linear precoded spatial multiplexing has been proposed for the transmitter with full CSI [45, 46, 47, 48, 49, 50], CMI[41, 51, 52], CCI[53, 54, 55], both CMI and CCI [56], or bandwidth-limited feedback from the receiver[40, 57]. The optimization objectives to choose the precoder include a wide array of criteria, for example, minimizing the weighted mean squared error(MSE) [53], [48], maximizing the received minimum Euclidean distance[50], minimizing the BER[58] and so on. Based on the powerful convex optimization theory and majorization theory, a unified framework is proposed in [49] for the joint linear transceiver design with full CSI at both sides. This framework includes all the design criteria by categorizing them into two families: Schur-concave and Schur-convex objective functions. For Schur-concave functions, eigen-decomposing the channel is optimal; for Schur-convex functions, diagonalizing the channel after rotating the transmit vector is the optimal solution. In [58], the converse formulation of the problem is studied: minimizing the transmit power given the QoS constraint on each substream. Not surprisingly, the optimal structure is eigenbeamforming(up to a unitary rotation in some cases). In the CMI case, grafting orthogonal space-time block coding (OSTBC) with the beamforming is a viable approach[51, 52, 56](see an example in Figure 3). The maximum likelihood (ML) decoding in this case is simplified to linear operations. Note all the above designs assume the constellations are fixed.

In multiple antenna systems with full CSI at the transmitter, it is proven in [59] that the classic capacity-achieving waterfilling and the gap approximation method on each eigenmode are virtually optimal in terms of minimizing the transmit power. Adaptive modulation with the channel mean feedback has been explored in the transceiver based on OSTBC and eigenbeamforming[60, 61]. Although the Alamouti code has been shown to achieve the full capacity of a 2×1 system, simple full rate OSTBCs have not yet been developed for the large number of transmit antennas. A bit and power loading scheme for eigen-beamforming with partial CSI is proposed in [62]. It assumes the maximum likelihood detector, which dramatically increases the complexity of the receiver. A robust adaptive modulation scheme is presented for eigenbeamforming taking into account the channel estimation error and the CSI delay in [63].

Partial CSIT resulting from finite rate feedback has attracted considerable attention lately due to its underlying practical constraints imposed by limited feedback channel capacity in transmitting forward channel measurements. When the feedback channel is assumed error-free with limited capacity, there are in general three approaches in exploiting partial CSIT at the transmitter side, namely, channel vector quantization, scalar quantization and quantized signal adaptation schemes [40]. The channel quantization scheme is to perform vector quantization of the measurements directly at the receiver end using the traditional MSE distortion metric, then optimize the transmission signaling at the transmitter conditioned directly on the quantizer output[64].

The quantized signal adaptation approach attracts more attention than the former two approaches. As described in [65], it is a hybrid beamforming and power control design framework that can achieve the optimal link capacity under the feedback-link capacity constraint. In this approach, CSIR is assumed to be perfectly estimated at the receiver, and the CSIR space is partitioned into $L = 2^{R_{fb}}$ non-overlapping regions, where R_{fb} is the finite rate constraint for the feedback channel. Given the perfect CSIR, the associated index $q \in \{1, \cdots, L\}$ of the CSIR region is located and fed back to the transmitter. At the transmitter, there is a table of transmit covariance matrices $Q_i, i = 1, \cdots, L$, each of which is equivalent to performing both power control and beamforming after eigen-decomposition. Upon receiving the index q, the transmitter sends a Gaussian codeword matrix with the covariance matrix Q_q. The objective function becomes throughput maximization by jointly designing CSIR space partition and the selection of L covariance matrices under a power constraint $\text{Tr}(Q_i) \leq P_0$, where P_0 is the maximum average total transmission power. A more general case with possibly imperfect CSIR is considered in [66].

The precedent optimization problems can in general be solved using an iterative Lloyd's algorithm with modified distortion measure [65], which unfortunately cannot yield analytical solutions. Recently, under the same framework in using partial CSIT as described above, the quantization problem associated with maximizing received SNR or minimizing outage probability with finite rate feedback is successfully formulated as quantization on the Grassmann manifold [57, 67]. The asymptotically equivalent lower and upper bounds for the rate-distortion tradeoff of quantization on the Grassmann manifold is lately obtained in [68]. Asymptotic capacity loss versus the number of feedback bits is also investigated in [69, 70, 71].

The major motivation of finite rate feedback is to avoid sending a larger number of fading parameters back to the transmitter[40]. When finite-rate feedback on transmitting CSIT is considered, it is quite common to adopt a block fading channel model. This model is more appropriate for slowly varying MIMO channels with a large number of antenna elements. However, for many wireless applications involving relatively fast varying fading variables, it is more suitable to adopt MIMO channel models with temporally correlated fading variables and delayed feedback. In reality, the number of antenna is relatively small; thus, it is not a huge burden on the feedback channel to carry full CSI. For instance, at most 4 antenna elements are equipped at either end of the wireless link in IEEE 802.11n and full CSI (may be compressed) is fed back[14]. In the sequel, we will focus on the design making use of delayed, yet full CSIT.

Notation: Underlined lower-case letters are vectors, boldface upper-case letters are matrices; $[\mathbf{A}]_{ij}$ means the (i, j)th element of the matrix $\mathbf{A}$; $(\cdot)^H$ and $(\cdot)^T$ denote the Hermitian and transpose of a matrix or vector, respectively; $\mathrm{diag}(\underline{a})$ denotes a diagonal matrix with the vector $\underline{a}$ on the diagonal; $\mathrm{E}_X(\cdot)$ represents the expectation with respect to the random variable X; $\mathbf{I}$ is the identity matrix; all vectors are column vectors.

3.1 System Model

An uncoded frequency non-selective MIMO system with m transmit antennas and n receive antennas($n \geq m$) is considered in this paper. Hence , the standard time-varying baseband model

$$\underline{y}(t) = \mathbf{H}(t)\underline{x}(t) + \underline{n}(t) \tag{6}$$

will be employed, where $\underline{y}(t)$ is the $n \times 1$ received vector, $\underline{x}(t)$ is the $m \times 1$ transmitted vector, $\mathbf{H}(t)$ is the $n \times m$ channel matrix and $\underline{n}(t)$ is the noise vector with each entry a proper complex Gaussian random process. The elements of $\mathbf{H}(t)$ will be assumed to be zero-mean stationary complex Gaussian random processes, and, because the array elements are assumed to be sufficiently separated, it will be assumed that these processes are independent of one another. Assume the channel is constant over a symbol interval, and define $\mathbf{H} \triangleq \mathbf{H}(t_0)$ to be the true channel matrix for the symbol at the time t_0 of data transmission and $\mathbf{H}' \triangleq \mathbf{H}(t_o - \tau)$ to be the outdated channel information that the transmitter utilizes to do eigenbeamforming and adaptive modulation. Then, the continuous-time model gives rise to a discrete-time model for the symbol of interest:

$$\underline{y} = \mathbf{H}\underline{x} + \underline{n}, \tag{7}$$

where the time index has been suppressed, and the entries of $\mathbf{H}$ will be assumed to have unit variance. Let $\rho(\tau) = \mathrm{E}\{[\mathbf{H}]_{ij}[\mathbf{H}']_{ij}\}$ for all i and j. Then, recalling $[\mathbf{H}]_{ij}$ and $[\mathbf{H}']_{ij}$ are jointly Gaussian,

$$\mathbf{H} = \rho(\tau)\mathbf{H}' + \mathbf{\Sigma}, \tag{8}$$

where entries of $\mathbf{\Sigma}$ are independently identically distributed(i.i.d.) zero-mean proper complex Gaussian random variables with variance $(1 - \rho^2(\tau))$. The delay τ will be suppressed for the rest of the paper.

The SVD of the outdated channel matrix can be written as:

$$\mathbf{H}' = \mathbf{U}'\mathbf{D}'\mathbf{V}'^{H},$$

where $\mathbf{U}'$ and $\mathbf{V}'$ are unitary matrices containing the left and right singular vectors of $\mathbf{H}'$, respectively, and $\mathbf{D}'$ is a diagonal matrix containing the singular values of $\mathbf{H}'$ in descending order. Let $\underline{x}$ and $\underline{s}$ represent the eigenbeamformed and the data-bearing signal, respectively. Then,

$$\underline{x} = \mathbf{V}'\mathbf{E}\underline{s}, \tag{9}$$

where $\mathbf{E} = \mathrm{diag}(\sqrt{E_1}, \sqrt{E_2}, \ldots, \sqrt{E_m})$, and E_i is the average energy on the i-th eigenmode. Hence the received signal $\underline{y}$ is

$$\underline{y} = \mathbf{H}\mathbf{V}'\mathbf{E}\underline{s} + \underline{n},$$

where entries of $\underline{n}$ are i.i.d. zero-mean complex Gaussian random variables with variance N_0.

