Variable-Rate Transmission for MIMO Time-Correlated Channels With Limited Feedback

Yuan-Pei Lin, *Senior Member, IEEE*

***Abstract*—In this paper we consider variable-rate transmission for time-correlated MIMO (multi-input multi-output) channels with limited feedback. The number of bits loaded on each subchannel of the MIMO system is dynamically assigned according to the current channel condition and fed back to the transmitter. As the channel is time-correlated, bit loading is a vector signal that is also time-correlated. We propose to feedback bit loading using predictive coding, which is known to be a powerful quantization technique when the underlying signal is correlated in time. Assuming the channel is a first-order Gauss-Markov random process, we derive the predictor for the bit loading to be coded and analyze the corresponding prediction error variance when the channel is varying slowly. By exploiting the prediction error variance, we adapt the quantizer of the prediction error to have a smaller quantization error. Furthermore we show that the prediction error variance is proportional to a term that depends only on the time-correlation coefficient. This leads to the conclusion that, a codebook that is designed for a particular time correlation coefficient can be easily modified to a codebook for a different correlation coefficient without redesign. Simulations are presented to demonstrate that the proposed predictive coding can achieve a very good approximation of the desired transmission rate with a very low feedback rate.**

***Index Terms*—MIMO, limited feedback, quantization of bit loading, variable rate transmission.**

I. Introduction

It has been shown that with limited amount of feedback information the performance of a transmission system can be enhanced greatly [1]. The transmission rate or error rate can be significantly improved when there is feedback from the receiver through a reverse channel. In general, the transmitter has no knowledge of the forward link channel and only the receiver has the channel state information. As the reverse channel can support only a limited transmission rate, it is desirable to have a low feedback rate.

Manuscript received July 13, 2013; revised November 14, 2013 and March 08, 2014; accepted July 14, 2014. Date of publication August 06, 2014; date of current version August 27, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Zhengdao Wang. This work was supported in part by the National Science Council, Taiwan, R.O.C. under Grant NSC 101-2221-E-009-091-MY3 and by the International Research Center of Excellence on Advanced Bioengineering under Grant NSC99-2911-I-009-101.

The author is with the Department of Electrical Engineering, National Chiao-Tung University, Hsinchu 300, Taiwan (e-mail: ypl@mail.nctu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2014.2345633

Various feedback schemes have been proposed for the case when the transmission rate is fixed. The feedback of precoder for spatial multiplexing MIMO systems with a fixed transmission rate has been extensively investigated [1]–[6]. The receiver chooses the optimal precoder from a codebook and sends the index back to the transmitter. Optimal codebook designs of unitary precoders using Grassmannian subspace packing for different criteria are developed in [2]. The optimal unitary precoder for minimizing BER (bit error rate) using infinite feedback rate is given in [4] and generalized Lloyd algorithm is used for constructing codebooks. In multimode precoding [3], the number of substreams can vary with the channel. The use of an identity precoder combined with the feedback of only bit loading is proposed in [5]. Quantization of bit loading using vector quantization is considered in [6] to reduce feedback rate for systems that feedback only bit loading.

For fading channels, adapting the transmission rate according to the channel state information can lead to considerable gain over fixed-rate systems [7]. Rate adaptation is also important for controlling frame error rate or packet error rate without using deep interleaving because we can adjust the rate to satisfy a given error rate constraint [8]. MIMO systems with variable transmission rate have been considered in the literature, e.g., [9]–[15]. In [9], the system switches between two transmission rates by using beamforming for low-rate transmission and spatial multiplexing for a higher rate, depending on the current channel condition. Switching between orthogonal space time block code and spatial multiplexing is proposed in [10] to improve the spectral efficiency. Spatial modulation for MIMO systems with adaptive constellation selection is considered in [11]. Feedback of quantized precoder, power allocation and/or bit allocation is considered in [12], [13]. Successive quantization of power and bit allocation is proposed in [14]. An efficient algorithm for per antenna power and rate control is developed in [15].

In practical transmission environments, a fading channel is usually correlated in time. Temporal correlation is considered in [16] to design the transmit beamforming vector by modeling the quantized channel as a finite-state Markov chain. Differential feedback of the channel Gram matrix based on geodesic curves is considered in [17]. When the channel is modeled as a first-order Gauss-Markov random process, the time correlation can be more directly exploited for analysis or for more efficient feedback [18]–[22]. For a given quantization error con-

1053-587X © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

straint, the minimum rate for feeding back differential channel information is derived in [18]. The design of polar-cap differential codebooks is addressed in [19] for beamforming systems. A feedback scheme that uses a differential rotation of the precoder is proposed in [20]. Based on random matrix quantization analysis, a spherical cap codebook can be generated systematically. For a given average feedback rate, the problem of choosing a feedback interval is addressed for beamforming systems in [21], [22]. The problem of optimal feedback interval for minimizing capacity loss is formulated in [21]. Comparison between frequent low resolution feedback and infrequent high resolution feedback is done in [22] by analyzing SNR loss due to quantization. Earlier results that exploit the time correlation of the MIMO channel are all for fixed-rate transmission systems to the best of our knowledge.

In this paper, we consider variable-rate transmission for limited-feedback MIMO systems over time-correlated channels using predictive quantization of bit loading (PQB). The transmission rate is adapted to the current channel condition by dynamically assigning bits to the subchannels of the MIMO systems. The bit loading, a vector signal, is fed back using predictive quantization, a coding scheme known to be very efficient for signals that are correlated in time.[1] Assuming the channel is modeled by a first-order Gauss-Markov process, we derive the predictor of the current bit loading given the previous channel and then derive the prediction error variance. We show that the predictor and prediction error variance can be approximated based on the previous bit assignment. Thus, without using extra feedback, we can quantize the prediction error based on its statistics so that a smaller quantization error can be achieved. It turns out that the prediction error variance is proportional to a term that depends only on the time correlation coefficient. Therefore, after we design a codebook for a particular time correlation coefficient, we can easily convert it to a codebook for a different correlation coefficient without redesign. Furthermore, we show how the derivation of predictor and analysis of quantization error can be modified for the case when the feedback rate is as low as one bit and interleaving of prediction error is needed for feedback. Simulations are presented to demonstrate that the proposed predictive coding of bit loading can achieve a very good performance for Gauss-Markov channels as well as more realistic time-correlated channels.

The sections are organized as follows. In Section 2, we give the system model of the time-correlated MIMO system. In Section 3., we derive the predictor of the current bit loading using past channel information. Predictive closed-loop quantization of bit loading is presented in Section 4. Simulation examples are given in Section 5 and a conclusion given in Section 6.

Notation: Boldfaced lower case letters represent vectors and boldfaced upper case letters are reserved for matrices. The notation $\mathbf{A}^\dagger$ denotes transpose-conjugate of $\mathbf{A}$ and $[\mathbf{A}]_{ij}$ denotes the (i,j)th entry of $\mathbf{A}$. The function $E[x]$ denotes the expected value of a random variable x.

[1]As the feedback information is bit loading, the antennas corresponding to subchannels loaded with zero bits are not used. Hence it can be considered a generalized antenna selection scheme, in which the active antennas can be loaded with different number of bits.

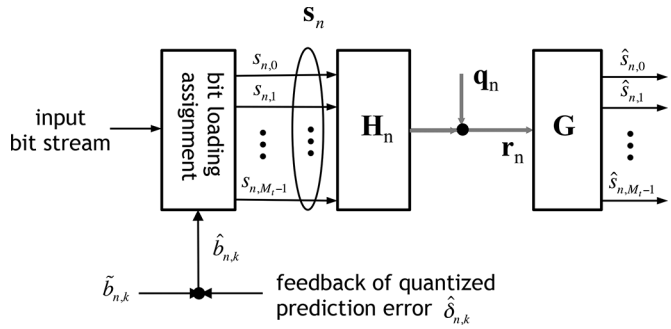


Fig. 1. A MIMO system with bit loading feedback.

II. System Model

Consider the wireless system with M_t transmit antennas and M_r receive antennas in Fig. 1. At time n, the channel is modeled by an $M_r \times M_t$ matrix $\mathbf{H}_n$ with $M_r \times 1$ channel noise $\mathbf{q}_n$. The input vector $\mathbf{s}_n$ and output vector $\mathbf{r}_n$ of the channel are related by $\mathbf{r}_n = \mathbf{H}_n\mathbf{s}_n + \mathbf{q}_n$. A simple but useful model to describe the time variation of a time-correlated MIMO channel is the first-order Gauss-Markov process [23]

$$\mathbf{H}_n = \sqrt{1-\epsilon^2}\mathbf{H}_{n-1} + \epsilon\mathbf{W}_n, \tag{1}$$

where $\mathbf{W}_n$ is independent of $\mathbf{H}_{n-1}$ and its entries are i.i.d. complex Gaussian random variables with zero mean and unit variance. The parameter ϵ is a coefficient that reflects the Doppler effect of the channel. With Jakes' model, $\epsilon = \sqrt{1-(J_0(2\pi f_d T))^2}$, where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind, f_d is the maximum Doppler frequency, T denotes the time interval between consecutive channel uses. The Doppler frequency can be computed using $f_d = vf_c/c$, where v is the mobile speed, f_c is the carrier frequency and $c = 3 * 10^8 m/s$ is the speed of light. When Doppler effect is not significant, the channel varies slowly, $J_0(2\pi f_d T) \approx 1$ and ϵ is a small number. The first-order Gauss-Markov model in (1) does not fully capture a practical time-correlated MIMO but it allows us to obtain a tractable solution, which will be tested on a more realistic time-correlated channel in simulations later. We assume the channel is slow fading so that the channel does not change during each channel use and it is known to the receiver. We also assume a delay free feedback channel with limited transmission rate is available. The noise vector $\mathbf{q}_n$ is additive white Gaussian with zero mean and variance N_0. We assume $M_r \geq M_t$ and the number of substreams transmitted M (i.e., the number of subchannels loaded with bits) is equal to M_t. (We consider the case when $M_r < M_t$ and $M = M_r$ in Section 4.2, then extend it to the more general case when there is no constraint on M_t, M_r and M, except that $M \leq \min\{M_t, M_r\}$.)

