

RESEARCH

Open Access



# Hybrid digital-analog coding with bandwidth expansion for correlated Gaussian sources under Rayleigh fading

Pradeepa Yahampath

## Abstract

Consider communicating a correlated Gaussian source over a Rayleigh fading channel with no knowledge of the channel signal-to-noise ratio (CSNR) at the transmitter. In this case, a digital system cannot be optimal for a range of CSNRs. Analog transmission however is optimal at all CSNRs, if the source and channel are memoryless and bandwidth matched. This paper presents new hybrid digital-analog (HDA) systems for sources with memory and channels with bandwidth expansion, which outperform both digital-only and analog-only systems over a wide range of CSNRs. The digital part is either a predictive quantizer or a transform code, used to achieve a coding gain. Analog part uses linear encoding to transmit the quantization error which improves the performance under CSNR variations. The hybrid encoder is optimized to achieve the minimum AMMSE (average minimum mean square error) over the CSNR distribution. To this end, analytical expressions are derived for the AMMSE of asymptotically optimal systems. It is shown that the outage CSNR of the channel code and the analog-digital power allocation must be jointly optimized to achieve the minimum AMMSE. In the case of HDA predictive quantization, a simple algorithm is presented to solve the optimization problem. Experimental results are presented for both Gauss-Markov sources and speech signals.

**Keywords:** Hybrid digital-analog coding, Predictive quantization, Transform coding, Fading channels, Speech coding

## 1 Introduction

In digital communication over a fading channel, the best performance is achieved when both the transmitter and the receiver are adapted to the channel state. If the channel-state information (CSI) is available, the transmitter can adapt coding and modulation to maintain the optimal performance at all times. However, there are common situations in which the transmitter adaptation is not an option. One obvious example is broadcasting where a single transmitter sends information to multiple receivers. Since the channels to different receivers may not be the same, it is not possible to adapt the transmitter to a specific channel state. Another example is when there is no possibility of CSI feedback from a mobile receiver to the transmitter. In either case, the receiver suffers from the “cliff effect” [1]—when channel signal-to-noise ratio (CSNR) decreases, at some point, a less than 1 dB drop

in CSNR can take the decoder from perfect operation to complete failure (threshold effect), and when the CSNR increases from this point, the decoder output quality remains fixed regardless of the CSNR (see for example [2] (Fig. 5)). One solution to this problem is multi-resolution coding and modulation [1, 3, 4]. This scheme does not entirely eliminate the cliff effect but improves it to a stair-case effect. For analog sources, a better alternative is *hybrid digital-analog* (HDA) coding [1, 5, 6] which is the focus of this paper.

It is known that uncoded or *analog* transmission achieves the optimal performance theoretically attainable (OPTA) in MMSE sense when both the source and the channel are Gaussian and memoryless and have the same bandwidths [7]. Clearly, uncoded transmission cannot be optimal for sources with memory and when the source and channel bandwidths are not matched. For sources with memory, widely used digital source-coding techniques such as predictive quantization (PQ) transform coding (TC) [8] exploit source memory to achieve a coding gain and will outperform uncoded transmission

Correspondence: Pradeepa.Yahampath@umanitoba.ca  
Department of Electrical and Computer Engineering, University of Manitoba,  
Winnipeg, MB R3T 5V6, Canada

if both the transmitter and the receiver have CSI. However, systems based on these techniques still suffer from the aforementioned cliff effect when the transmitter has no CSI. On the other hand, implementing good analog codes for sources with memory is difficult. A promising approach to benefit from both the robustness of analog transmission against CSNR variations and the source-coding gain due to source correlation is HDA coding. Fundamentally, HDA transmission involves the simultaneous transmission of a source in both digital and analog forms. Most previous work on HDA coding have used a form of layered transmission in which the base layer is digitally coded, and the quantization error of the base layer is transmitted as a refinement layer, using analog pulse amplitude (PAM) modulation [2, 9–12]. While a considerable amount of research has focused on HDA transmission of memoryless sources, much less work has been devoted to developing good HDA codes for sources with memory. In particular, when the source has memory, the optimal HDA coding involves a very different design trade-off compared to coding a memoryless source. The main goal of this paper is to design HDA systems which can simultaneously benefit from high coding gain of PQ or TC and the CSNR-independent optimality of a parallel analog transmission. PQ is the standard technique for moderate to high bit-rate (16–40 kbs) speech coding [13] while TC is a staple in image and video compression.

