PREDICTIVE QUANTIZATION OF BIT LOADING FOR MIMO TIME-CORRELATED CHANNELS WITH LIMITED FEEDBACK

Yuan-Pei Lin and Chi-Chang Wu

Dept. Electrical Engr., National Chiao Tung Univ., Hsinchu, Taiwan

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, so is bit loading. We propose to feedback bit loading using predictive coding, which is known to be a powerful technique for quantizing correlated signals. Assuming the channel is a first-order Gauss-Markov random process, we derive the optimal predictor for the bit loading to be coded. We show that the subchannels prediction errors are approximated Gaussian and thus can be quantized using quantizers designed for Gaussian random variables. Simulations demonstrate that the proposed predictive coding can achieve a very good approximation of the desired transmission rate with a very low feedback rate.

1. INTRODUCTION

It has been shown that with limited amount of feedback information the performance of a transmission system can be enhanced greatly. In general, the transmitter has no knowledge of the forward link channel and only the receiver has the channel state information. 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]. For fading channels, adapting the transmission rate according to the channel state information can lead to considerable gain over fixed-rate systems [2]. Rate adaptation is also important for controlling frame error rate or packet error rate without using deep interleaving because we can adjust the rate for a given target error rate [3]. MIMO systems with variable transmission rate have been considered in the literature, e.g, [4], which uses beamforming for low-rate transmission and spatial multiplexing for a higher rate.

In practical transmission environments, a fading channel is usually correlated in time. When the channel is modeled as a first-order Gauss-Markov random process [5], the time correlation can be more directly exploited for analysis or for more efficient feedback [6]-[8]. For a given quantization error constraint, the minimum rate for feeding back differential channel information is derived in [6]. The design of polar-cap differential codebook is addressed in [7] for beamforming systems. A feedback scheme that uses a differential rotation of the precoder is proposed in [8]. 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. The transmission rate is adapted to the current channel condition by assigning bits to the subchannels of the MIMO systems. The bit loading is fed back using predictive quantization, a scheme known to be very efficient for coding signals that are correlated in time. Assuming the channel is modeled by a first-order Gauss-Markov process, we derive the predictor of the bit loading to be coded when the channel is varying slowly. Furthermore we show that the prediction error is approximately Gaussian and can be quantized using quantizers designed for a Gaussian source. Simulations are presented to demonstrate that the proposed predictive coding of bit loading can achieve a very good approximation of the desired transmission rate using a low feedback rate. The sections are organized as follows. In Sec. 2, we give the system model of the time-correlated MIMO system. In Sec. 3., we present the proposed predictive quantization of bit loading. Simulation examples are given in Sec. 4 and a conclusion given in Sec. 5.

2. 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 $M_r \times 1$ chan-

This work was supported in parts by National Science Council, Taiwan, R. O. C., under NSC 101-2221-E-009-091-MY3 and Aiming for the Top University Plan of National Chiao Tung University, the Ministry of Education, Taiwan, under Contract 100W9633 and 101W963.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

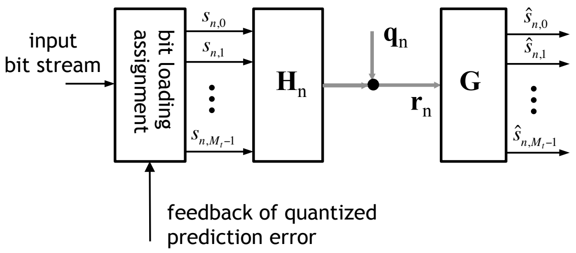

Figure 1: A MIMO system with bit loading feedback.

nel noise $\mathbf{q}_n$ and $M_r \geq M_t$. A useful model of a time-correlated channel is the first-order Gauss-Markov process [5]

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

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. 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.