The inputs of the channel as indicated in Fig. 1 are modulation symbols $s_{n,k}$, for $k = 0, 1, \ldots, M_t - 1$, assumed to be uncorrelated with zero mean. For the subchannels loaded with bits, the variance of $s_{n,k}$ is equal to P_t/M, where P_t is the total transmission power. The $M_t \times M_r$ linear receiver $\mathbf{G}_n$

is zero forcing and $\mathbf{G}_n = (\mathbf{H}_n^\dagger \mathbf{H}_n)^{-1}\mathbf{H}_n^\dagger$ [24]. The receiver output error $\mathbf{e}_n = \mathbf{G}_n\mathbf{q}_n$ is Gaussian distributed when the channel $\mathbf{H}_n$ is given. The error $\mathbf{e}_n$ has autocorrelation matrix $\mathbf{R}_n = E[\mathbf{e}_n\mathbf{e}_n^\dagger]$ given by [24]

$$\mathbf{R}_n = N_0(\mathbf{H}_n^\dagger \mathbf{H}_n)^{-1}. \tag{2}$$

The kth subchannel error variance at time n is $\sigma^2_{e_{n,k}} = [\mathbf{R}_n]_{kk}$. When the error $\mathbf{e}$ is Gaussian and QAM modulation symbols are used, the number of bits that can be loaded on the kth subchannel at time n is [24]

$$b_{n,k} = \log_2\left(1 + \frac{P_t/(M\Gamma)}{\sigma^2_{e_{n,k}}}\right), \tag{3}$$

where Γ is the SNR gap that depends on the desired symbol error rate (SER)[2], $\Gamma = \frac{1}{3}[Q^{-1}(\frac{\mathrm{SER}}{4})]^2$, and $Q(y) = \frac{1}{\sqrt{2\pi}}\int_y^\infty e^{-t^2/2}dt$, $y \geq 0$. The number of bits transmitted at time n is $\mathcal{R}_n = \sum_{k=0}^{M-1} b_{n,k}$.

III. Prediction of Bit Loading

When the channel is time-correlated, we can expect the bit loading to be time-correlated as well. In this section we show how to predict the bit loading at time n given the channel at time $n-1$. Based on the results we derive an approximation of the predictor using the bit loading at time $n-1$.

Using the Gauss-Markov channel model in (1), the output noise autocorrelation matrix $\mathbf{R}_n = N_0(\mathbf{H}_n^\dagger\mathbf{H}_n)^{-1}$ in (2) can be written as

$$\mathbf{R}_n = N_0\left((1-\epsilon^2)\mathbf{H}_{n-1}^\dagger\mathbf{H}_{n-1} + \epsilon\sqrt{1-\epsilon^2}\left(\mathbf{H}_{n-1}^\dagger\mathbf{W}_n + \mathbf{W}_n^\dagger\mathbf{H}_{n-1}\right) + \epsilon^2\mathbf{W}_n^\dagger\mathbf{W}_n\right)^{-1} \tag{4}$$

The following lemma gives an approximation of $\mathbf{R}_n$ that allows us to make a useful connection between $\mathbf{R}_n$ and $\mathbf{R}_{n-1}$.

Lemma 1: Consider the output noise autocorrelation matrix $\mathbf{R}_n$ in (4). For a small ϵ, $\mathbf{R}_n$ has the following second-order approximation:

$$\begin{aligned}
&\mathbf{R}_n \approx \mathbf{R}_{n-1} + \epsilon\mathbf{A}_n + \epsilon^2\mathbf{B}_n,\\
\text{where}\quad &\mathbf{A}_n = -\frac{1}{N_0}\mathbf{R}_{n-1}\left(\mathbf{H}_{n-1}^\dagger\mathbf{W}_n + \mathbf{W}_n^\dagger\mathbf{H}_{n-1}\right)\mathbf{R}_{n-1},\\
\text{and}\quad &\mathbf{B}_n = \mathbf{R}_{n-1} - \frac{1}{N_0}\mathbf{R}_{n-1}\mathbf{W}_n^\dagger\mathbf{W}_n\mathbf{R}_{n-1}\\
&\qquad + \frac{1}{N_0^2}\mathbf{R}_{n-1}\left(\mathbf{H}_{n-1}^\dagger\mathbf{W}_n + \mathbf{W}_n^\dagger\mathbf{H}_{n-1}\right)\mathbf{R}_{n-1}\\
&\qquad \times\left(\mathbf{H}_{n-1}^\dagger\mathbf{W}_n + \mathbf{W}_n^\dagger\mathbf{H}_{n-1}\right)\mathbf{R}_{n-1}.
\end{aligned} \tag{5}$$

A proof is given in Appendix A. Note that the two matrices $\mathbf{A}_n$ and $\mathbf{B}_n$ depend on the channel $\mathbf{H}_{n-1}$ and $\mathbf{W}_n$ but not ϵ. The differential error covariance matrix $\mathbf{R}_n - \mathbf{R}_{n-1}$ can be approximated as $\epsilon\mathbf{A}_n + \epsilon^2\mathbf{B}_n$, which is equal to zero when $\epsilon = 0$. Using the expression of $\mathbf{R}_n$ in Lemma 1, we can derive the following approximation of bit loading at time n.

[2]When a desired bit error rate (BER) is given, we can use $\Gamma = -\ln(5\mathrm{BER})/1.5$ in (3) to approximate $b_{n,k}$ [7]. For a coded performance, adaptive coded modulation can be applied to each subchannel as in a single-input-single-output fading channel [25]. In this case, $\Gamma = -\ln(5\mathrm{BER})/(1.5G)$, where G is the channel coding gain [25].

Lemma 2: At time n, the number of bits that can be loaded on the kth subchannel can be approximated in terms of the bit loading at time $n-1$ as

$$b_{n,k} \approx b_{n-1,k} - \frac{\epsilon}{\ln 2}\frac{[\mathbf{A}_n]_{kk}}{\sigma^2_{e_{n-1,k}}} + \frac{\epsilon^2}{\ln 2}\left(\frac{[\mathbf{A}_n]^2_{kk}}{2\sigma^4_{e_{n-1,k}}} - \frac{[\mathbf{B}_n]_{kk}}{\sigma^2_{e_{n-1,k}}}\right), \tag{6}$$

when ϵ is small, where $[\mathbf{X}]_{i,j}$ denotes the (i,j)th entry of a matrix $\mathbf{X}$.

A proof is given in Appendix B. For the predictive quantization of $b_{n,k}$, we would like to have a predictor $\tilde{b}_{n,k}$ of $b_{n,k}$ so that the prediction error $\delta_{n,k} = b_{n,k} - \tilde{b}_{n,k}$ is small. It is known that, given the previous channel $\mathbf{H}_{n-1}$, the best predictor that minimizes the mean squared prediction error is the conditional mean [26] $\tilde{b}^{\mathrm{opt}}_{n,k} = E_{\mathbf{W}_n}[b_{n,k}|\mathbf{H}_{n-1}]$, where $E_{\mathbf{W}_n}[x]$ denotes the expectation of a random variable x averaged over the random matrix $\mathbf{W}_n$. The result in Lemma 2 leads to the following expression of predictor for $b_{n,k}$,

$$\tilde{b}^{\mathrm{opt}}_{n,k} \approx b_{n-1,k} + \frac{\epsilon^2}{\ln 2}E_{\mathbf{W}_n}\left[\frac{[\mathbf{A}_n]^2_{kk}}{2\sigma^4_{e_{n-1,k}}} - \frac{[\mathbf{B}_n]_{kk}}{\sigma^2_{e_{n-1,k}}}\middle|\mathbf{H}_{n-1}\right],$$

where we have used the fact that the entries of $\mathbf{W}_n$ have zero mean, and thus the conditional mean $E_{\mathbf{W}_n}[\mathbf{A}_n|\mathbf{H}_{n-1}] = \mathbf{0}$. It turns out that the above expectation term can be expressed in a closed form as described in Lemma 3.