We consider the transmission of a correlated Gaussian source over a block-fading Gaussian channel whose bandwidth is greater than or equal to the source bandwidth (channel memory is however not considered). In the proposed approach, the source is digitally transmitted using either PQ or TC. The quantization error of the digital encoder is transmitted by linear analog coding over the same channel bandwidth as the digital transmission, by using superposition and power sharing. Given that the transmitter cannot be adapted to the instantaneous CSNR at the receiver, we determine the best analog-digital power allocation by minimizing the average MMSE (AMMSE) with respect to the receiver-CSNR distribution. A closer look at this problem reveals an interesting trade-off between digital and analog transmissions when the source has memory. On the one hand, allocating more power to the digital transmission allows a higher quantization rate and hence a higher predictive or transform coding gain. On the other hand, allocating more power to the analog transmission makes it possible to achieve a greater reduction in distortion as the CSNR increases. The not so obvious variable here that also affects this trade-off is the *outage CSNR* which is the lowest CSNR at which a receiver can decode the digital signal. For the same power allocation, a higher quantization rate can be chosen at the expense of increased outage

CSNR. Therefore, there exists a non-trivial trade-off between the power allocation, quantization rate, and the outage CSNR.

We also address the problem of determining the power allocation and the outage CSNR (or equivalently the quantization rate) in HDA-PQ and HDA-TC systems to achieve optimal (in AMMSE sense) trade-off. To this end, we obtain analytical expressions for the AMMSE of HDA-PQ and HDA-TC systems by relying on the *high-rate* model of entropy constrained scalar quantizers [14]. Our solutions are therefore asymptotically (in rate) optimal. In general, finding a closed-form solution for the optimal power allocation and outage CSNR appears intractable. However, in the case of HDA-PQ, we identify a simple co-ordinate descent algorithm [15] to determine the optimal solution. This algorithm converges rapidly, typically in 2–3 iterations. We demonstrate that it is quite possible to implement good practical finite-rate HDA-PQ and HDA-TC systems using the asymptotically optimal solutions. Experimental results obtained with Gauss-Markov processes as well as speech signals modeled as a Gaussian auto-regressive (AR) process show that both the system AMMSE and the MMSE of a receiver operating at a given CSNR of practical designs closely match those given by the asymptotic expressions, when the quantization rate is higher than about 1 bit/sample. Our results show that, for highly correlated sources, the HDA systems can substantially outperform both purely digital and purely analog transmission over a wide range of receiver CSNRs.

### 1.1 Main contribution and related previous work

Compared to previous work on HDA coding of Gaussian sources with memory, the main contribution of this paper is the joint optimization of power allocation and quantization rate of HDA systems based on PQ or TC, with respect to the AMMSE criterion. This optimization problem does not arise when the source is memoryless. We also provide a lower bound to the AMMSE achievable for source with memory, which can be numerically computed for a Gauss-Markov source.

Previously, HDA coding of correlated sources have appeared in [2, 9–12, 16–18]. With the exception of [18], none of these work uses the AMMSE as a criterion for power allocation. While [18] uses the AMMSE, their problem is analog-only transmission of unquantized video DCT coefficients over a fast fading channel. The objective of the power allocation in that case is to benefit from channel-diversity. Therefore, power is allocated among consecutive analog transmissions. As a result, their formulation leads to a mixed discrete and continuous optimization problem which has been solved by a heuristic approach unrelated to ours. The other work cited above does not consider the joint optimization of the power allocation and the quantization rate. Phamdo and Mittal [2]