At the receiver, ML (e.g. sphere decoding[72]) or other suboptimal receivers (e.g. MMSE with successive interference cancellation) can be utilized for signal detection as in standard spatially multiplexing MIMO systems since the equivalent channel matrix $\mathbf{HV}'$ is known to the receiver. In order to reduce the complexity burden on the receiver, however, we retain the original simple receiver of the perfect CSI case, since the proposed scheme is intended for the situation where the correlation between the outdated CSI and the current CSI is fairly high. Recall that the major motivation for beamforming is its low complexity. Hence, the left singular vectors $\mathbf{U}'^H$ are used to combine the received vector:

$$\begin{aligned} \underline{z} &= \mathbf{U}'^H \underline{y} \\ &= \mathbf{U}'^H \mathbf{HV}'\mathbf{E}\underline{s} + \mathbf{U}'^H \underline{n} \\ &= \mathbf{GE}\underline{s} + \underline{w}, \end{aligned} \tag{10}$$

where $\mathbf{G}$ is defined as $\mathbf{U}'^H\mathbf{HV}'$, and since $\mathbf{U}'^H$ is unitary, the noise vector $\underline{w}$ has the same statistical properties as $\underline{n}$. The MIMO channel is decomposed into n parallel non-interfering channels if the transmitter knows the true channel matrix $\mathbf{H}$(i.e. $\mathbf{H} = \mathbf{H}'$), and detection can be done separately on different eigenmodes. However, perfect CSI is not available at the transmitter. The treatment of different eigenmodes as scalar channels will allow the preservation of the small complexity that is one of the main motivations for eigenbeamforming. Hence, scalar detection on each element of $\underline{z}$ will be assumed throughout this paper.

The overall strategy is the same as in the single antenna case: the transmitter predicts the error performance(*not* the channel, which would be suboptimal) given the outdated CSI and maximizes the channel throughput based on this information. Therefore, it is first necessary to characterize the error performance given the outdated CSI at the transmitter: $\mathrm{E}_{\mathbf{H}|\mathbf{H}'}[\mathrm{BER}(\mathrm{SINR}_i)|\mathbf{H}'], i = 1, \ldots, m$, where SINR_i stands for the signal to interference and noise ratio on the i-th eigenmode. Note that $\mathbf{H}'$ and its eigen-decomposition are known to the transmitter. Hence, for the sake of neatness, the conditioning on such is dropped for the rest of the paper.

3.2 BER Approximation

First consider the statistics of the matrix $\mathbf{G}$. Define $\underline{g}_i^T$ to be the i-th row of the matrix $\mathbf{G}$, and $g_{ij} \triangleq [\mathbf{G}]_{ij}$. Clearly, $\underline{g}_i$ is a proper complex Gaussian random vector. Also,

$$\begin{aligned}\mathbf{G} &\triangleq \mathbf{U}'^H \mathbf{H} \mathbf{V}' \\ &= \rho \mathbf{U}'^H \mathbf{H}' \mathbf{V}' + \mathbf{U}'^H \boldsymbol{\Sigma} \mathbf{V}' \\ &= \rho \mathbf{D}' + \boldsymbol{\Xi},\end{aligned}$$

where $\boldsymbol{\Xi}$ has the same statistical properties as $\boldsymbol{\Sigma}$, since $\mathbf{U}'$ and $\mathbf{V}'$ are unitary. Let $\{\lambda_i', i = 1, \ldots, m\}$ be the singular values of $\mathbf{H}'$ in decreasing order, and $\underline{\xi}_i^T$ be the i-th row of the matrix $\boldsymbol{\Xi}$; then,

$$\underline{g}_i^T = [0, \ldots, \lambda_i', \ldots, 0]\rho + \underline{\xi}_i^T,$$

and the received signal on the i-th eigenmode is

$$z_i = \underbrace{g_{ii}\sqrt{E_i}s_i}_{signal} + \underbrace{\sum_{j \neq i} g_{ij}\sqrt{E_j}s_j}_{interference} + w_i, \quad i = 1, \ldots, m, \tag{11}$$

where g_{ii} is a complex Gaussian random variable, and each $g_{ij}, i \neq j$, is a zero mean complex Gaussian random variable given the outdated channel matrix. Hence, the interference is Gaussian given the transmitted signal, and the average probability of error for independent data streams can be obtained by averaging an error function over all potential transmitted vectors. Such is readily accomplished for small n. If the eigenmode number m is large, the interference can be considered complex Gaussian according to the central limit theorem(CLT). Thus, assuming uncorrelated unit-variance s_i and s_j, $i \neq j$,

$$\text{SINR}_i = \frac{Y_i E_i}{(1-\rho^2)(E - E_i) + N_0}$$

where E is the total average energy on all of the eigenmodes per channel use, and $Y_i \triangleq |g_{ii}|^2$.

The bit error rate(BER) for square M-ary quadrature amplitude modulation(M-QAM) with Gray coding can be approximated by [26]

$$\text{BER} = 0.2 \exp\left(-\frac{1.5}{M-1}\text{SINR}\right). \tag{12}$$

Therefore, the expected BER of the i-th eigenmode can be calculated as follows,

$$\overline{\text{BER}}_i = 0.2\text{E}_{Y_i}\left\{\exp\left(-\frac{1.5}{M_i-1}\text{SINR}_i\right)\right\},$$

Calculating the expectation yields

$$\overline{\text{BER}}_i = \frac{0.2\exp\left[-\frac{\rho^2\lambda_i'^2}{2(1-\rho^2)}(1-\frac{1}{A_i})\right]}{A_i}, i=1,\ldots,m, \quad (13)$$

where

$$A_i = 1 + \frac{3E_i(1-\rho^2)}{(M_i-1)[(1-\rho^2)(E-E_i)+N_0]}, \quad i=1,\ldots,m. \quad (14)$$

If $\rho = 1$, there is no interference from other eigenmodes and $Y_i = \lambda_i'^2$ with probability 1. In this case, the approximate expected BER is

$$\overline{\text{BER}}_i = 0.2\exp\left(-\frac{1.5}{M_i-1}\frac{\lambda_i'^2 E_i}{N_0}\right). \quad (15)$$

3.3 Adaptive MQAM

It is shown in [42] that, without temporal power loading, waterfilling over the outdated channel is very close to optimal in terms of channel capacity when $\rho = 0.9$. In the design here, temporal power adaption is just "power pruning" motivated by [17], i.e. extra power in one time slot is allocated to the next time slot, and the rate for the specific channel matrix is maximized with the given average power at the moment.

The optimization problem is expressed as:

$$\text{maximize} \sum_i \log_2 M_i \quad (16)$$
$$\text{subject to } \overline{\text{BER}}_i \le P_b, \text{and} \sum_i E_i \le E$$

where P_b is the prescribed BER, and M_i is the size of the constellation on the i-th eigenmode. Notice that spatial power and bit loading in eigenbeamforming transmission with perfect CSI are very similar to that in multi-carrier communication systems; hence, optimal solutions exist when the transmitter has perfect CSI. Manipulating (15) yields

$$M_i = 1 - \frac{1.5\lambda_i'^2 E_i}{\ln(5P_b)N_0}, i = 1, \ldots, m \tag{17}$$

Assuming the energy and constellation size are continuous, the optimal spatial power loading is

$$E_i^* = \left(\nu - \frac{-2\ln(5P_b)}{3}\frac{N_0}{\lambda_i'^2}\right)^+, i = 1, \ldots, m \tag{18}$$

where ν is a constant that satisfies $\sum E_i^* = E$ and $(x)^+$ means $\max(0, x)$. Note that "waterfilling" over eigenmodes throughout this paper means (18) if not in the information theory context. Substituting the optimal energy value in (17) yields the optimal continuous constellation size, which is then rounded to a discrete constellation size. Note that this rounding scheme is not the optimal discrete bit loading solution. If the constellation size on one eigenmode is rounded to 0, the energy on it will be put on other eigenmodes. Per above, "power pruning" as described in [17] is then applied; the extra energy from this channel use will be added to that for the next channel use. The algorithm for the case of perfect feedback can be summarized as following steps:

1. *Allocate the energy(sum of the energy from last time and the average energy) according to (18).*
2. *Calculate continuous constellation sizes on all eigenmodes according to (17).*
3. *Reallocate the energy from the higher-ordered eigenmodes for which the rounded constellation is of size zero to the remaining lower-ordered eigenmodes equally.*
4. *Calculate continuous constellation sizes again on the remaining eigenmodes with the new power loading according to (17).*
5. *Round the continuous constellation sizes to the nearest discrete constellation sizes and obtain the minimum energy values to maintain the prescribed BER with the selected discrete constellation sizes.*

6. *Save the extra energy for the next use of the channel.*

Due to the complicated relationship between power, BER and M, the optimal spatial power and bit loading in the delayed feedback case is very difficult to derive. Existing algorithms derived for different frameworks for the problem either just allocate equal power on all active eigenmodes[63], or use a high-complexity brute-force searching method to do joint power and bit loading [62]. Here, initial spatial power loading is "waterfilling" over the active eigenmodes assuming the CSI at the transmitter is perfect. Next, with knowledge of the power on an eigenmode, one can select the maximum constellation size that keeps the BER below the prescribed value. Because M_i cannot be written as an explicit function of E_i, this has to be done by off-line numerical search and stored in a lookup table. The constellation size of the first eigenmode is determined first. Then the minimum energy to support the required BER is obtained numerically. Extra energy is reallocated to the next eigenmode. This process continues until power and bit loading is done on all eigenmodes. Because numerical results show that it works well, (13) is used regardless of the number of eigenmodes. Similar to the perfect CSI feedback case, if 0-QAM is assigned to an eigenmode, no power is transmitted and the total power will be waterfilled to other lower-ordered eigenmodes, still according to the outdated CSI. The extra energy from this channel use will be used for the next channel use. The initial number of active eigenmodes is m. After the above procedure is done, the total bit number is recorded and the weakest eigenmode among the active eigenmodes is removed from the active eigenmode set. Then this procedure is repeated until only one eigenmode is active. The number of active eigenmodes that maximize the total bit number, as well as the corresponding power and bit loading, is chosen. The algorithm contains the following steps:

1. *Select the active eigenmode number $N = m$(m is the maximum number of eigenmodes).*

2. *Waterfill the total energy E (the average energy plus extra energy from the last time of channel use) over these N eigenmodes.*

3. *Use the BER approximation formula (13) to load bits on the l-th eigenmode(initial value $l = 1$) so that the BER is below the prescribed value.*

4. *Calculate the minimum energy to keep the required BER with the selected constellation size.*

5. *if $l = N$, go to Step 6; otherwise $l = l + 1$, add the extra energy from Step 4 to the l-th eigenmode and go to Step 3.*

6. *Subtract the extra power on the last eigenmode from the total power E, and repeat Step 2-5 to do bit loading again.*

7. *Add the extra power from the first and second round of bit loading. This extra energy will be allocated to the next channel use.*

8. *Sum the bit number on all eigenmodes, and store this value.*

9. *$N = N - 1$. If $N = 0$, go to Step 10; Otherwise, go to Step 2.*

10. *Select the number of eigenmodes that supports the maximum number of bits. The corresponding bit and power loading is also selected.*

Note that Step 6 is used to consider the cross interference between the eigenmodes.

3.4 Numerical results

A 4×4 MIMO system is used in the simulation. In the simulation, the variance of the complex AWGN noise is normalized to 1, and hence SNR manipulation is done via the average transmitted power setting. The proposed scheme will be compared with two schemes. One is the "delay disregarding" scheme which assumes the transmitter CSI is perfect. Thus, the power and bit loading are done exactly like in the $\rho = 1$ case except that power pruning is not done. The "energy margin" scheme also employs design equations with $\rho = 1$, but now a link margin is employed to compensate for the imperfect CSI. In other words, the $\rho = 1$ equations are used with a more conservative energy substituted until the BER requirement is met.

First, consider the accuracy of the BER approximation formula as shown in Figure 8. The BER depends on the high-dimension $\mathbf{H}$, so it is impossible to verify every case. Thus, one specific realization of the outdated channel matrix is randomly generated and fixed during the simulation. The constellations are 16-QAM, QPSK, PSK and PSK on eigenmode 1, 2, 3, 4, respectively, and the corresponding transmit SNRs are 14.7dB, 13.0dB, 11.8dB and 11.8dB. For other figures, the average total transmit SNR is 17.8dB, and the available constellation sets for adaptive modulation are square M-QAM with $M \in \{0, 2, 4, 16, 64, 256\}$. Figure 9 indicates that the proposed scheme meets the prescribed BER 10^{-3} while the delay disregarding scheme greatly misses the target. It is shown in Figure 11 that the proposed scheme provides higher data rate than the "energy margin" scheme with similar error performance. Not surprisingly, the data rate gain is even higher than that in [17].

4 Conclusions

Wireless communications systems form a critical part of the communications infrastructure, particularly in disaster situations, but multipath fading makes communication over wireless channels challenging. One promising method of improving performance over such channels is to employ channel state information at the transmitter to tune the transmitted signal to the channel state. Such an approach has not only been widely studied in academic circles but also has recently found penetration into numerous standards. However, the channel state information available at the transmitter is inherently outdated and noisy, which makes system design challenging. In this paper, we have reviewed and extended an approach to this problem where statistical information about the current channel state based on the available measurements is explicitly exploited. For single-antenna systems, employing such an approach rather than simply relying on extra link margin to compensate for such has been shown to result in significant gains. These results were then extended to the multiple-antenna case, where the loss due to inaccuracy in the channel measurements is higher due to the resulting cross-talk between transmission modes. However, the available gains are also higher, and an algorithm has been presented to efficiently achieve such. Numerical results support this analytic line of thought.

There is considerable interest to extend the line of research of this paper to the multiple-antenna broadcast case, where the gains of transmitter CSI are enormous even from a Shannon capacity point of view. In particular, unlike the point-to-point MIMO channel, transmitter CSI in the broadcast case provides an effective increase in the degrees of freedom, and thus it is of interest to consider both how these degrees of freedom decay as CSI becomes outdated and/or noisy and how to efficiently signal in such a scenario.

References

[1] I. Kalet, "The multitone channel." *IEEE Trans. Commun.*, vol. 37, no. 2, pp. 119–124, Feb. 1989.

[2] P. Chow, J. Cioffi, and J. Bingham, "A practical discrete multitone transceiver loading algorithm fordata transmission over spectrally shaped channels," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 773–775, Feb.-Apr. 1995.

[3] J. Hayes, "Adaptive Feedback Communications," *IEEE Trans. Commun.*, vol. 16, no. 1, pp. 29–34, Feb. 1968.

[4] J. Cavers, "Variable-rate transmission for rayleigh fading channels," *IEEE Trans. Commun.*, vol. 20, no. 1, pp. 15–22, Feb. 1972.

[5] ETSI EN 301 709, *Digital Cellular Telecommunication System (Phase 2+); Link Adaptation*, Std., Dec. 1999.

[6] TIA/EIA/IS-95-B, *Mobile Station-Base Station Compatibility Standard for Wideband Spread Spectrum Cellular Systems*, Telecommunications Industry Association Std., Mar. 1999.

[7] S. Nanda, K. Balachandran, and S. Kumar, "Adaptation techniques in wireless packet data services," *IEEE Commun. Mag.*, vol. 38, no. 1, pp. 54–64, Jan. 2000.

[8] 3GPP2 C.S0002-D, *Physical Layer Standard for cdma2000 Spread Spectrum Systems*, 3rd Generation Partnership Project 2(3GPP2) Std., Sept. 2005.

[9] 3GPP TS 25.212, *Technical Specification Group Radio Access Netork; Multiplexing and channel coding(FDD)*, 3rd Generation Partnership Project(3GPP) Std., June 2006.

[10] CTWTS TSM 03.64 v3.0.0, *TD-SCDMA Packet Radio Service(TD-PRS); Overall Description of TD-PRS Radio Interface*, China Wireless Telecommunicaton Standard(CWTS) Std., Aug. 2002.

[11] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, High-speed Physical Layer in the 5 GHz band*, IEEE Std. 802.11a, 1999.

[12] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Amendment 2: Higher-speed Physical Layer(PHY) extension in the 2.4 GHz Band*, IEEE Std. 802.11b, 2001.

[13] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Amendment 4:Further Higher Data Rate Extension in the 2.4 GHz Band*, IEEE Std. 802.11g, 2003.

[14] *High Throughput Physical Layer (PHY) Specification, v1.27*, Enhanced Wireless Consortium Working Draft Proposed Standard, 2005. [Online]. Available: http://www.enhancedwirelessconsortium.org

[15] *Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Std. 802.16, 2004.