Lemma 3: Consider the matrices $\mathbf{A}_n$ and $\mathbf{B}_n$ in (5). When the entries of $\mathbf{W}_n$ are independent Gaussian random variables with zero mean and unit variance, $E_{\mathbf{W}_n}[[\mathbf{A}_n]^2_{kk}|\mathbf{H}_{n-1}]$ and $E_{\mathbf{W}_n}[[\mathbf{B}_n]_{kk}|\mathbf{H}_{n-1}]$ are given, respectively, by

$$\begin{aligned}
E_{\mathbf{W}_n}[[\mathbf{A}_n]^2_{kk}|\mathbf{H}_{n-1}] &= \frac{2}{N_0}\sigma^2_{e_{n-1,k}}[\mathbf{R}^2_{n-1}]_{kk},\\
E_{\mathbf{W}_n}[[\mathbf{B}_n]_{kk}|\mathbf{H}_{n-1}] &= \sigma^2_{e_{n-1,k}}(1 + \mathrm{tr}(\mathbf{R}_{n-1})/N_0)\\
&\quad + \frac{(M-M_r)}{N_0}[\mathbf{R}^2_{n-1}]_{kk},
\end{aligned} \tag{7}$$

where $\mathrm{tr}(\mathbf{R}_{n-1})$ is the trace of $\mathbf{R}_{n-1}$. In this case, the predictor that minimizes the mean squared prediction error $E[|b_{n,k} - \tilde{b}_{n,k}|^2]$ is given by

$$\tilde{b}^{\mathrm{opt}}_{n,k} \approx b_{n-1,k} + \epsilon^2\eta_{n,k}, \tag{8}$$

$$\text{where } \eta_{n,k} = \frac{1}{N_0\ln 2}\times\left(\sigma^{-2}_{e_{n-1,k}}\left[\mathbf{R}^2_{n-1}\right]_{kk}(M_r + 1 - M) - \mathrm{tr}(\mathbf{R}_{n-1}) - N_0\right).$$

A proof is given in Appendix C. With the above results, the prediction error is given by

$$\delta_{n,k} \approx -\frac{\epsilon}{\ln 2}\frac{[\mathbf{A}_n]_{kk}}{\sigma^2_{e_{n-1,k}}} + \frac{\epsilon^2}{\ln 2}\left(\frac{[\mathbf{A}_n]^2_{kk}}{2\sigma^4_{e_{n-1,k}}} - \frac{[\mathbf{B}_n]_{kk}}{\sigma^2_{e_{n-1,k}}}\right) - \epsilon^2\eta_{n,k}. \tag{9}$$

We see the first term in (9) is in the order of ϵ as both $\mathbf{A}_n$ and $\sigma^2_{e_{n-1,k}}$ are independent of ϵ, while the rest of the expression is in the order of ϵ^2. The error $\delta_{n,k}$ is dominated by the first term. This means that we do not need to redesign the bit loading

codebook as ϵ varies. We can design the codebook for a particular ϵ. When ϵ changes to a different value, known to both the transmitter and receiver, we only need to scale the codewords in the codebook accordingly. Therefore, the codebook can be easily adapted as ϵ changes. This is similar to the result obtained in [28] for quantizing beamforming vectors in a time-correlated environment. As the optimal predictor is the conditional mean of $b_{n,k}$, the conditional mean of the prediction error $\delta_{n,k}$ in this case is equal to zero, i.e.,$E_{\mathbf{W}_n}[\delta_{n,k}|\mathbf{H}_{n-1}] = 0$. Given the previous channel, the variance of $\delta_{n,k}$ is the mean squared prediction error $\mathcal{E}_{n,k} = E_{\mathbf{W}_n}[\delta_{n,k}^2|\mathbf{H}_{n-1}]$. Based on (9), $\mathcal{E}_{n,k}$ has the second-order approximation $\mathcal{E}_{n,k} \approx (\epsilon/\ln 2)^2 E_{\mathbf{W}_n}[[\mathbf{A}_n]_{kk}^2|\mathbf{H}_{n-1}]\sigma_{e_{n-1,k}}^{-4}$, which can be expressed as

$$\mathcal{E}_{n,k} \approx \frac{2\epsilon^2}{(\ln 2)^2 N_0}\sigma_{e_{n-1,k}}^{-2}[\mathbf{R}_{n-1}^2]_{kk}, \tag{10}$$

using (7). We see that the variance of $\delta_{n,k}$ is in the order of ϵ^2.

Prediction using previous bit loading

The predictor in (8) is derived assuming $\mathbf{H}_{n-1}$ is known. It contains the term $[\mathbf{R}_{n-1}^2]_{kk}$ that is not available to the transmitter as only bit loading is fed back. Nonetheless we can use (8) to obtain an estimate for $\tilde{b}_{n,k}^{\text{opt}}$. To do this, note that $[\mathbf{R}_{n-1}^2]_{kk}$ is given by $[\mathbf{R}_{n-1}^2]_{kk} = \sum_{\ell=0}^{M-1} |E[e_{n-1,k}e_{n-1,\ell}^*]|^2$, where $e_{n-1,\ell}$ is the ℓth output subchannel error at time $n-1$. It can be bounded as

$$\sigma_{e_{n-1,k}}^4 \le [\mathbf{R}_{n-1}^2]_{kk} \le \sigma_{e_{n-1,k}}^2 \sum_{\ell=0}^{M-1} \sigma_{e_{n-1,\ell}}^2. \tag{11}$$

The lower bound is achieved when the subchannel errors are uncorrelated. The upper bound is obtained by using $|E[e_{n-1,k}e_{n-1,\ell}^*]| \le \sigma_{e_{n-1,k}}\sigma_{e_{n-1,\ell}}$, where equality holds if $e_{n-1,k}$ is a scaled version of $e_{n-1,\ell}$. The lower and upper bounds in (11) depend only on $\sigma_{e_{n-1,\ell}}^2$ for $\ell = 0, 1, \ldots, M-1$. The subchannel error variances $\sigma_{e_{n-1,\ell}}^2$ can be obtained from the bit loading information that is fed back of the transmitter using (3), i.e., $\sigma_{e_{n-1,\ell}}^2 = P_t/(M\Gamma(2^{b_{n-1,\ell}} - 1))$. Equation (11) leads to the following upper and lower bounds.

$$f_{n,k} \lesssim \tilde{b}_{n,k}^{\text{opt}} \lesssim g_{n,k}, \tag{12}$$

where

$$f_{n,k} = b_{n-1,k} + \frac{\epsilon^2}{\ln 2}\left((M_r + 1 - M)\frac{P_t/(M\Gamma N_0)}{2^{b_{n-1,k}} - 1} - \sum_{\ell=0}^{M-1}\frac{P_t/(M\Gamma N_0)}{2^{b_{n-1,\ell}} - 1} - 1\right),$$

$$g_{n,k} = b_{n-1,k} + \frac{\epsilon^2}{\ln 2}\left((M_r - M)\sum_{\ell=0}^{M-1}\frac{P_t/(M\Gamma N_0)}{2^{b_{n-1,\ell}} - 1} - 1\right).$$

The above upper and lower bounds depend only on the bit loading of the previous time instant and we can use them to approximate the predictor in (8).

To examine the difference between the lower bound predictor and the upper bound predictor, we conduct the following empirical comparison. The mean squared prediction error $E[\delta_{n,0}^2]$ is plotted as a function of ϵ assuming $\hat{b}_{n-1,0} = b_{n,0}$ in Fig. 2 for $M_r = M_t = 4$. We have obtained the predictor using the lower bound, the upper bound, and the mean of the two. We can see in Fig. 2 that that the predictor based on the lower bound has the smallest prediction error, the one based on the upper bound has the largest prediction, while the one based on the mean of the two bounds is in the middle. However the difference is a small one. We will use the lower bound for the prediction of bit loading in the next section.

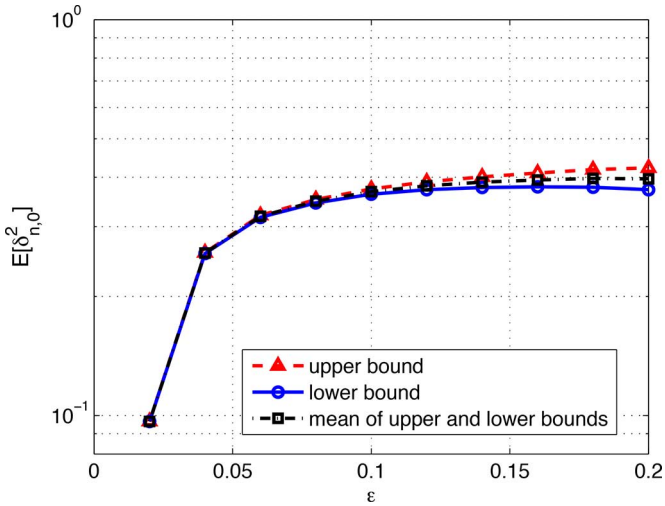


Fig. 2. Mean squared prediction errors using different predictors.

On the other hand, it is also desirable to have the information of the prediction error variance $\mathcal{E}_{n,k}$ available to the transmitter. The prediction error variance $\mathcal{E}_{n,k}$, if known to the transmitter, is a useful reference for designing the quantizer so that a smaller quantization error can be achieved [26]. The expression of $\mathcal{E}_{n,k}$ in (10) contains the term $[\mathbf{R}_{n-1}^2]_{kk}$ that is not known to the transmitter. To have an estimate of $\mathcal{E}_{n,k}$ based on the previous bit loading, we can use (10) and the bounds in (11) as in the above discussion to obtain

$$\frac{2\epsilon^2}{(\ln 2)^2}\frac{P_t/(M\Gamma N_0)}{2^{b_{n-1,k}} - 1} \lesssim \mathcal{E}_{n,k} \lesssim \frac{2\epsilon^2}{(\ln 2)^2}\sum_{\ell=0}^{M-1}\frac{P_t/(M\Gamma N_0)}{2^{b_{n-1,\ell}} - 1}, \tag{13}$$

where we have replaced $\sigma_{e_{n-1,\ell}}^2$ by $P_t/(M\Gamma(2^{b_{n-1,\ell}} - 1))$ as before. The above lower and upper bounds depend only on the previous bit loading and can be used to design the quantizer for quantizing the prediction error in Section 4.