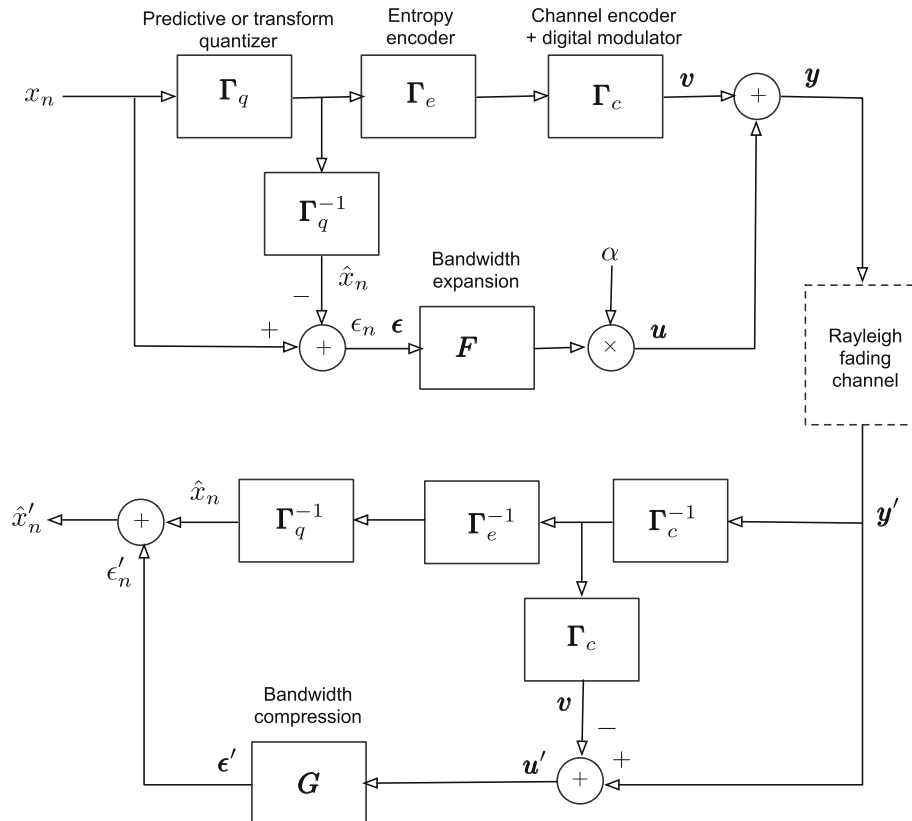
present an implementation of an HDA system for low-bit-rate speech transmission based on the standard FS 1016 CELP codec, by using two independent channels with identical CSNRs for digital and analog transmissions (hence identical power allocations). Yu et al. [9] present similar HDA scheme for video transmission based on H.264/AVC codec but use channel superposition of analog and digital components the power allocation between which is determined by assuming a worst-case CSNR. In [10–12], channel optimized vector quantizers (COVQ) are used as the digital encoder whose quantization error is transmitted in analog form. However, no method for optimizing the power allocation is given. An HDA transform coding scheme is considered in [16], where the analog and digital components are transmitted by time-division multiplexing using equal powers. To the authors' knowledge, HDA schemes based on linear predictive quantization have not been reported so far.

The rest of this paper is organized as follows. Section 2 describes the HDA system considered in this paper and derives an expression for the decoder MSE. Section 3 finds expressions for the MMSE of asymptotically optimal HDA-PQ and HDA-TC over a Rayleigh fading channel.

Section 4 considers the main optimization problem and presents a simple algorithm for solving the problem in the case of HDA-PQ. Section 5 presents some performance bounds for HDA-PQ and HDA-TC systems. Section 6 presents numerical and experimental results, and concluding remarks are given in Section 7.

## 2 HDA transmission of correlated Gaussian sources over fading channels

A block diagram of the HDA transmission system considered in this paper is shown in Fig. 1. Let the source  $\{x_n\}$  be a discrete-time Gaussian process obtained by Nyquist sampling of a correlated analog signal with bandwidth  $W_s$ . Let  $E\{x_n\} = 0$ ,  $E\{x_n^2\} = \sigma_X^2$ , and the correlation coefficients  $r_X(i) = E\{x_n x_{n-i}\} / \sigma_X^2$ ,  $i = 1, 2, \dots$ . This source is to be transmitted over a Rayleigh fading channel with bandwidth  $W_c$  using an average power of  $\mathcal{P}_T$ . It is assumed that the channel has *slow fading* so that the channel gain does not significantly change during a single codeword. We are concerned with systems which allow bandwidth expansion. That is,  $W_c \geq W_s$ , and each source sample is transmitted in  $b = \frac{W_c}{W_s}$  channel uses, where  $b$  is the *bandwidth expansion factor*.