The inputs of the channel as indicated in Fig. 1 are modulation symbols $s_{n,k}$, for $k = 0, 1, \cdots, M_t - 1$, assumed to be uncorrelated with zero mean and variance P_t/M_t, where P_t is the total transmission power. The $M_t \times M_r$ linear receiver $\mathbf{G}_n$ is zero forcing and $\mathbf{G}_n = (\mathbf{H}_n^\dagger \mathbf{H}_n)^{-1}\mathbf{H}_n^\dagger$ [9], where † denotes conjugate transpose. In this case, the receiver output error $\mathbf{e}_n = \mathbf{G}_n \mathbf{q}_n$ has autocorrelation matrix $\mathbf{R}_n = E[\mathbf{e}_n \mathbf{e}_n^\dagger]$ given by $\mathbf{R}_n = N_0(\mathbf{H}_n^\dagger \mathbf{H}_n)^{-1}$. The kth subchannel error variance at time n is $\sigma^2_{e_{n,k}} = [\mathbf{R}_n]_{kk}$. When QAM modulation symbols are used, the number of bits that can be loaded on the kth subchannel at time n is [2]

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

where $\Gamma = -\ln(5BER)/1.5$ depends on the target bit error rate (BER). The number of bits transmitted at time n is $\mathcal{R}_n = \sum_{k=0}^{M-1} b_{n,k}$.

3. PREDICTIVE QUANTIZATION OF BIT LOADING

When the channel is time-correlated, we can expect the bit loading to be time-correlated as well. For the predictive quantization of $b_{n,k}$, both the transmitter and receiver compute a predicted value $\widetilde{b}_{n,k}$. Quantization is applied on the prediction error $\delta_{n,k} = b_{n,k} - \widetilde{b}_{n,k}$ and the quantized version $\widehat{\delta}_{n,k}$ is fed back to the transmitter. The transmitter reproduces the quantized bit loading using $\widehat{b}_{n,k} = \widetilde{b}_{n,k} + \widehat{\delta}_{n,k}$. In this case the quantization error $b_{n,k} - \widehat{b}_{n,k}$ is the same as $\delta_{n,k} - \widehat{\delta}_{n,k}$. With judicious design, the prediction error has a smaller variance than $b_{n,k}$ and a smaller quantization error can be achieved as quantization error is proportional to the variance of the signal to be quantized [10].

We first show how to predict the bit loading at time n given the channel at time $n-1$. Using the Gauss-Markov channel model in (1), the bit loading at time n can be approximated as in the following lemma.

Lemma 1 *For a small ϵ, the number of bits that can be loaded on the kth subchannel at time n can be approximated in terms of the bit loading at time $n-1$ as*

$$b_{n,k} \approx b_{n-1,k} - \frac{\epsilon}{\ln 2}\frac{[\mathbf{A}_n]_{kk}}{\sigma^2_{e_{n-1,k}}}, \tag{3}$$

where

$$\mathbf{A}_n = -\frac{1}{N_0}\mathbf{R}_{n-1}(\mathbf{H}_{n-1}^\dagger \mathbf{W}_n + \mathbf{W}_n^\dagger \mathbf{H}_{n-1})\mathbf{R}_{n-1} \tag{4}$$

and $[\mathbf{X}]_{i,j}$ denotes the (i,j)th entry of a matrix $\mathbf{X}$.

Proof. We can rewrite $\mathbf{R}_n = N_0(\mathbf{H}_n^\dagger \mathbf{H}_n)^{-1}$ as $\mathbf{R}_n = \frac{1}{1-\epsilon^2}\mathbf{R}_{n-1}(\mathbf{I}+\mathbf{E})^{-1}$, 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, [12]) when $\mathbf{E}$ satisfies $||\mathbf{E}|| < 1$, where $||\mathbf{E}||$ denotes a certain matrix norm of $\mathbf{E}$, e.g., Frobenius norm, then $(\mathbf{I}+\mathbf{E})^{-1}$ 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 $(\mathbf{I}+\mathbf{E})^{-1} \approx \mathbf{I} - \mathbf{E}$ and ignoring higher-order terms of ϵ, we have the approximation $\mathbf{R}_n \approx \mathbf{R}_{n-1} + \epsilon \mathbf{A}_n$. It follows that the kth subchannel error variance at time n is given by $\sigma^2_{en,k} \approx \sigma^2_{e_{n-1,k}} + \epsilon[\mathbf{A}_n]_{kk}$. Using this expression and (2) we obtain