IV. Predictive Quantization of Bit Loading

For the predictive quantization of bit loading $b_{n,k}$ at time n, the receiver computes a predicted value $\tilde{b}_{n,k}$ and apply quantization on the prediction error $\delta_{n,k} = b_{n,k} - \tilde{b}_{n,k}$. The quantized prediction error $\hat{\delta}_{n,k}$ is then sent back to the transmitter. The transmitter computes the predicted value $\tilde{b}_{n,k}$ just like the receiver and reproduces the quantized bit loading using

$$\hat{b}_{n,k} = \tilde{b}_{n,k} + \hat{\delta}_{n,k}.$$

The bounds derived in (12) can be used as a predictor for $b_{n,k}$. Using the lower bound of (12), we have the following predictor,

$$\tilde{b}_{n,k} = \hat{b}_{n-1,k} + \frac{\epsilon^2}{\ln 2}\left((M_r+1-M)\frac{P_t/(M\Gamma N_0)}{2^{\hat{b}_{n-1,k}}-1} - \sum_{\ell=0}^{M-1}\frac{P_t/(M\Gamma N_0)}{2^{\hat{b}_{n-1,\ell}}-1}\right), \quad (14)$$

where we have replaced $b_{n-1,k}$ by the quantized bit loading $\hat{b}_{n-1,k}$ as only the quantized bit loading is available to the transmitter. The above predicted value $\tilde{b}_{n,k}$ is computed at both the transmitter and the receiver. The kth subchannel at time n is loaded with $\lfloor \hat{b}_{n,k} \rfloor$ bits, where $\lfloor x \rfloor$ denotes the largest integer smaller or equal to x. Note that the quantization error $\hat{b}_{n,k} - b_{n,k}$ is the same as the quantization error of prediction error, i.e., $\hat{b}_{n,k} - b_{n,k} = \hat{\delta}_{n,k} - \delta_{n,k}$. With a proper design of the predictor, the prediction error $\delta_{n,k}$ has a smaller variance than $b_{n,k}$. Thus a smaller quantization error can be achieved for the same quantization bit rate as the quantization error variance is proportional to the variance of the random variable to be quantized [26].

Quantization of Prediction Error

For the quantization of $\delta_{n,k}$, a smaller quantization error can be achieved if the quantizer is designed according to the statistics of $\delta_{n,k}$ [30]. Observe that the elements of $\mathbf{A}_n$ are Gaussian random variables when $\mathbf{H}_{n-1}$ is given. Therefore in the expression of prediction error in (9), the first-order term of ϵ, i.e., the dominating term, is Gaussian. Thus $\delta_{n,k}$ is approximately Gaussian and it can be quantized using quantizers designed for Gaussian random variables. For example, when we use one bit to quantize a Gaussian random variable, the optimal reproduction points are $\pm\sqrt{\frac{2}{\pi}}$ times its standard deviation [30]. That is $\hat{\delta}_{n,k} = \sqrt{\frac{2\mathcal{E}_{n,k}}{\pi}}$, for $\delta_{n,k} \geq 0$ and $\hat{\delta}_{n,k} = -\sqrt{\frac{2\mathcal{E}_{n,k}}{\pi}}$ for $\delta_{n,k} < 0$. For higher resolution quantization of Gaussian random variables, the optimal reconstruction points can also be expressed in terms of the variance [30]. Thus we can easily adapt the reconstruction points according to the variance of the prediction error $\delta_{n,k}$. With such an adaptive quantization scheme the transmitter needs to know the prediction error variance for reconstructing $\hat{\delta}_{n,k}$. For this, we can use the bounds in (13) to approximate the prediction error variance $\mathcal{E}_{n,k}$. An approximation of $\mathcal{E}_{n,k}$ using the quantized bit loading and the upper bound in (13) is

$$\tilde{\mathcal{E}}_{n,k} = \frac{2\epsilon^2}{(\ln 2)^2}\sum_{\ell=0}^{M-1}\frac{P_t/(M\Gamma N_0)}{2^{\hat{b}_{n-1,\ell}}-1}. \quad (15)$$

Such an expression depends only on the previous quantized bit loading. When the receiver uses the approximated $\tilde{\mathcal{E}}_{n,k}$ to quantize the prediction errors in an adaptive manner, the transmitter can compute $\tilde{\mathcal{E}}_{n,k}$ using the quantized bit loading and reconstruct $\hat{\delta}_{n,k}$ just like the receiver. Using the upper bound in (13) as an estimate of the prediction error variance, the prediction error is somewhat overestimated and a larger quantization step size is used. In this case we can keep better track of the unquantized bit loading when there is a sudden change in the channel. This leads to a better error rate performance as we will see in simulations.

A. Low Feedback Rate

For a given feedback rate B, the prediction errors are each quantized using B/M bits. When the feedback rate B is small, e.g., $B = 1$, each prediction error can only be quantized using less than one bit. To quantize the prediction errors with a higher resolution, we can interleave the feedback of the prediction errors. For example, if we interleave the prediction errors and feedback each prediction error every M channel uses, each subchannel can be quantized using one bit when $B = 1$. More generally, suppose each prediction error is quantized using B_δ bits. Thus MB_δ bits are needed for quantizing all the prediction errors and the number of times of feedback required is $\tau = \lceil MB_\delta/B \rceil$, where $\lceil MB_\delta/B \rceil$ denotes the smallest integer larger or equal to x. Thus we need to predict the bit loading at time n using the bit loading at time $n-\tau$, instead of using the bit loading at time $n-1$ as discussed in Section 3. The derivation of predictor in this case is given below.

Let us examine the channel model in (1). Using $\mathbf{H}_{n-1} = \sqrt{1-\epsilon^2}\mathbf{H}_{n-2} + \epsilon\mathbf{W}_{n-1}$, we can write $\mathbf{H}_n$ as $\mathbf{H}_n = (1-\epsilon^2)\mathbf{H}_{n-2} + \epsilon\sqrt{1-\epsilon^2}\mathbf{W}_{n-1} + \epsilon\mathbf{W}_n$. Continue in the same manner, we have

$$\mathbf{H}_n = (1-\epsilon^2)^{\ell/2}\mathbf{H}_{n-\ell} + \epsilon(1-\epsilon^2)^{(\ell-1)/2}\mathbf{W}_{n-\ell+1} + \epsilon(1-\epsilon^2)^{(\ell-2)/2}\mathbf{W}_{n-\ell+2} + \cdots + \epsilon\mathbf{W}_n. \quad (16)$$

The first term $(1-\epsilon^2)^{\ell/2}\mathbf{H}_{n-\ell}$ is a matrix whose elements are Gaussian with variance $1-\ell\epsilon^2+\mathcal{O}(\epsilon^4)$, where $\mathcal{O}(\epsilon^4)$ denotes a term that is in the order of ϵ^4. The last ℓ terms of (16) add to a random matrix whose elements are independent and identically distributed Gaussian random variables with variance $\ell\epsilon^2 + \mathcal{O}(\epsilon^4)$. Thus, we can write $\mathbf{H}_n$ as $\mathbf{H}_n = \sqrt{1-\ell\epsilon^2+\mathcal{O}(\epsilon^4)}\mathbf{H}_{n-\ell} + \sqrt{\epsilon^2\ell+\mathcal{O}(\epsilon^4)}\mathbf{Z}_n$, where $\mathbf{Z}_n$ is an $M_r \times M$ matrix whose elements are independent Gaussian random variables with zero mean and unit variance. When ϵ is small, we have

$$\mathbf{H}_n \approx \sqrt{1-\ell\epsilon^2}\mathbf{H}_{n-\ell} + \sqrt{\ell}\epsilon\mathbf{Z}_n. \quad (17)$$

We see that (17) is similar to the Gauss-Markov channel model in (1) but the coefficient ϵ is replaced by $\sqrt{\ell}\epsilon$. This means that when we use the information at time $n-\tau$ to predict $b_{n,k}$, we can simply replace ϵ with $\epsilon' = \sqrt{\tau}\epsilon$ in (14) to obtain $\tilde{b}_{n,k}$. In this case the variance of the prediction error $\delta'_{n,k} = b_{n,k} - \tilde{b}'_{n,k}$ is roughly τ times that when there is no delay in prediction ($\tau = 1$). This is because the prediction error variance in (10) is proportional to ϵ^2. When the prediction error variance is employed for adaptive quantization of the prediction error, we can obtain an approximation of the variance by scaling the expression in (15) by τ.

B. Removing the Constraint $M_t \leq M_r$ and $M = M_t$

First let us consider $M_r < M_t$ and the number of substreams transmitted $M = M_r$. Then we can extend it to the more general case when there is no constraint on M_t, M_r and M, except that $M \leq \min\{M_t, M_r\}$. When $M_r < M_t$ and $M = M_r$, some

$M_t - M$ subchannels are not loaded with bits. The bit loading vector $\mathbf{b}_n = [\, b_{n,0} \quad b_{n,1} \quad \cdots \quad b_{n,M_t-1} \,]^T$ has $M_t - M$ zero entries. Removing the zero entries, we get an $M \times 1$ reduced bit loading vector $\mathbf{b}'_n$. For the predictive quantization of bit loading, we only need to consider the quantization of the reduced bit loading $\mathbf{b}'_n$. Let $\mathbf{H}'_n$ be the $M_r \times M$ submatrix of $\mathbf{H}_n$ by keeping the columns that correspond to the nonzero entries of $\mathbf{b}_n$. In a similar manner, we define $\mathbf{W}'_n$. From (1), the effective channel $\mathbf{H}'_n$ can be written as $\mathbf{H}'_n = \sqrt{1-\epsilon^2}\mathbf{H}'_{n-1} + \epsilon\mathbf{W}'_n$ and the subchannel error autocorrelation matrix is $\mathbf{R}'_n = N_0(\mathbf{H}'^{\dagger}_n\mathbf{H}'_n)^{-1}$. Replacing $\mathbf{H}_n$ and $\mathbf{W}_n$, respectively, by $\mathbf{H}'_n$ and $\mathbf{W}'_n$ everywhere in earlier derivations, we can predict bit loading and quantize prediction error as before. As M is less than M_t, we can choose M out of M_t subchannels for transmission, according to the channel state information. The transmitter can be informed of the change of loaded subchannels periodically, say, every S channel uses. To avoid extra feedback, the first few feedback bits of each S-period can be used to inform the transmitter. More generally we can also change M according to the channel condition. This can be done by choosing the optimal subset of M subchannels that maximizes the transmission rate at the beginning of each S-period, where $1 \le M \le \min\{M_r, M_t\}$. In this case, there is no constraint on M_t and M_r.