**Fig. 1** The baseband equivalent of the HDA system considered in this paper

In our HDA system, the source  $\{x_n\}$  is quantized by using either a PQ or a TC ( $\Gamma_q$  in Fig. 1) to exploit the memory. The resulting bitstream is entropy coded ( $\Gamma_e$ ) and transmitted after channel coding and modulation ( $\Gamma_c$ ). In addition, analog quantization error  $\epsilon_n = x_n - \hat{x}_n$  ( $\hat{x}_n$  is the quantized value of  $x_n$ ) is also transmitted over the same channel bandwidth as the digital modulator output by using superposition. This is achieved as follows. Since the sequence  $\{\epsilon_n\}$  has a bandwidth of  $W_s$ , bandwidth expansion by a factor of  $b$  is first applied to  $\{\epsilon_n\}$ . The expanded sequence of samples are then converted to a channel signal by using pulse amplitude modulation (PAM) which is superimposed on the digital modulator output for transmission. Bandwidth expansion can be achieved by using an  $L \times M$  linear transform matrix  $\mathbf{F}$  such that  $\frac{L}{M} = b$ , which maps a vector of  $M$  consecutive quantization error samples  $\boldsymbol{\epsilon}$  to an  $L$ -dimensional channel sample vector. In this paper, the frame operator of a uniform tight frame (UTF) [19] is used as  $\mathbf{F}$ . For a UTF, one has  $\mathbf{F}^T \mathbf{F} = b \mathbf{I}_M$ , where  $\mathbf{I}_M$  denotes the  $M \times M$  identity matrix. A simple class of UTFs is the harmonic frames; see [19] for details. The PAM channel input vector (in discrete-time baseband equivalent form)  $\mathbf{v} = \alpha \mathbf{F} \boldsymbol{\epsilon}$  is superimposed on the  $L$ -dimensional digitally modulated vector  $\mathbf{u}$  which is the result of applying entropy coding, channel coding, and modulation to  $M$  quantizer outputs whose errors are in  $\boldsymbol{\epsilon}$ . The amplification factor  $0 \leq \alpha \leq 1$  controls the power output of the analog modulator. Given a total average transmitter power of  $\mathcal{P}_T$ , let  $\mathcal{P}_a = \rho \mathcal{P}_T$  and  $\mathcal{P}_d = (1 - \rho) \mathcal{P}_T$  be the fractions of total power allocated to analog and digital transmissions, respectively, where  $0 \leq \rho < 1$ ,

$$\mathcal{P}_d = \frac{1}{L} E\{\|\mathbf{u}\|^2\}, \quad (1)$$

$$\mathcal{P}_a = \frac{1}{L} E\{\|\mathbf{v}\|^2\} = \alpha^2 \sigma_\epsilon^2, \quad (2)$$

and  $\sigma_\epsilon^2 = E\{\epsilon_n^2\}$  is the quantization error variance. Therefore, the amplification factor  $\alpha = \sqrt{\frac{\rho \mathcal{P}_T}{\sigma_\epsilon^2}}$ .

The channel input is the sum  $\mathbf{y} = \mathbf{u} + \mathbf{v}$  (see Fig. 1). For simplicity, we will assume that the baseband equivalent of the channel input and output are real valued, but they could equally well be complex valued. The channel output is given by

$$\mathbf{y}' = g\mathbf{u} + g\mathbf{v} + \mathbf{w}, \quad (3)$$

where  $\mathbf{w}$  is the  $L$ -dimensional Gaussian channel noise vector with the covariance matrix  $\mathbf{C}_w = \sigma_c^2 \mathbf{I}_L$  and  $g$  is the channel gain which is assumed to remain constant for the duration of an  $L$ -dimensional channel symbol  $\mathbf{y}$ . Let the CSNR at the receiver input be  $\theta = \frac{\gamma \mathcal{P}_T}{\sigma_c^2}$ , where  $\gamma = g^2$  is the channel power gain. It is assumed that  $\theta$  is known to the decoder (but not to the transmitter).

The total noise component at the input to the digital channel decoder  $\Gamma_c^{-1}$  consists of Gaussian channel noise  $\mathbf{w}$  and the interference  $g\mathbf{v}$  from the analog transmission which in general will not be Gaussian<sup>1</sup>. The distribution of the combined noise  $\mathbf{z} = g\mathbf{v} + \mathbf{w}$  is difficult to find. However, for a given noise variance, the capacity of a channel is the lowest when  $\mathbf{z}$  is an iid Gaussian vector ([20], Theorem 7.4.3). This capacity lower bound can be found by evaluating the capacity of an AWGN channel at the CSNR  $(1 - \rho)\gamma \mathcal{P}_T / (\sigma_c^2 + \rho\gamma \mathcal{P}_T)$ , which is given by  $C_{min}(\rho, \theta) = \frac{1}{2} \log_2 \left( 1 + \frac{(1-\rho)\theta}{1+\rho\theta} \right)$  bits/channel use [21]. We assume that, for a given  $\rho$ , the maximum allowable transmission rate at a given  $\theta$  is  $C_{min}(\rho, \theta)$ . Suppose the channel code used for digital transmission is designed for some channel state  $\theta_o \leq \theta$ . It follows that the maximum allowable bit-rate (in bits/sample) of the quantizer is given by