[16] R. Laroia, S. Uppala, and J. Li, "Designing a mobile broadband wireless access network," *IEEE Signal Processing Mag.*, vol. 1053, no. 5888/04, Sept. 2004.

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

[18] J. Proakis, *Digital Communications, 4-th Edition*. New York: McGraw-Hill, 2001.

[19] A. Bahai and B. Saltzberg, *Multi-Carrier Digital Communications: Theory and Applications of Ofdm*. Springer, 1999.

[20] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[21] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 2007–2019, Sept. 1999.

[22] G. Caire, G. Taricco, E. Biglieri, and D. di Elettronica, "Optimum power control over fading channels," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1468–1489, July 1999.

[23] T. Cover and J. Thomas, *Elements of information theory*. New York: Wiley, 1991.

[24] H. Viswanathan, "Capacity of Markov channels with receiver CSI and delayed feedback," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 761–771, 1999.

[25] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, Oct. 1998.

[26] A. Goldsmith and S. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.

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

[28] P. Ormeci, X. T. Liu, D. L. Goeckel, and R. D. Wesel, "Adaptive bit-interleaved coded modulation," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1572–1581, Sept. 2001.

[29] D. Goeckel, "Adaptive coding for fading channels using outdated channelestimates," in *Proc. Veh. Technol. Conf. (VTC)*, vol. 3, May 1998.

[30] T. Yang, A. Duel-Hallen, and H. Hallen, "Reliable adaptive modulation aided by observations of another fading channel," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 605–611, Apr. 2004.

[31] S. Falahati, A. Svensson, T. Ekman, and M. Sternad, “Adaptive modulation systems for predicted wireless channels,” *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 307–316, Feb. 2004.

[32] X. D. Cai and G. B. Giannakis, “Adaptive psam accounting for channel estimation and prediction errors,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 1, pp. 246–256, Jan. 2005.

[33] I. S. Gradshteyn and I. M. Ryzhik, *Table of integreals, series and products*. New York: Academic press, 1980.

[34] E. Feig, “Practical aspects of DFT-based frequency division multiplexing fordata transmission,” *Communications, IEEE Transactions on*, vol. 38, no. 7, pp. 929–932, July 1990.

[35] 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, Sept. 2001.

[36] D. Goeckel, “Robust Adaptive Coded Modulation for Time-Varying Channels with Delayed Feedback,” in *Proc. 35th Annu. Allerton Conf. Communication, Control, and Computing*, Sept. 1997.

[37] I. Telatar, “Capacity of multi-antenna Gaussian channels,” *Euro. Trans. on Telecom.*, vol. 10, no. 6, pp. 585–595, 1999.

[38] G. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multiple antennas,” *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[39] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, “Capacity limits of MIMO channels,” *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 684–702, June 2003.

[40] D. Love, R. Heath Jr, W. Santipach, and M. Honig, “What is the value of limited feedback for MIMO channels?” *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54–59, Oct. 2004.

[41] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2632–2639, Sept. 2001.

[42] S. Venkatesan, S. Simon, and R. Valenzuela, "Capacity of a Gaussian MIMO channel with nonzero mean," in *Proc. IEEE 58th Vehicular Technology Conference*, vol. 3.

[43] S. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, 2004.

[44] S. Simon and A. Moustakas, "Optimality of beamforming in multiple transmitter multiple receiver communication systems with partial channel knowledge," in *Proc. of DIMACS Workshop on Signal Processing Wireless Communications*, DIMACS, Rutgers Univ.

[45] A. Narula, M. Lopez, M. Trott, G. Wornell, M. Inc, and M. Mansfield, "Efficient use of side information in multiple-antenna datatransmission over fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, 1998.

[46] F. Rey, M. Lamarca, and G. Vazquez, "Robust power allocation algorithms for MIMO OFDM systems with imperfect CSI," *IEEE Trans. Signal Processing*, vol. 53, no. 3, pp. 1070–1085, Mar. 2005.

[47] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channelsusing the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.

[48] A. Scaglione, P. Stoica, S. Barbarossa, G. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[49] D. Palomar, J. Cioffi, and M. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Processing*, vol. 51, no. 9, pp. 2381–2401, Sept. 2003.

[50] L. Collin, O. Berder, P. Rostaing, and G. Burel, "Optimal minimum distance-based precoder for MIMO spatial multiplexing systems," *IEEE Trans. Signal Processing*, vol. 52, no. 3, pp. 617–627, Mar. 2004.

[51] S. Zhou and G. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel mean feedback," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2599–2613, Oct. 2002.

[52] G. Jongren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 611–627, 2002.

[53] H. Sampath and A. Paulraj, "Linear precoding for space-time coded systems with known fading correlations," *IEEE Commun. Lett.*, vol. 6, no. 6, pp. 239–241, June 2002.

[54] Z. Hong, K. Liu, R. Heath Jr, and A. Sayeed, "Spatial multiplexing in correlated fading via the virtual channel representation," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 856–866, June 2003.

[55] S. Zhou and G. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel correlations," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1673–1690, July 2003.

[56] M. Vu and A. Paulraj, "Optimal linear precoders for mimo wireless correlated channels with non-zero mean in space-time coded systems,."

[57] D. Love, R. Heath Jr, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[58] D. Palomar, M. Lagunas, and J. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Trans. Signal Processing*, vol. 52, no. 5, pp. 1179–1197, May 2004.

[59] D. Palomar and S. Barbarossa, "Designing MIMO communication systems: constellation choice and linear transceiver design," *IEEE Trans. Signal Processing*, vol. 53, no. 10, pp. 3804–3818, Oct. 2005.

[60] S. Zhou and G. Giannakis, "Adaptive modulation for multiantenna transmissions with channel mean feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1626–1636, Sept. 2004.

[61] P. Xia, S. Zhou, and G. Giannakis, "Adaptive MIMO-OFDM based on partial channel state information," *IEEE Trans. Signal Processing*, vol. 52, no. 1, pp. 202–213, Jan. 2004.

[62] S. Bergman, C. Martin, and B. Ottersten, "Bit and power loading for spatial multiplexing using partial channel state information," in *Proc. ITG Workshop on Smart Antennas*, Mar. 2004, pp. 152–159.

[63] S. H. Ting, K. Sakaguchi, and K. Araki, "A robust and low complexity adaptive algorithm for mimo eigenmode transmission system with experimental validation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 7, pp. 1775–1784, July 2006.

[64] G. Jongren and M. Skoglund, "Quantized feedback information in orthogonal space-time block coding," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2473–2486, Oct. 2004.

[65] V. Lau, Y. Liu, and T. Chen, "On the design of MIMO block-fading channels with feedback-link capacity constraint," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 62–70, Jan. 2004.

[66] ——, "Capacity of memoryless channels and block-fading channels with designable cardinality-constrained channel state feedback," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 2038–2049, Sept. 2004.

[67] K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.

[68] W. Dai, Y. Liu, and B. Rider, "Quantization bounds on grassmann manifolds and applications to mimo systems," *IEEE Trans. Inform. Theory*, submitted for publication.

[69] W. Santipach and M. L. Honig, "Asymptotic capacity of beamforming with limited feedback," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, Chicago, Illinois, June 2004.

[70] ——, "Capacity of beamforming with limited training and feedback," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, Seattle, Washington, July 2006.

[71] A. Dabbagh and D. Love, "Feedback rate-capacity loss tradeoff for limited feedback MIMO systems," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 2190–2202, May 2006.

[72] O. Damen, A. Chkeif, and J. Belfiore, "Lattice code decoder for space-time codes," *IEEE Commun. Lett.*, vol. 4, no. 5, pp. 161–163, May 2000.

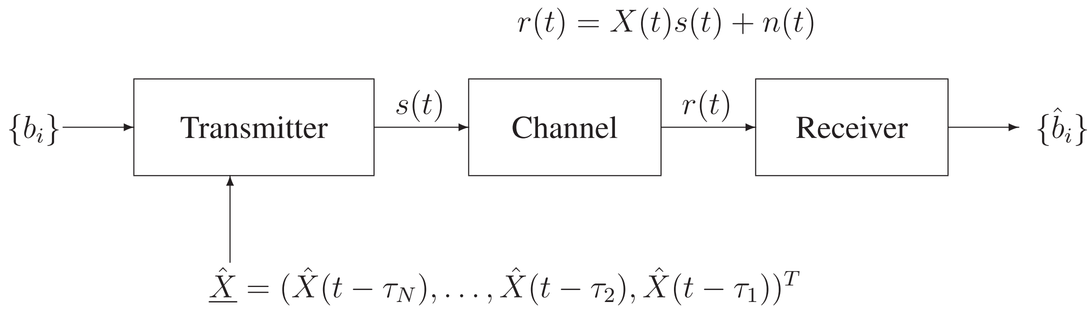

Figure 1: System Diagram.