V. Simulation Examples

In the following examples $M_r = 4$, $M_t = 4$. We have used $f_c = 2.5 * 10^9$ Hz, and $T = 2$ ms as suggested in [32]. In a microcellular transmission scenario, the terminal speed of interest is 3 km/hr [32], which results in maximum Doppler frequency $f_D = 7$ Hz and $\epsilon = 0.06$. In Examples 1–2, $P_t/N_0 = 15$ dB and $\Gamma = 5.05$, which corresponds to a symbol error rate of 10^{-4}. We have used $5 * 10^6$ channel realizations in Examples 2–4. The channel $\mathbf{H}_n$ is generated using the Gauss-Markov model in (1) for Examples 1–3 and generated using the filtering method [33] for Example 4.

Example 1. Quantized Transmission Rate: Fig. 3(a) and (b) show the unquantized and quantized transmission rates of the proposed PQB system for, respectively, $\epsilon = 0.06$ and $\epsilon = 0.1$. The feedback rate $B = 1$. The prediction error $\delta_{n,k}$ is quantized using a one-bit codebook for Gaussian random variables. As the feedback rate $B = 1$, the prediction errors are interleaved for feedback as discussed in Section 4.1. The unquantized rate is the sum of unquantized bit assignments $\mathcal{R}_n = \sum_{k=0}^{M-1} b_{n,k}$, where $b_{n,k}$ is computed according to (3). For the quantized rate $\hat{\mathcal{R}}_n = \sum_{k=0}^{M-1} \hat{b}_{n,k}$, predictive quantization of bit loading is applied assuming the transmitter knows only the quantized bit loading $\hat{b}_{n-1,k}$; $b_{n,k}$ is predicted using (14) and $\mathcal{E}_{n,k}$ estimated using (15). The curve "$\hat{\mathcal{R}}_n$ ($\mathbf{H}_{n-1}$ known)" is the quantized transmission rate assuming $\mathbf{H}_{n-1}$ is known to the transmitter so that the bit loading is predicted with the knowledge of $\mathbf{H}_{n-1}$ in (8) and the quantizer is adapted according to (10). Although the predictor and prediction error variances are approximated for $\hat{\mathcal{R}}_n$, the curves of the two quantized rates are very close. We can also see that the quantized rate $\hat{\mathcal{R}}_n$ is a good approximation of $\mathcal{R}_n$ even for $B = 1$.

Example 2. Quantization Error: We show the mean squared quantization error (MSE) MSE $= E[\sum_{k=0}^{M_t-1}(\hat{b}_{n,k} - b_{n,k})^2]$ as a function of ϵ in Fig. 4(a) for $B = 1, 2, 6$ and as a function of the feedback rate B in Fig. 4(b) for $\epsilon = 0.02, 0.06, 0.1$. The number of substreams transmitted M is 3 and a two-bit Gaussian quantizer is used for quantizing the prediction errors. Fig. 4(a) demonstrates that the MSE is roughly proportional to ϵ^2 for the same B. For example, observe the curve of $B = 6$; when $\epsilon = 0.02$, MSE is around 0.0015 and when $\epsilon = 0.04$, MSE is around 0.006, roughly 4 times that of $\epsilon = 0.02$. From Fig. 4(b) we see that the MSE decreases with the feedback rate B. For example, when $\epsilon = 0.06$, the MSE of $B = 1, 2$ and 6 is, respectively, around 0.1, 0.05 and 0.015. That is, the MSE of $B = 1$ is approximately 2 times that of $B = 2$ and 6 times that of $B = 6$. This is consistent with what we have derived in Section 4.1. To see this, observe that for $B = 1$ and 2, interleaving of the prediction errors is applied because B is not large enough to feedback all the prediction errors in one feedback. As the prediction errors are quantized using a 2-bit codebook and $M = 3$, the feedback of all three prediction errors requires 6 bits, which needs, respectively, 6, 3 and 1 times of feedback for $B = 1, 2$ and 6. Thus the prediction error variance of $B = 1$ is

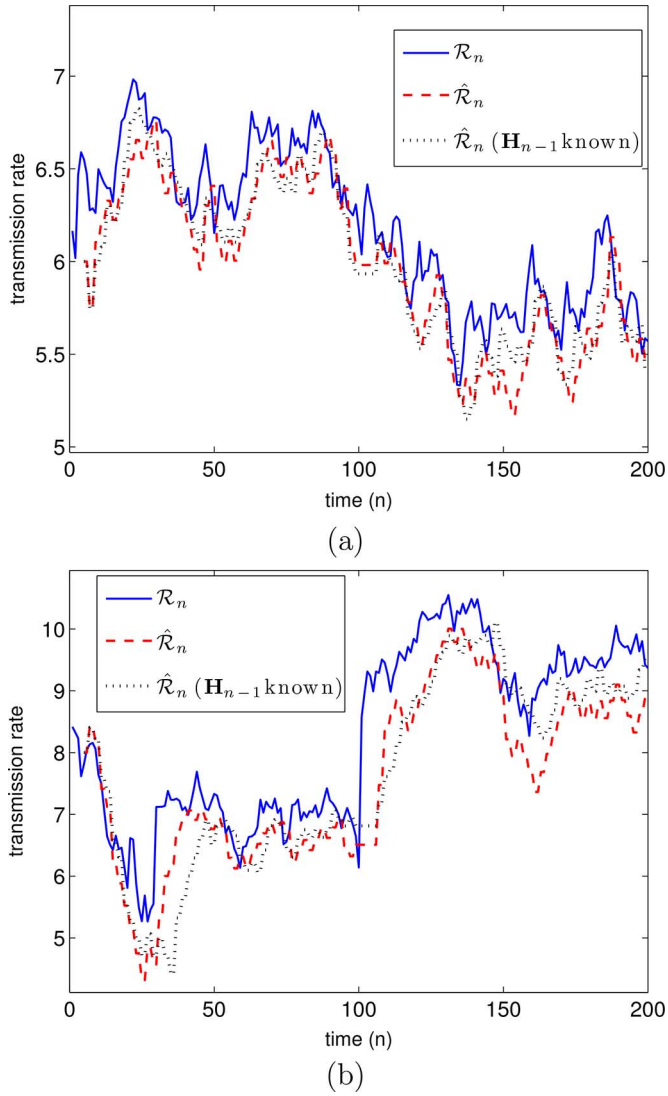


Fig. 3. The unquantized and quantized transmission rates for (a) $\epsilon = 0.06$ and (b) $\epsilon = 0.1$.

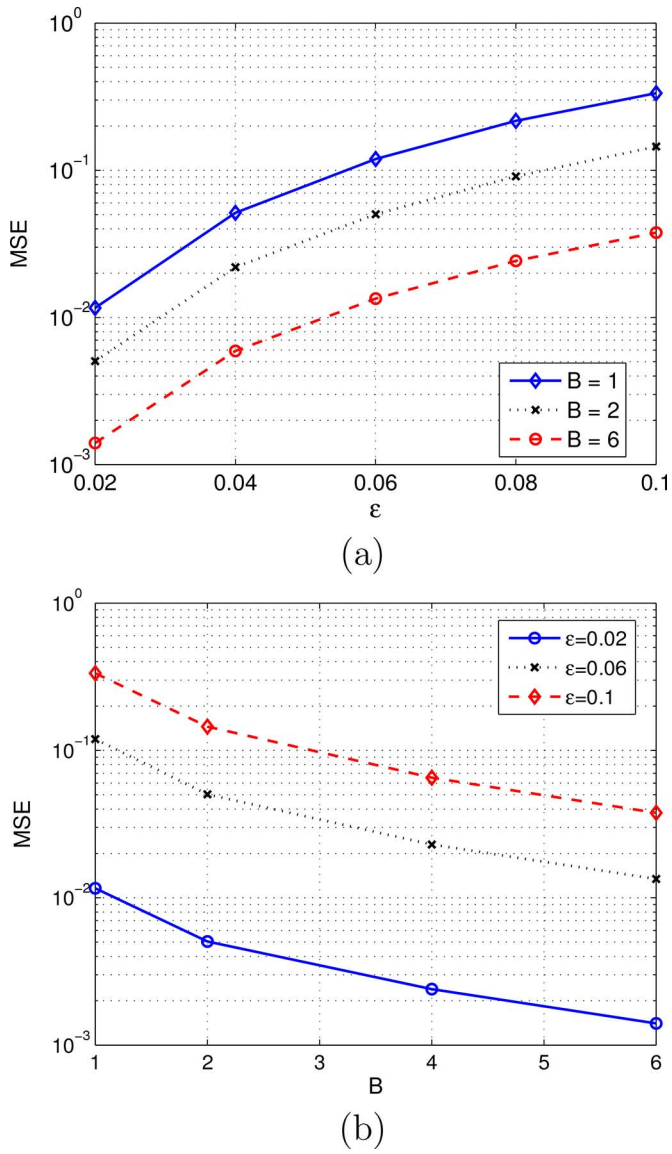


Fig. 4. Mean squared quantization error as a function of (a) ϵ and (b) feedback rate B.

around 6 times that of $B = 6$ and 2 times that of $B = 2$ based on the discussion in Section 4.1.