$$R(\rho, \theta_o) = \frac{b}{2} \log_2 \left( \frac{1 + \theta_o}{1 + \rho\theta_o} \right). \quad (4)$$

Therefore, the quantization error variance is a function both  $\rho$  and  $\theta_o$ , which we denote by  $\sigma_\epsilon^2(\rho, \theta_o)$ .

For estimating the analog quantization error at the receiver, the digital signal is first canceled out from the channel output by using a locally generated digital channel signal. The quantization error is then linearly estimated from the residual  $\mathbf{v}' = g\alpha \mathbf{F} \boldsymbol{\epsilon} + \mathbf{w}$  as

$$\boldsymbol{\epsilon}' = \mathbf{G} \mathbf{v}', \quad (5)$$

where  $\mathbf{G}$  is a  $M \times L$  matrix. Finally, the source samples are reconstructed as  $\hat{x}'_n = \hat{x}_n + \epsilon'_n$ . From [22] (Theorem 11.1), it follows that the optimal estimator which minimizes  $E\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}'\|^2$  is given by

$$\mathbf{G}^* = g\alpha \mathbf{C}_\epsilon \mathbf{F}^T \left( g^2 \alpha^2 \mathbf{F} \mathbf{C}_\epsilon \mathbf{F}^T + \mathbf{C}_w \right)^{-1}, \quad (6)$$

where  $\mathbf{C}_\epsilon$  is the covariance matrix of  $\boldsymbol{\epsilon}$ . Assuming that the quantization error vector  $\boldsymbol{\epsilon}$  is uncorrelated, we have  $\mathbf{C}_\epsilon = \sigma_\epsilon^2 \mathbf{I}_M$ , and hence,

$$\mathbf{G}^* = \frac{1}{g\alpha} \mathbf{F}^T \left( \mathbf{F} \mathbf{F}^T + \frac{1}{\rho\theta} \mathbf{I}_L \right)^{-1}. \quad (7)$$

The covariance matrix of the corresponding estimation error is given by ([22], Eq. 11.35)

$$\mathbf{C}_{err} = \left( \frac{1}{\sigma_\epsilon^2} \mathbf{I}_M + \frac{g^2 \alpha^2}{\sigma_c^2} \mathbf{F}^T \mathbf{F} \right)^{-1}. \quad (8)$$

The minimum possible end-to-end source reconstruction MSE at CSNR  $\theta$  for given  $\rho$  and  $\theta_o$  is

$$\begin{aligned} D(\rho, \theta_o, \theta) &= \frac{1}{M} \text{trace} \{ \mathbf{C}_{err} \} \\ &= \frac{\sigma_\epsilon^2}{1 + \frac{bg^2 \alpha^2 \sigma_\epsilon^2}{\sigma_\epsilon^2}} \\ &= \frac{\sigma_\epsilon^2(\rho, \theta_o)}{1 + b\rho\theta}, \theta \geq \theta_o. \end{aligned} \quad (9)$$

Notice that the numerator is independent of the receiver CSNR  $\theta$ . The denominator is the factor by which the overall distortion at the receiver is reduced due to the analog transmission of the quantization error. Unlike a purely digital system whose MSE will depend only on  $\theta_o$  independent of the instantaneous CSNR  $\theta$ , an HDA system will have  $D(\rho, \theta_o, \theta) \rightarrow 0$  as  $\theta \rightarrow \infty$ . In the following, we first derive expressions for  $\sigma_\epsilon^2(\rho, \theta_o)$  of asymptotically optimal PQ and TC. We then consider determining optimal  $\rho$  and  $\theta_o$  for a Rayleigh fading channel which minimize the AMMSE.