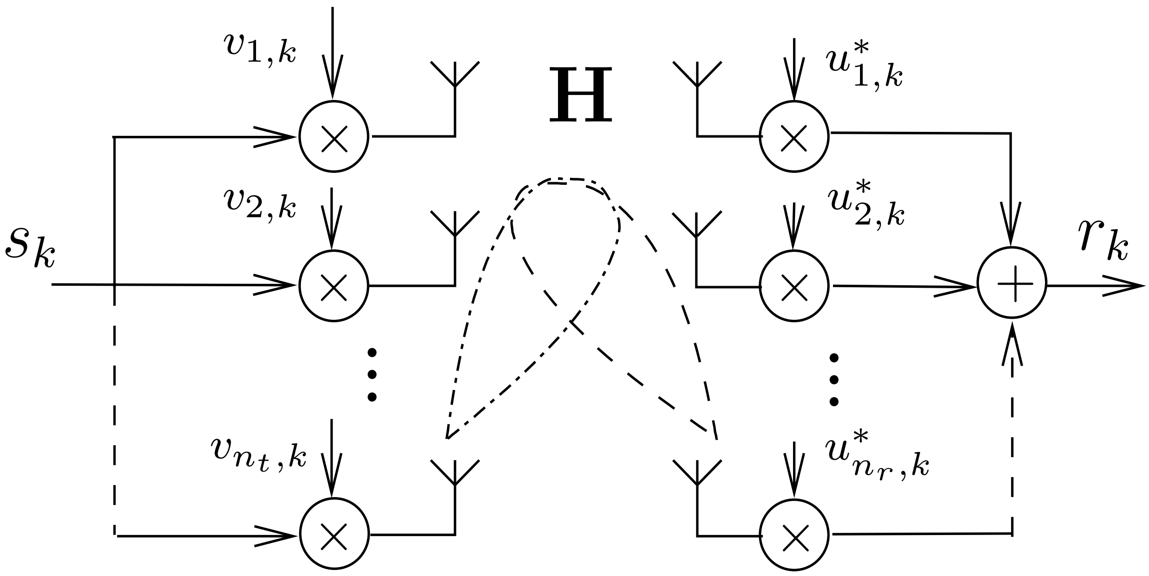

Figure 2: Illustration of the k-th subchannel in MIMO eigenbeamforming.

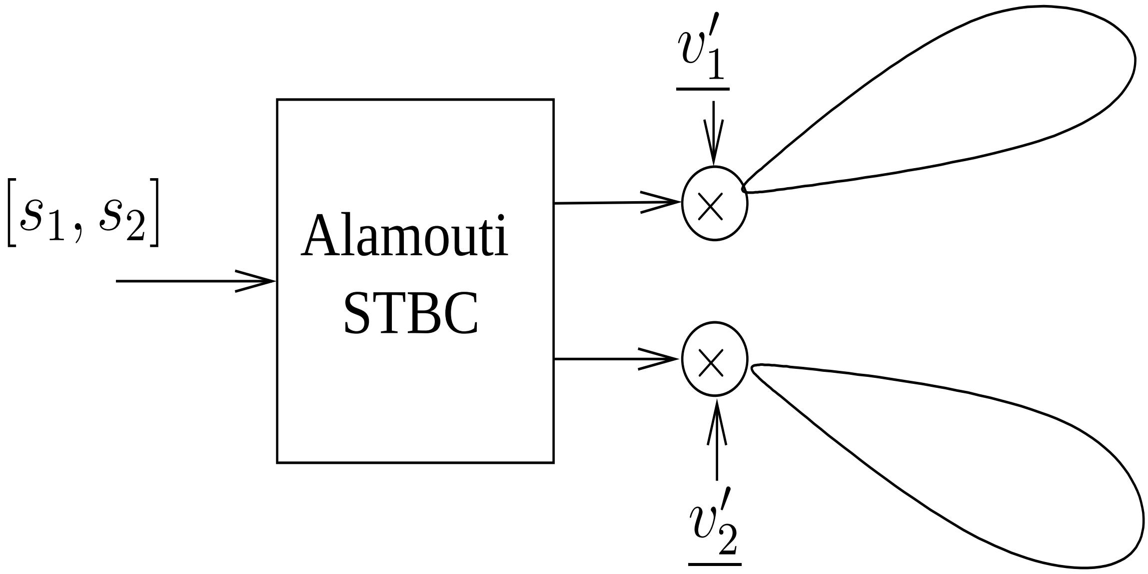

Figure 3: STBC and beamforming are cascaded to make use of imperfect CSIT.

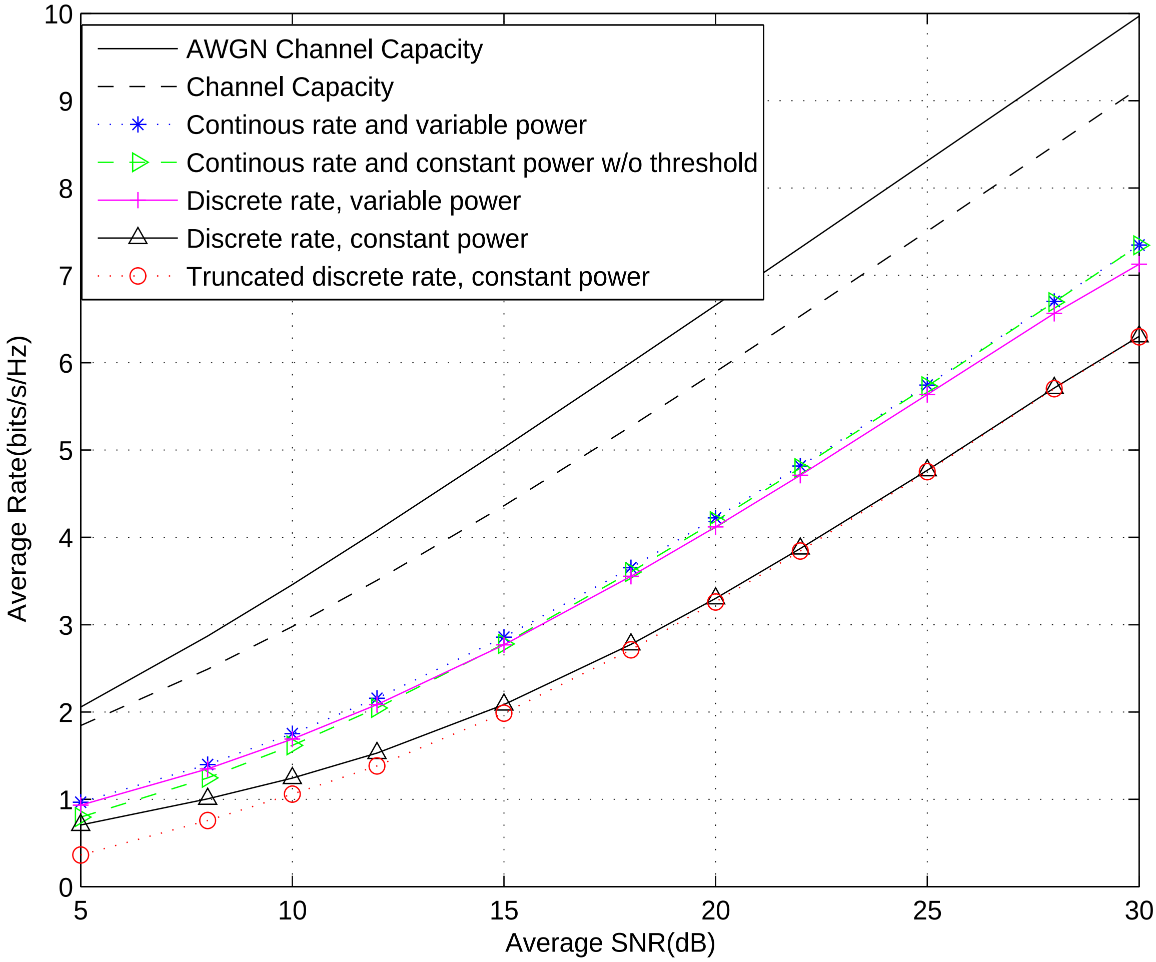

Figure 4: Spectral efficiencies of various adaptive uncoded QAM designs with perfect CSIT in single-antenna systems