Example 3. BER Performance: Fig. 5 show for $\epsilon = 0.06$ the BER performance of the PQB system for a given transmission rate of 12 bits per channel use. As the transmission rate of the PQB system varies with the channel, we adjust the target error rate for each data point so that the average transmission rate is equal to or slightly over 12. The feedback rate $B = 2$ and the number of substreams transmitted is updated periodically without using extra feedback as described in Section 4.2, and a two-bit quantizer is used for quantizing the prediction errors. The quantizer is adapted according to the upper bound of prediction error variance in (13), labeled as 'upper bound', and according to the lower bound ('lower bound'). We have also shown PQB for the case the quantizer is not adapted ('fixed quantizer'), and for the case the bit loading is the integer part of the unquantized bit loading in (3), labeled as 'unquantized'.

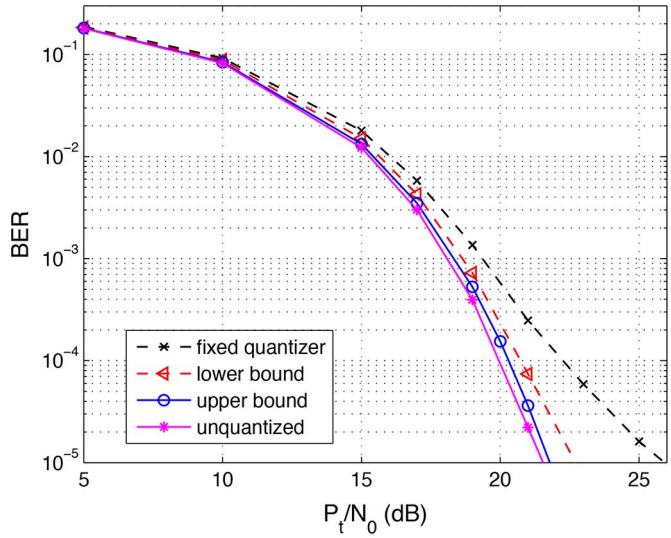


Fig. 5. BER performance of PQB using different quantizers.

For 'fixed quantizer', the quantizer is designed for a Gaussian random variable with a fixed standard deviation, which is 0.01 in the simulations. Using the upper bound in (13) for quantizer adaptation, a smaller BER can be obtained then the lower bound. This is because, with the upper bound, a larger quantization step size is used. This allows better tracking of the channel when the channel changes suddenly. With the lower bound, such a sudden change can lead to a larger quantization error and hence a larger error rate, which usually dominates the overall BER performance. For $B = 2$ the performance of PQB using the upper bound adaptation is close to that of the unquantized bit loading; the difference is 0.5 dB when BER is 10^{-4}.

Example 4. Performance Comparison: In this example, we use the filtering method [33] to generate the time-correlated MIMO channel. For the same maximum Doppler frequency $f_D = 7$ Hz and $\epsilon = 0.06$ used in Example 3, the BER performance is shown as a function of P_t/N_0 in Fig. 6(a). We use feedback rate $B = 2$ for all the curves shown and there is no extra feedback overhead. The transmission rate is 12 bits per channel in Fig. 6(a). For 'PQB', the quantizer is adapted on the fly using the prediction error variance approximation in (15). The curve "full bit loading" corresponds to the case that the best integer bit loading is computed and fed back slowly without quantization. Also shown in Fig. 6(a) is the BER of the differential precoder systems in [17] and [20], and the bit allocation (BA) system in [6] for the same transmission rate $R_b = 12$. All of these three are fixed-rate systems. Time correlation of the channel is exploited in [17] using quantization of the Gram matrix $\mathbf{H}_n^\dagger \mathbf{H}_n$ based on geodesic curve and also exploited in [20] using differential rotation of the previous precoder. The BA system feedbacks the bit loading vector from a codebook with a fixed transmission rate to minimize BER. We can see that PQB requires a smaller power for the same BER. As an example, for BER $= 10^{-4}$, it has a 1.7 dB gain over other systems. In Fig. 6(b) we show the BER performance as a function of transmission rate for the systems in (a) when $P_t/N_0 = 20$ dB and $B = 2$. To obtain the performance for PQB, we choose different target error rates and for each case

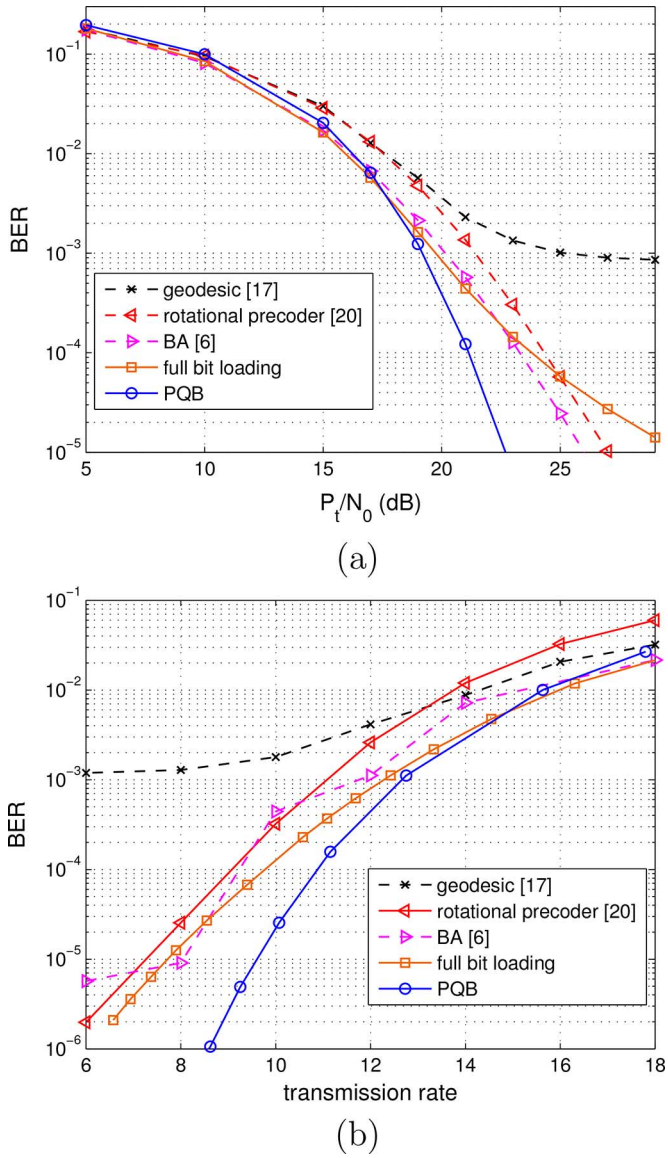


Fig. 6. Performance comparison for $\epsilon = 0.06$, (a) BER versus SNR, and (b) BER versus transmission rate.

compute the resulting average transmission rate and average BER. For the fixed rate systems, different transmission rates are chosen and the corresponding BER computed. With PQB, a smaller BER can be achieved for the same average transmission rate.

In the above comparisons, the "full bit loading" case is not as good as PQB. To understand this better, we plot the BER v.s. transmission curves for $\epsilon = 0.02$ in Fig. 7. For the almost time-invariant case $\epsilon = 0.02$, the 'full bit loading' case has very good performance. But as the channel varies faster ($\epsilon = 0.06$ in Fig. 6(b)), the performance deteriorates rapidly. This is expectable because the feedback of full bit loading requires more feedback time. When the channel changes faster, the feedback information becomes out of date by the time the information is used for transmission. The performance of the other schemes degrades in a more graceful manner as ϵ increases.

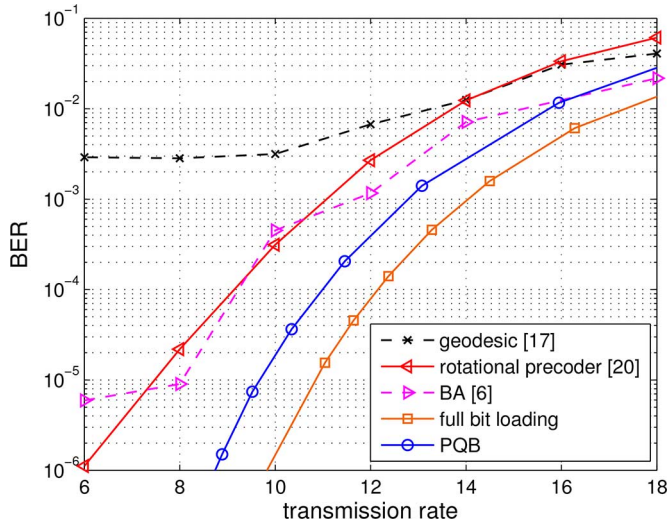


Fig. 7. BER versus transmission rate for $\epsilon = 0.02$.

VI. Conclusion

In this paper we considered variable-rate transmission for time-correlated MIMO channels with limited feedback using predictive quantization of bit loading. The transmission rate is adapted to the current channel condition by dynamically assigning bits to the subchannels of the MIMO system. We assumed the channel is described by a first-order Gauss-Markov model. The predictor is derived and the corresponding prediction errors shown to be approximately Gaussian when the channel is slow fading. For the quantization of prediction errors, we show the quantizer can be adapted according to the prediction error variance to achieve a smaller quantization error. Furthermore the predictor and the prediction error quantizers can be easily modified when there is a change in the time correlation coefficient. Simulation examples have been presented to demonstrate that a very small mean squared error can be obtained and a good performance achieved using a low feedback rate.