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

When ϵ is small, we can approximate the numerator $\sigma^2_{e_{n-1,k}} + \epsilon[\mathbf{A}_n]_{kk} + P_t/(M_t\Gamma)$ as $\sigma^2_{e_{n-1,k}} + P_t/(M\Gamma)$. In this case we can rearrange it as

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

We recognize the first term is equal to $b_{n-1,k}$, the number of bits loaded on the kth subchannel at time $n-1$ in (2). Using the Taylor series approximation of the second term at $\epsilon = 0$, we arrive at (3). $\triangle$

We would like to design the predictor $\widetilde{b}_{n,k}$ so that the mean squared prediction error $E[\delta_{n,k}^2]$ is minimized. It is known that [10], given the previous channel $\mathbf{H}_{n-1}$, the best predictor is the conditional mean, $\widetilde{b}_{n,k}^{opt} = E_{\mathbf{W}_n}[b_{n,k}|\mathbf{H}_{n-1}]$, 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 1 leads to the following expression of the optimal predictor,

$$\widetilde{b}_{n,k}^{opt} \approx b_{n-1,k}, \tag{5}$$

where we have used the fact that the entries of $\mathbf{W}_n$ have zero mean, and thus the conditional mean $E_{\mathbf{W}_n}[\mathbf{A}_n|\mathbf{H}_{n-1}] = \mathbf{0}$. Although the optimal predictor is the condition mean given the previous channel, it turns out to depend on the previous bit loading only and is independent of ϵ. With the optimal prediction, the prediction error is given by

$$\delta_{n,k} \approx -\frac{\epsilon}{\ln 2}\frac{[\mathbf{A}_n]_{kk}}{\sigma_{e_{n-1,k}}^2}. \tag{6}$$

Observe that the entries of $\mathbf{A}_n$ are Gaussian. Thus $\delta_{n,k}$ is approximately Gaussian and can be quantized using quantizers designed for Gaussian sources [11]. For the quantization of Gaussian random variables, the optimal quantizer is well known [11]. Note that the prediction error is in the order of ϵ as both $\mathbf{A}_n$ and $\sigma_{e_{n-1,k}}^2$ are independent of ϵ. The result 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.

On the other hand, the prediction error variance, if known to the transmitter, is a useful reference for designing the quantizer so that a smaller quantization error can be achieved [11]. In particular the optimal reconstruction points for quantizing a Gaussian random variable can be explicitly expressed in terms of its variance [11]. Based on (6), it can be shown that the variance of $\delta_{n,k}$, denoted as $\mathcal{E}_{n,k}$, has the second-order approximation

$$\mathcal{E}_{n,k} \approx \frac{2\epsilon^2}{(\ln 2)^2 N_0}\sigma_{e_{n-1,k}}^{-2}[\mathbf{R}_{n-1}^2]_{kk}. \tag{7}$$

It contains the term $[\mathbf{R}_{n-1}^2]_{kk}$ that is not available to the transmitter as only bit loading is fed back. Note that $[\mathbf{R}_{n-1}^2]_{kk}$ can be bounded as

$$\sigma_{e_{n-1,k}}^4 \leq [\mathbf{R}_{n-1}^2]_{kk} \leq \sigma_{e_{n-1,k}}^2 \sum_{\ell=0}^{M-1} \sigma_{e_{n-1,\ell}}^2. \tag{8}$$