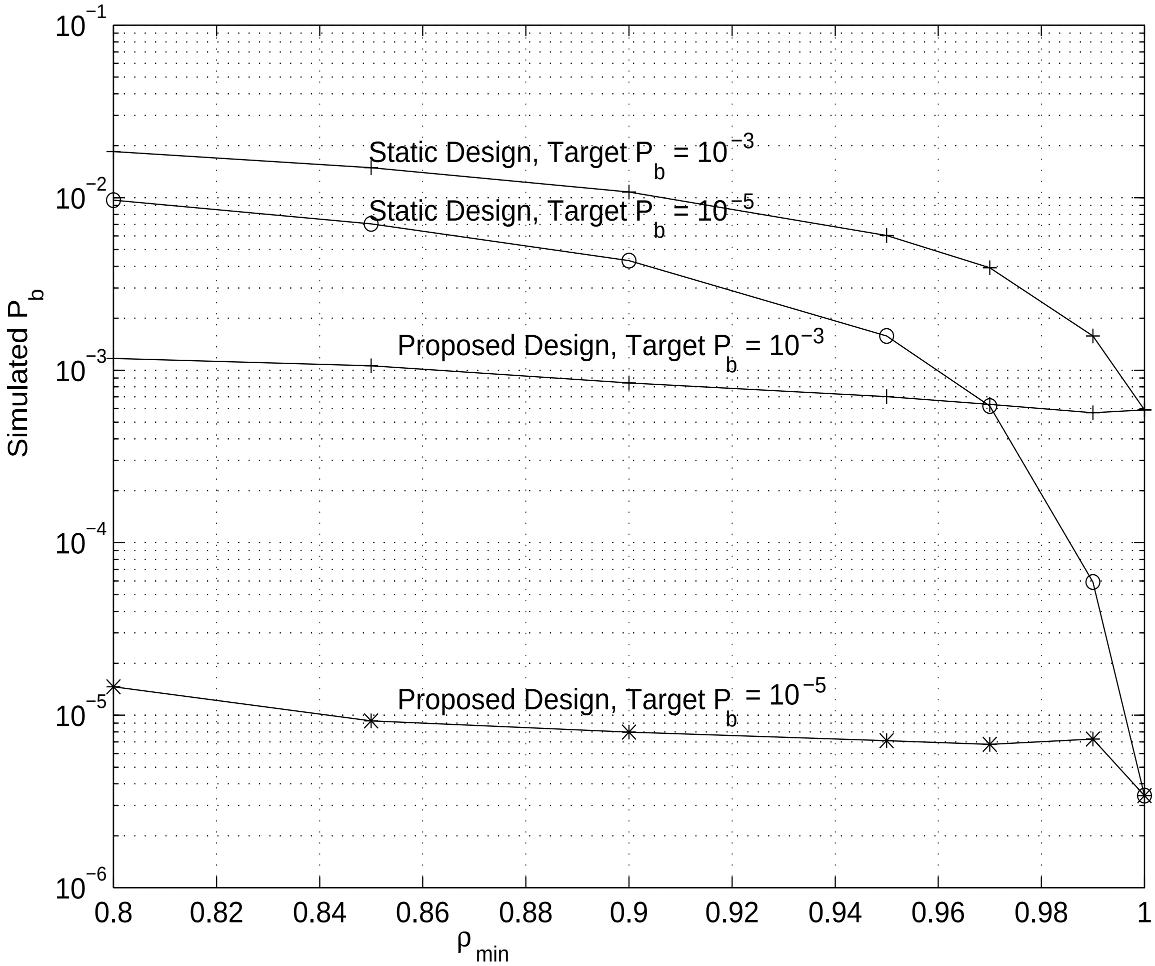

Figure 5: Simulated BER of adaptive uncoded QAM design schemes versus ρ_{min} at average SNR = 15dB [17].

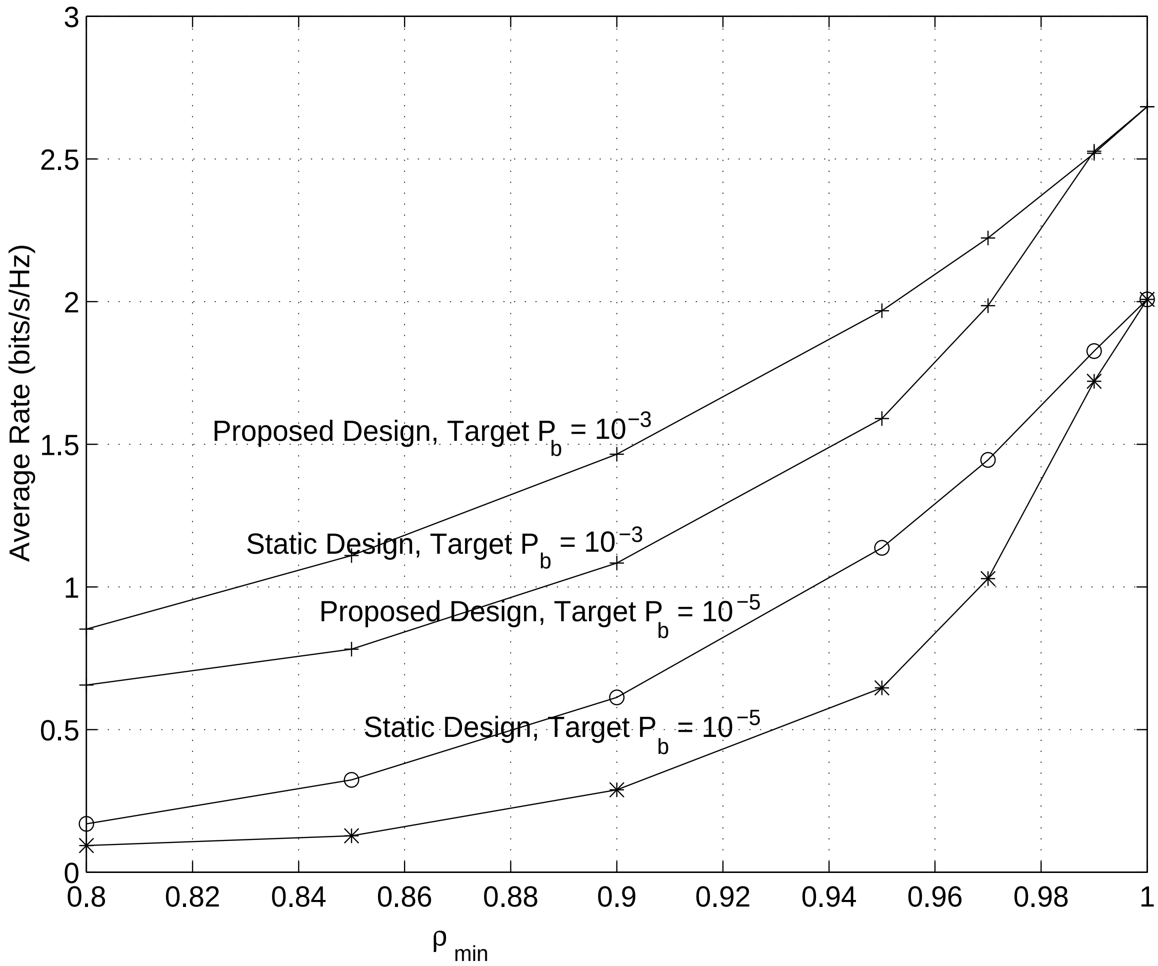

Figure 6: Average data rates of adaptive uncoded QAM design schemes versus ρ_{min} at average SNR = 15dB [17].