Appendix

Proof of Lemma 1: The expression of $\mathbf{R}_n$ in (4) can be rearranged as

$$\mathbf{R}_n = N_0 \left\{ \left[\mathbf{I}_M + \left(\epsilon\sqrt{1-\epsilon^2}(\mathbf{H}_{n-1}^\dagger \mathbf{W}_n + \mathbf{W}_n^\dagger \mathbf{H}_{n-1}) + \epsilon^2 \mathbf{W}_n^\dagger \mathbf{W}_n \right) \left((1-\epsilon^2)\mathbf{H}_{n-1}^\dagger \mathbf{H}_{n-1} \right)^{-1} \right] \times (1-\epsilon^2)\mathbf{H}_{n-1}^\dagger \mathbf{H}_{n-1} \right\}^{-1}. \quad (18)$$

by pulling out the term $(1-\epsilon^2)\mathbf{H}_{n-1}^\dagger \mathbf{H}_{n-1}$, where $\mathbf{I}_M$ denotes the $M \times M$ identity matrix. We can further rewrite it as $\mathbf{R}_n = \frac{1}{1-\epsilon^2}\mathbf{R}_{n-1}(\mathbf{I}_M + \mathbf{E})^{-1}$, where $\mathbf{E} = \left(\epsilon\sqrt{1-\epsilon^2}(\mathbf{H}_{n-1}^\dagger \mathbf{W}_n + \mathbf{W}_n^\dagger \mathbf{H}_{n-1}) + \epsilon^2 \mathbf{W}_n^\dagger \mathbf{W}_n \right) \times \left((1-\epsilon^2)\mathbf{H}_{n-1}^\dagger \mathbf{H}_{n-1} \right)^{-1}$ and we have used $\mathbf{R}_{n-1} =$

$N_0(\mathbf{H}_{n-1}^\dagger \mathbf{H}_{n-1})^{-1}$. It is known that (page 301, [29]) when $\mathbf{E}$ satisfies $\|\mathbf{E}\| < 1$, where $\|\mathbf{E}\|$ denotes a certain matrix norm of $\mathbf{E}$, e.g., Frobenius norm, then $(\mathbf{I}_M + \mathbf{E})^{-1}$ can be written as a power series $\sum_{k=0}^{\infty}(-1)^k \mathbf{E}^k$. We see that $\mathbf{E}$ has norm smaller than unity when ϵ is sufficiently small. Using the approximation $(\mathbf{I}_M + \mathbf{E})^{-1} \approx \mathbf{I}_M - \mathbf{E} + \mathbf{E}^2$, we obtain

$$\begin{aligned}\mathbf{R}_n \approx \frac{\mathbf{R}_{n-1}}{1-\epsilon^2}\Bigg\{\mathbf{I} - \Big[\epsilon\sqrt{1-\epsilon^2}\left(\mathbf{H}_{n-1}^\dagger \mathbf{W}_n + \mathbf{W}_n^\dagger \mathbf{H}_{n-1}\right) \\ +\epsilon^2 \mathbf{W}_n^\dagger \mathbf{W}_n\Big]\frac{1}{1-\epsilon^2}\left(\mathbf{H}_{n-1}^\dagger \mathbf{H}_{n-1}\right)^{-1}\Bigg\} \\ + \frac{\mathbf{R}_{n-1}}{1-\epsilon^2}\Bigg(\Big[\epsilon\sqrt{1-\epsilon^2}\left(\mathbf{H}_{n-1}^\dagger \mathbf{W}_n + \mathbf{W}_n^\dagger \mathbf{H}_{n-1}\right) \\ +\epsilon^2 \mathbf{W}_n^\dagger \mathbf{W}_n\Big]\frac{1}{1-\epsilon^2}(\mathbf{H}_{n-1}^\dagger \mathbf{H}_{n-1})^{-1}\Bigg)^2. \qquad (19)\end{aligned}$$

With the second order Taylor approximation, we have $(1-\epsilon^2)^{-1} \approx 1+\epsilon^2$ and $(1-\epsilon^2)^{1/2} \approx 1 - \epsilon^2/2$. Using these approximations and ignoring the third and higher-order terms of ϵ in (19), we have the second-order approximation of $\mathbf{R}_n$ in (5).

APPENDIX

Proof of Lemma 2: With (5) we can approximate the kth subchannel error variance at time n by $\sigma_{n,k}^2 \approx \sigma_{n-1,k}^2 + \epsilon[\mathbf{A}_n]_{kk} + \epsilon^2[\mathbf{B}_n]_{kk}$. Using this expression and (3) we obtain

$$b_{n,k} \approx \log_2\left(\frac{\sigma_{e_{n-1,k}}^2 + \epsilon[\mathbf{A}_n]_{kk} + \epsilon^2[\mathbf{B}_n]_{kk} + P_t/(M\Gamma)}{\sigma_{e_{n-1,k}}^2 + \epsilon[\mathbf{A}_n]_{kk} + \epsilon^2[\mathbf{B}_n]_{kk}}\right).$$

When ϵ is small, we can approximate the numerator $\sigma_{e_{n-1,k}}^2 + \epsilon[\mathbf{A}_n]_{kk} + \epsilon^2[\mathbf{B}_n]_{kk} + P_t/(M\Gamma)$ as $\sigma_{e_{n-1,k}}^2 + P_t/(M\Gamma)$. Then we have

$$\begin{aligned}b_{n,k} &\approx \log_2\left(\frac{\sigma_{e_{n-1,k}}^2 + P_t/(M\Gamma)}{\sigma_{e_{n-1,k}}^2 + \epsilon[\mathbf{A}_n]_{kk} + \epsilon^2[\mathbf{B}_n]_{kk}}\right) \\ &= -\log_2\left(\frac{\sigma_{e_{n-1,k}}^2}{\sigma_{e_{n-1,k}}^2 + P_t/(M\Gamma)}\right) \\ &\quad - \log_2\left(1 + \frac{\epsilon[\mathbf{A}_n]_{kk} + \epsilon^2[\mathbf{B}_n]_{kk}}{\sigma_{e_{n-1,k}}^2}\right).\end{aligned}$$

We recognize the first term $-\log_2\left(\frac{\sigma_{e_{n-1,k}}^2}{\sigma_{e_{n-1,k}}^2 + P_t/(M\Gamma)}\right)$ is equal to $b_{n-1,k}$, the number of bits loaded on the kth subchannel at time $n-1$ in (3). The second term of the above equation is of the form $-\log_2(1 + f\epsilon + g\epsilon^2)$, where $f = \sigma_{e_{n-1,k}}^{-2}[\mathbf{A}_n]_{kk}$ and $g = \sigma_{e_{n-1,k}}^{-2}[\mathbf{B}_n]_{kk}$. Using the Taylor series of $\log_2(1 + f\epsilon + g\epsilon^2)$ at $\epsilon = 0$, we have the second-order approximation $\frac{1}{\ln 2}(f\epsilon + (g - f^2/2)\epsilon^2)$, which leads to (6).

APPENDIX

Proof of Lemma 3: For brevity of notation, the (i,j)th element of $\mathbf{R}_{n-1}$, $\mathbf{H}_{n-1}$, $\mathbf{W}_n$ are denoted, respectively, as r_{ij}, h_{ij} and w_{ij} in this proof. Using (5), the kth diagonal element of $\mathbf{A}_n$ is given by

$$[\mathbf{A}_n]_{kk} = -\frac{1}{N_0}\sum_{i=0}^{M-1}\sum_{j=0}^{M_r-1}\sum_{\ell=0}^{M-1} r_{ki}h_{ji}^* w_{j\ell} r_{\ell k} + r_{ki} w_{ji}^* h_{j\ell} r_{\ell k}.$$

The conditional mean square of $[\mathbf{A}_n]_{kk}$ is

$$\begin{aligned}&E_{\mathbf{W}_n}\left[[\mathbf{A}_n]_{kk}^2 | \mathbf{H}_{n-1}\right] \\ &= \frac{1}{N_0^2} E_{\mathbf{W}_n}\Bigg[\left(\sum_{i=0}^{M-1}\sum_{j=0}^{M_r-1}\sum_{\ell=0}^{M-1} r_{ki}h_{ji}^* w_{j\ell} r_{\ell k} + r_{ki} w_{ji}^* h_{j\ell} r_{\ell k}\right) \\ &\times \left(\sum_{m=0}^{M-1}\sum_{p=0}^{M_r-1}\sum_{q=0}^{M-1} r_{km}h_{pm}^* w_{pq} r_{qk} + r_{km} w_{pm}^* h_{pq} r_{qk}\right)\Bigg| \mathbf{H}_{n-1}\Bigg]. \qquad (20)\end{aligned}$$

As the entries of $\mathbf{W}_n$ are independent circularly symmetric Gaussian random variables with zero mean and unit variance, $E_{\mathbf{W}_n}[w_{j\ell} w_{j\ell}^*] = 1$, $E_{\mathbf{W}_n}[w_{j\ell} w_{pm}^*] = 0$ when $p \neq j$ or $m \neq \ell$, and $E_{\mathbf{W}_n}[w_{j\ell} w_{pm}] = 0$ for all $0 \leq j, p < M_r$, and $0 \leq \ell, m < M$. Using these properties we can verify that (20) reduces to

$$\begin{aligned}&E_{\mathbf{W}_n}\left[[\mathbf{A}_n]_{kk}^2 | \mathbf{H}_{n-1}\right] \\ &\qquad = \frac{2}{N_0^2}\sum_{\ell=0}^{M-1}|r_{k\ell}|^2 \sum_{i=0}^{M-1}\sum_{j=0}^{M_r-1}\sum_{q=0}^{M-1} r_{ki}h_{ji}^* h_{jq} r_{qk}.\end{aligned}$$

We observe that $\sum_{i=0}^{M-1}\sum_{j=0}^{M_r-1}\sum_{q=0}^{M-1} r_{ki}h_{ji}^* h_{jq} r_{qk}$ is the kth diagonal element of the product $\mathbf{R}_{n-1}\mathbf{H}_{n-1}^\dagger \mathbf{H}_{n-1}\mathbf{R}_{n-1}$. The above equation can be further simplified as in (7) by using $\mathbf{R}_{n-1} = N_0(\mathbf{H}_{n-1}^\dagger \mathbf{H}_{n-1})^{-1}$ and $\sum_{\ell=0}^{M-1}|r_{k\ell}|^2 = [\mathbf{R}_{n-1}^2]_{kk}$.