The upper bound is obtained by using $|E[e_{n-1,k}e_{n-1,\ell}^*]| \leq \sigma_{e_{n-1,k}}\sigma_{e_{n-1,\ell}}$. The variance $\sigma_{e_{n-1,\ell}}^2$ can be obtained from the bit loading using (2), i.e., $\sigma_{e_{n-1,\ell}}^2 = P_t/(M_t\Gamma(2^{b_{n-1,\ell}} - 1))$. Using the bounds in (8), we have

$$\frac{2\epsilon^2}{(\ln 2)^2}\frac{P_t/(M_t\Gamma N_0)}{2^{b_{n-1,k}}-1} \lesssim \mathcal{E}_{n,k} \lesssim \frac{2\epsilon^2}{(\ln 2)^2}\sum_{\ell=0}^{M_t-1}\frac{P_t/(M_t\Gamma N_0)}{2^{b_{n-1,\ell}}-1}. \tag{9}$$

With the above bounds, the transmitter can approximate $\mathcal{E}_{n,k}$. The approximation can then be used to design the reconstruction points of the quantizers. For example, when we use one bit to quantize, the optimal reproduction points are $\pm\sqrt{2\mathcal{E}_{n,k}/\pi}$ [11].

In practical predictive quantization, only the past quantized bit loading is available to the transmitter. Replacing $b_{n-1,k}$ by the quantized version $\widehat{b}_{n-1,k}$ in (5), we have $\widetilde{b}_{n,k} = \widehat{b}_{n-1,k}$. The receiver applies quantization on the prediction error $\delta_{n,k} = b_{n,k} - \widehat{b}_{n-1,k}$ and sends back the quantized prediction error $\widehat{\delta}_{n,k}$ to the transmitter. The transmitter reconstructs the quantized bit loading by $\widehat{b}_{n,k} = \widehat{b}_{n-1,k} + \widehat{\delta}_{n,k}$. The kth subchannel at time n is loaded with $\lfloor\widehat{b}_{n,k}\rfloor$ bits, where $\lfloor x\rfloor$ denotes the largest integer smaller or equal to x.

4. SIMULATION EXAMPLES

In the following examples, the channel $\mathbf{H}_n$ is generated using the Gauss-Markov model in (1) for $M_r = 6$, $M_t = 4$ and $M = 4$. We have used $f_c = 2.5 * 10^9\ Hz$, and $T = 2\ ms$ as suggested in [13]. In an indoor or microcellular transmission scenario, the terminal speed of interest is $3\ km/hr$ [13], which corresponds to $\epsilon = 0.06$. In the simulations, $P_t/N_0 = 15$ dB and the target BER is 10^{-4}.

Example 1. In this example we demonstrate that the prediction error $\delta_{n,k}$ is approximately Gaussian, given the previous channel $\mathbf{H}_{n-1}$. Fig. 2 shows the histogram of $\delta_{n,0}$ for a randomly chosen $\mathbf{H}_{n-1}$ using 10^5 realizations of $\mathbf{W}_n$. The histogram is plotted for $\epsilon = 0.06$ and 0.1. The dash and solid lines correspond to the pdf (probability density function) of zero-mean Gaussian random variables with variance computed using (7). For $\epsilon = 0.06$, the histogram is very close to the Gaussian pdf. We see that the prediction errors are well approximated by Gaussian random variables, especially for a small ϵ.

Example 2. *Quantized transmission rate.* Fig. 3 shows the unquantized and quantized transmission rates for $\epsilon = 0.06$ and feedback rate $B = 1$. The prediction error $\delta_{n,k}$ is quantized using a one-bit codebook for Gaussian random variables and the prediction errors are interleaved for feedback. 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 (2). For the quantized rate $\widehat{\mathcal{R}}_n = \sum_{k=0}^{M-1}\widehat{b}_{n,k}$,

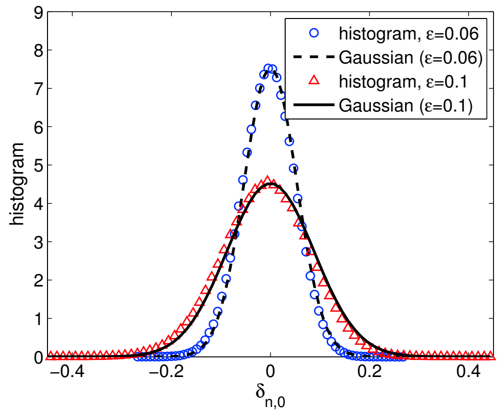

Figure 2: The histograms of $\delta_{n,0}$ and pdf of zero-mean Gaussian random variables with the same variances for $\epsilon = 0.06$ and 0.1.