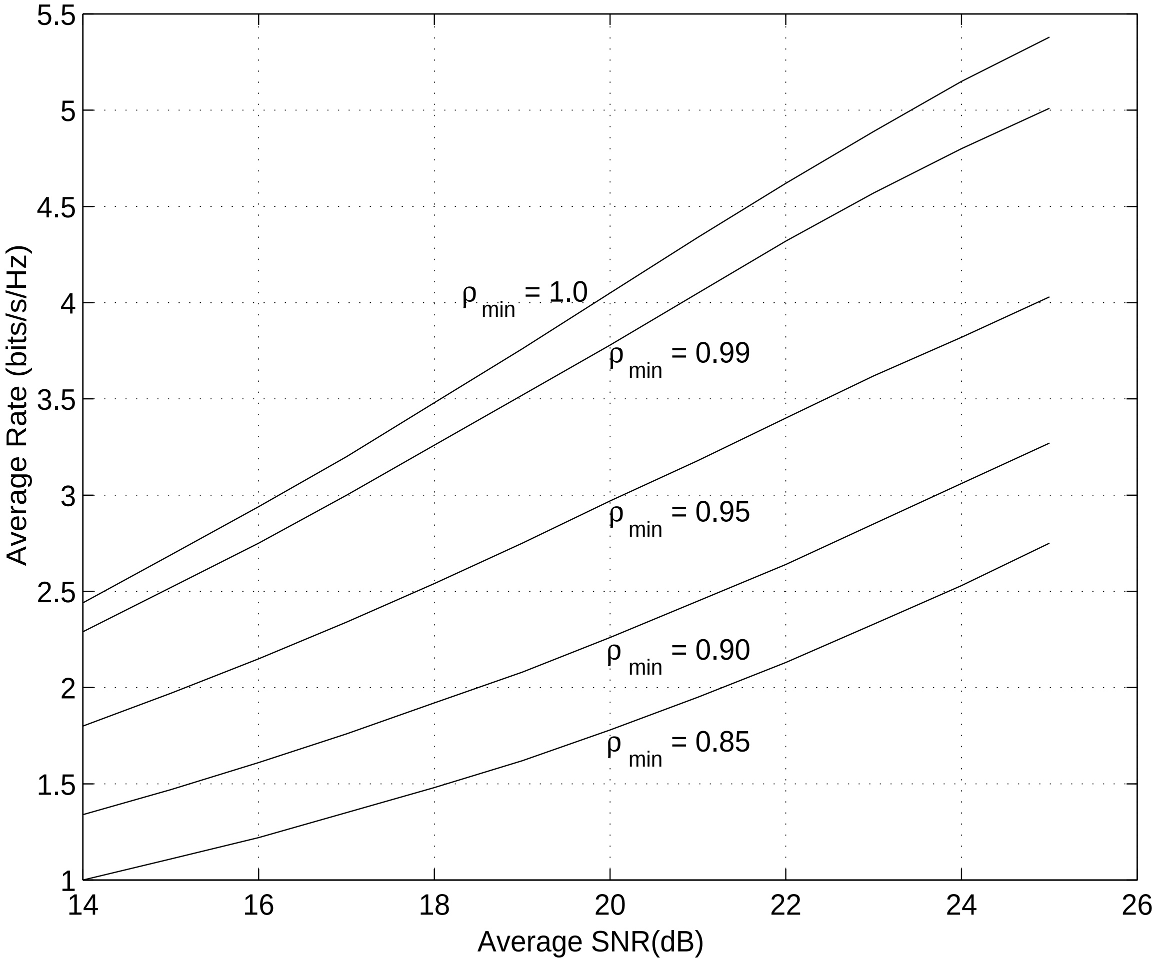

Figure 7: Average data rates versus average SNR at design BER equal to 10^{-3} [17].

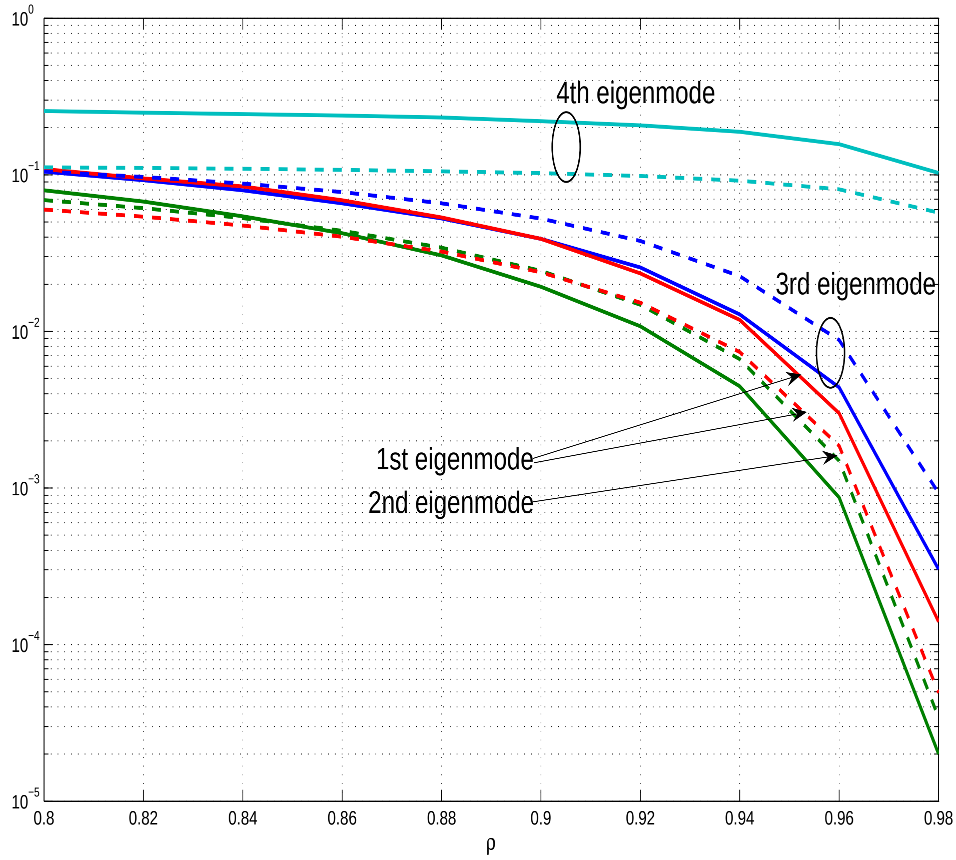

Figure 8: Evaluation of the BER approximation formula. 10^4 channel uses are utilized in the simulation to generate the curves. Dashed lines are from the formula, and solid lines are from the simulation. Note the BER calculated from the formula are in the same order as the simulated ones. More importantly, the formula are useful for design, as demonstrated in succeeding figures.

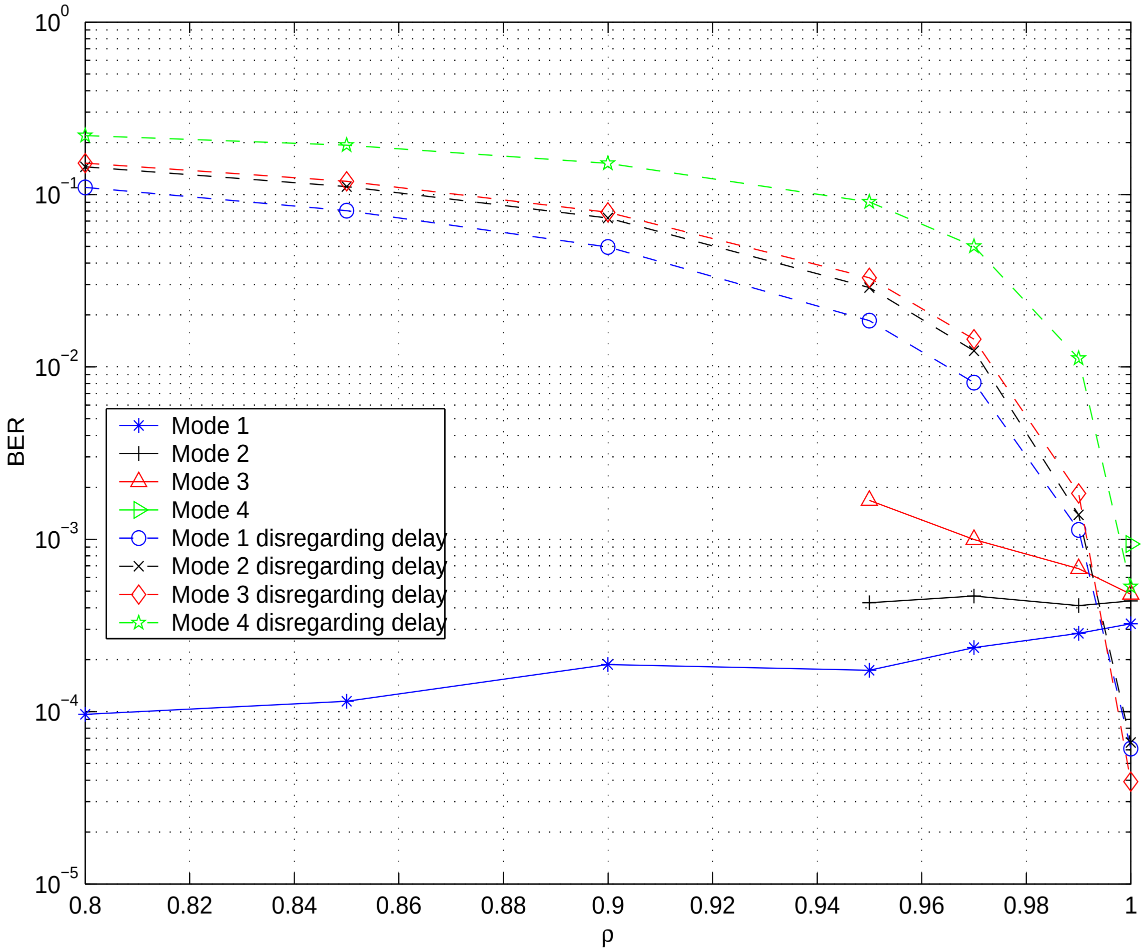

Figure 9: Comparison of the two schemes. The top four curves are from the delay-disregarding scheme. The four curves in the bottom come from the proposed scheme. Due to the small number of bits transmitted on the higher-ordered eigenmodes for small ρ values, BERs of those eigenmodes are not available. We can see the scheme disregarding the delay misses the BER target 10^{-3}, while BERs in the proposed scheme are below 10^{-3}. Also, note that the 4th eigenmode is never used in our scheme when $\rho < 1$.