With the expression of $\mathbf{B}$ in (5), we get

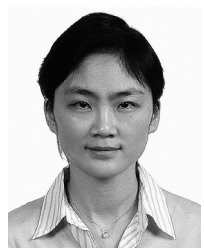
$$\begin{aligned}E[\mathbf{B}_n|\mathbf{H}_{n-1}] &= \mathbf{R}_{n-1} - \frac{1}{N_0}\mathbf{R}_{n-1}E_{\mathbf{W}_n}[\mathbf{W}_n^\dagger \mathbf{W}_n|\mathbf{H}_{n-1}]\mathbf{R}_{n-1} \\ &\quad + \frac{1}{N_0^2}\mathbf{R}_{n-1}E_{\mathbf{W}_n}[(\mathbf{W}_n^\dagger \mathbf{H}_{n-1} + \mathbf{H}_{n-1}^\dagger \mathbf{W}_n)\mathbf{R}_{n-1} \\ &\quad \times \left(\mathbf{W}_n^\dagger \mathbf{H}_{n-1} + \mathbf{H}_{n-1}^\dagger \mathbf{W}_n\right)|\mathbf{H}_{n-1}]\mathbf{R}_{n-1}.\end{aligned}$$

As the elements of $\mathbf{W}_n$ are independent and of unit variance, $E[\mathbf{W}_n^\dagger \mathbf{W}_n] = M_r \mathbf{I}_M$. Observe that $E[\mathbf{W}_n^\dagger \mathbf{Q}\mathbf{W}_n] = \mathrm{tr}(\mathbf{Q})\mathbf{I}_M$ for an $M_r \times M_r$ deterministic matrix $\mathbf{Q}$. This is because the (m,n)th element of $E[\mathbf{W}_n^\dagger \mathbf{Q}\mathbf{W}_n]$ is $E[\sum_{k=0}^{M_r-1}\sum_{j=0}^{M_r-1} w_{km}^*[\mathbf{Q}]_{kj} w_{jn}]$, which is non zero only when $m = n$ and $j = k$, and the mth diagonal element of $E[\mathbf{W}_n^\dagger \mathbf{Q}\mathbf{W}_n]$ is equal to $\sum_{k=0}^{M_r-1}[\mathbf{Q}]_{kk}$. In a similar manner, it can be verified that $E[\mathbf{W}_n \mathbf{F}\mathbf{W}_n^\dagger] = \mathrm{tr}(\mathbf{F})\mathbf{I}_{M_r}$ for an $M \times M$ deterministic matrix $\mathbf{F}$ and $E[\mathbf{W}_n \mathbf{P}\mathbf{W}_n] = 0$, for an $M \times M_r$ deterministic matrix $\mathbf{P}$. Using these identities, we can verify that $E[\mathbf{B}_n|\mathbf{H}_{n-1}]$ is as given in (7).

REFERENCES

[1] D. Love, R. W. Heath, Jr., V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[2] D. J. Love and R. W. Heath, Jr., "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.
[3] D. J. Love and R. W. Heath, Jr., "Multimode precoding for MIMO wireless systems," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3674–3687, Oct. 2005.
[4] S. Zhou and B. Li, "BER criterion and codebook construction for finite-rate precoded spatial multiplexing with linear receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1653–1665, May 2006.
[5] C. C. Weng, C. Y. Chen, and P. P. Vaidyanathan, "MIMO transceivers with decision feedback and bit loading: Theory and optimization," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1334–1346, Mar. 2010.
[6] Y.-P. Lin, H.-C. Chen, and P. Jeng, "Bit allocation and statistical precoding for correlated MIMO channels with limited feedback," *IEEE Trans. Veh. Technol.*, vol. 61, no. 2, pp. 597–606, Feb. 2012.
[7] A. J. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.
[8] T. Wu and V. K. N. Lau, "Robust rate, power and precoder adaptation for slow fading MIMO channels with noisy limited feedback," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2360–2367, Jun. 2008.
[9] M. R. McKay, I. B. Collings, A. Forenza, and R. W. Heath, Jr., "Multiplexing/beamforming switching for coded MIMO in spatially correlated channels based on closed-form BER approximations," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 2555–2567, Sep. 2007.
[10] J. Huang and S. Signell, "On spectral efficiency of low-complexity adaptive MIMO systems in Rayleigh fading channel," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4369–4374, Sep. 2009.
[11] P. Yang, Y. Xiao, L. Li, Q. Tang, Y. Yu, and S. Li, "Link adaptation for spatial modulation with limited feedback," *IEEE Trans. Veh. Technol.*, vol. 61, no. 8, pp. 3808–3813, Oct. 2012.
[12] R. Yellapantula, Y. Yao, and R. Ansari, "Unitary precoding and power control in MIMO systems with limited feedback," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2006, pp. 1221–1226.
[13] M. A. Sadrabadi, A. K. Khandani, and F. Lahouti, "Channel feedback quantization for high date rate MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3335–3338, Dec. 2006.
[14] S. T. Chung, A. Lozano, H. C. Huang, A. Sutivong, and J. M. Cioffi, "Approaching the MIMO capacity with a low-rate feedback channel in V-BLAST," *EURASIP J. Appl. Signal Process.*, no. 5, pp. 762–771, 2004.
[15] H. Zhuang, L. Dai, S. Zhou, and Y. Yao, "Low complexity per-antenna rate and power control approach for closed-loop V-BLAST," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1783–1787, Nov. 2003.
[16] K. Huang, R. W. Heath, Jr., and J. G. Andrews, "Limited feedback beamforming over temporally-correlated channels," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1959–1975, May 2009.
[17] D. Sacristán-Murga and A. Pascual-Iserte, "Differential feedback of MIMO channel Gram matrices based on Geodesic Curves," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3714–3727, Dec. 2010.
[18] L. Zhang, L. Song, M. Ma, and B. Jiao, "On the minimum differential feedback for time-correlated MIMO Rayleigh block-fading channels," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 411–420, Feb. 2012.
[19] J. Choi, B. Clerckx, N. Lee, and G. Kim, "A new design of polar-cap differential codebook for temporally/spatially correlated MISO channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 703–711, Feb. 2012.
[20] T. Kim, D. J. Love, and B. Clerckx, "MIMO systems with limited rate differential feedback in slowly varying channels," *IEEE Trans. Commun.*, vol. 59, no. 4, pp. 1175–1189, Apr. 2011.
[21] W. Santipach and K. Mamat, "Optimal feedback interval for temporally-correlated multiantenna channel," in *Proc. Global Telecommun. Conf.*, 2011, pp. 5–9.
[22] T. Kim, D. J. Love, and B. Clerckx, "Does frequent low resolution feedback outperform infrequent high resolution feedback for multiple antenna beamforming systems?," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1654–1669, Apr. 2011.
[23] R. Etkin and D. N. C. Tse, "Degrees of freedom in some underspread MIMO fading channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1576–1608, Apr. 2006.
[24] P. P. Vaidyanathan, S.-M. Phoong, and Y.-P. Lin, *Signal Processing and Optimization for Transceiver Systems*. Cambridge, U.K.: Cambridge Univ. Press, Apr. 2010.
[25] A. J. Goldsmith and S.-G. Chua, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 595–602, May 1998.
[26] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. New York, NY, USA: Kluwer Academic Publishers, 1991.
[27] T. Rappaport, *Wireless Communications-Principles and Practice*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice Hall, 2001.
[28] K. Kim, T. Kim, D. J. Love, and I. H. Kim, "Differential feedback in codebook-based multiuser MIMO systems in slowly varying channels," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 578–588, Feb. 2012.
[29] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[30] K. Sayood, *Introduction to Data Compression*. New York, NY, USA: Morgan Kaufmann, 2005.
[31] G. Gore, R. W. Heath, Jr., and A. Paulraj, "One performance of the zero forcing receive in presence of transmit correlation," in *Proc. Int. Symp. Inf. Theory*, 2002, p. 159.
[32] "ITU-R, Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced," Tech. Rep. ITU-R M.2135-1, Dec. 2009.
[33] G. L. Stüber, *Principles of Mobile Communication*. Berlin, Germany: Springer-Verlag, 2000.



Yuan-Pei Lin (S'93–M'97–SM'03) was born in Taipei, Taiwan, 1970. She received the B.S. degree in control engineering from the National Chiao-Tung University, Taiwan, in 1992, and the M.S. degree and the Ph.D. degree, both in electrical engineering from California Institute of Technology, in 1993 and 1997, respectively.

She joined the Department of Electrical and Control Engineering of National Chiao-Tung University, Taiwan, in 1997. Her research interests include digital signal processing, multirate filter banks, and signal processing for digital communications. She has coauthored two books, "Signal Processing and Optimization for Transceiver Systems," and "Filter Bank Transceivers for OFDM and DMT Systems," (Cambridge University Press, 2010).

She was a recipient of Ta-You Wu Memorial Award in 2004. She served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II, the IEEE SIGNAL PROCESSING LETTERS, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I, the *EURASIP Journal on Applied Signal Processing*, and Multidimensional Systems and Signal Processing, Academic Press. She was a Distinguished Lecturer of the IEEE Circuits and Systems Society for 2006-2007.