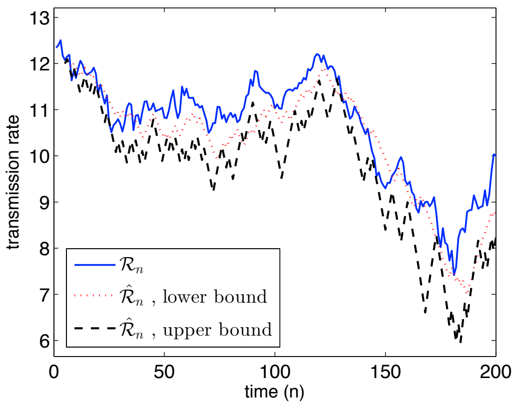

Figure 3: The unquantized and quantized transmission rates.

we have designed the quantizers using the lower and upper bounds in (9). The mean squared quantization error $E[(\mathcal{R}_n - \widehat{\mathcal{R}}_n)^2]$ obtained by averaging over 10^5 channel realization is 0.1582 when the upper bound is used and 0.0571 when the lower bound is used; the lower bound is a better approximation of the prediction error variance and a smaller quantization error can be achieved. In this case the quantized rate is a good approximation of $\mathcal{R}_n$ even for $B = 1$.

5. CONCLUSION

In this paper, we considered variable-rate transmission for time-correlated MIMO channels with limited feedback. The transmission rate is adapted through dynamic bit assignment according to the current channel. We proposed the use of predictive quantization for the quantization of bit loading. Given the previous channel, the optimal predictor turns out to depend only on the previous bit loading. We show through simulations that a very small quantization error can be achieved with a small feedback rate.

6. 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 J-SAC*, vol. 26, no. 8, pp. 1341-1365, Oct. 2008.

[2] A. J. Goldsmith and S.-G. Chua, “Variable-rate variable-power MQAM for fading channels,” *IEEE Trans. Communi.*, vol. 45, no. 10, pp. 1218-1230, Oct. 1997.

[3] T. Wu and V. K. N. Lau, “Robust rate, power and precoder adaptation for slow fading MIMO channels with noisy limited feedback ,” *IEEE Trans. on Wireless Communi.*, vol. 7, no. 6, pp. 2360-2367, June 2008.

[4] 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. Vehi. Tech.*, vol. 56, no. 5, pp. 2555-2567, Sept. 2007.

[5] 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, April 2006.

[6] L. Zhang, L. Song, M. Ma, and B. Jiao, “On the minimum differential feedback for time-correlated MIMO Rayleigh block-fading channels,” *IEEE Trans. Communi.*, vol. 60, no. 2, pp. 411-420, Feb. 2012.

[7] 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 Communi.*, vol. 11, no. 2, pp. 703-711, Feb. 2012.

[8] T. Kim, D. J. Love, and B. Clerckx, “MIMO systems with limited rate differential feedback in slowly varying channels” *IEEE Trans. Communi.*, vol. 59, no. 4, pp. 1175-1189, April 2011.

[9] P. P. Vaidyanathan, See-May Phoong, and Yuan-Pei Lin, Signal processing and optimization for transceiver systems, Cambridge University Press, April 2010.

[10] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic, 1991.

[11] K. Sayood, *Introduction to Data Compression*, 3rd. ed., Morgan Kaufmann, 2005.

[12] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge, U.K.: Cambridge Univ. Press, 1985.

[13] ITU-R, Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced, Report ITU-R M.2135-1, Dec. 2009.