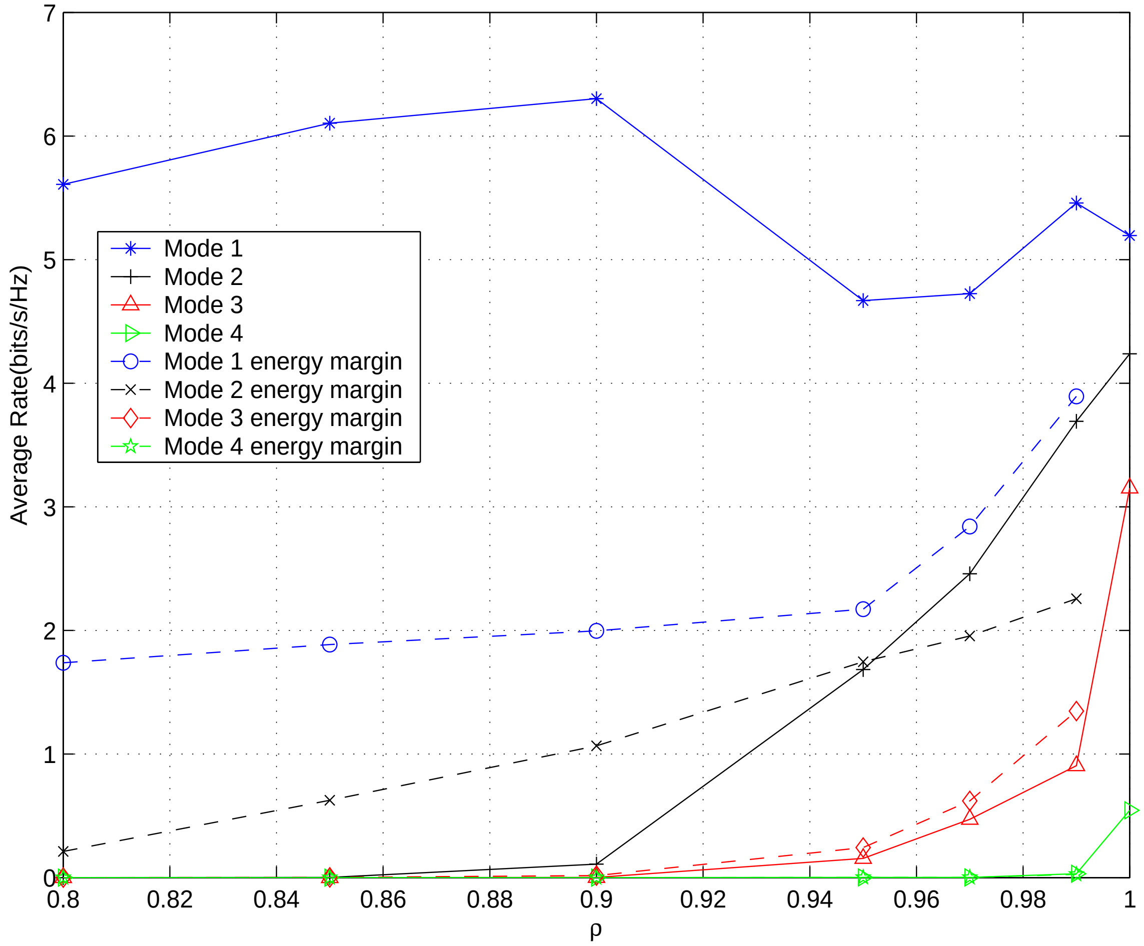

Figure 10: Date rates on different eigenmodes. Solid lines are the proposed scheme, and dashed lines are the "energy margin" scheme. The 4th eigenmode carries very little throughput for either scheme, even when the ρ is fairly large. The decline in the date rate on the first eigenmode is because power is poured to other eigenmodes when the CSI gets better. It is observed that using only the strongest eigenmode(beamforming) gives negligible rate loss when $\rho < 0.9$.

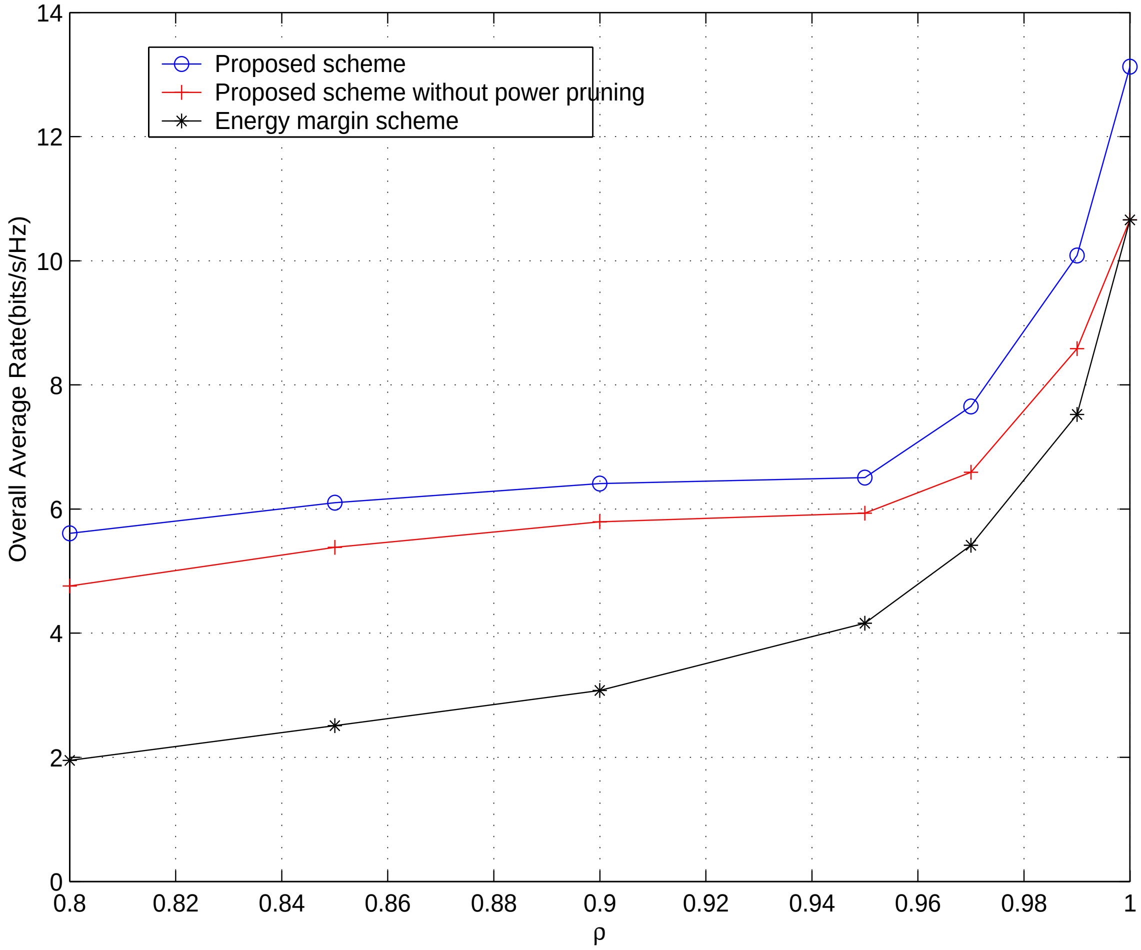

Figure 11: Comparison of overall date rates of different schemes. This figure clearly shows that the proposed scheme is superior when $\rho < 0.95$. The rate of the proposed scheme is almost 4 bits/s/Hz more than the energy margin scheme, and 1 bit/s/Hz more than the scheme without power pruning.