### 3 Asymptotically optimal quantization in HDA systems

#### 3.1 HDA-PQ

A detailed description of predictive quantization (PQ) can be found in [8]. In summary, a PQ quantizes the prediction error  $e_n = x_n - \tilde{x}_n$  rather than the input sample  $x_n$ , where  $\tilde{x}_n = \sum_{i=1}^K a_i \hat{x}_{n-i}$  is the predicted value of  $x_n$  using a  $K$ -th order linear predictor with coefficients  $a_1, \dots, a_K$ . As usual, the prediction is carried out using the quantized values of the past inputs. Let the quantized value of  $e_n$  be  $\hat{e}_n$  and the quantization error be  $\epsilon_n = e_n - \hat{e}_n$ . The quantized value of  $x_n$  is given by  $\hat{x}_n = \tilde{x}_n + \hat{e}_n$ , and the overall quantization error is  $x_n - \hat{x}_n = \epsilon_n$ . The optimal quantizer and the predictor can be found by minimizing the MSE  $\sigma_\epsilon^2 = E \{ \epsilon_n^2 \}$ . Owing to its non-linear feedback structure, the exact analysis of a PQ is a well-known difficult problem [23, 24]. However, an analytical expression to which the MSE of the optimal PQ converges as the quantizer rate  $R$  grows can be found noting the fact that, as  $R \rightarrow \infty$  (that is, as the size of the maximum quantization interval approaches zero), the closed-loop prediction error for a Gaussian source is also Gaussian and therefore the quantizer MSE approaches the Gish-Pierce asymptotic [14]. In this case, the quantization error variance is given by

$$\sigma_\epsilon^2 = h\sigma_e^2 2^{-2R}, \quad (10)$$

where  $R$  is the rate and  $h = \frac{\sqrt{3}\pi}{2}$  for fixed-rate scalar quantization and  $h = \frac{\pi e}{6}$  for entropy constrained scalar quantization [8] (in the latter case,  $R = H(q)$  is the entropy of the quantizer output  $q_n$ ).

At high rate, the optimal closed-loop predictor approaches the optimal (open-loop) predictor for the

source. It can be shown that the error of the optimal infinite memory linear predictor is an uncorrelated process (error whitening property) [8]. We assume that  $K$  is chosen large enough so that the optimal  $K$ -th order predictor is close to the infinite memory predictor. In this case, the closed-loop prediction error variance  $\sigma_e^2 = E \{ e_n^2 \}$  can be given by

$$\sigma_e^2 = A^2 \sigma_\epsilon^2 + \sigma_o^2, \quad (11)$$

where  $\sigma_o^2$  is the prediction error variance of the optimal predictor and  $A^2 = \sum_{i=1}^K a_i^2$  is the energy of the predictor impulse response. The coefficients of the optimal  $K$ -th order predictor for  $\{x_n\}$  are given by  $\mathbf{a} = \mathbf{R}_X^{-1} \mathbf{r}_X$ , where  $\mathbf{a} = (a_1, \dots, a_K)^T$ ,  $(i, j)$  element of the  $K \times K$  Toeplitz matrix  $\mathbf{R}_X$  is  $r_X(|i - j|)$ , and  $\mathbf{r}_X = [r_X(1), \dots, r_X(K)]^T$  [8]. The variance of the optimal prediction error is given by  $\sigma_o^2 = \sigma_X^2 (1 - \mathbf{r}_X^T \mathbf{R}_X^{-1} \mathbf{r}_X)$ . Now, from (10) and (11), it follows that the MSE of the optimal PQ as  $R \rightarrow \infty$  is given by

$$\sigma_\epsilon^2 = \frac{h\sigma_o^2 2^{-2R}}{1 - c_0 2^{-2R}}, \quad (12)$$

where for convenience, we define the constant  $c_0 \triangleq hA^2$ . We refer to a PQ which satisfies (12) as an *asymptotically optimal* PQ. The related work on high-rate analysis of predictive quantizers can be found in [23–25].

The maximum allowable quantization rate of the HDA system is given by (4). Relying on the asymptotic expression (12), the minimum possible quantization error variance for given  $\rho$  and  $\theta_o$  is therefore given by

$$\sigma_\epsilon^2(\rho, \theta_o) = \frac{h\sigma_o^2}{\phi(\rho, \theta_o)^b - c_0}, \quad (13)$$

where  $\phi(\rho, \theta_o) = \left( \frac{1+\theta_o}{1+\rho\theta_o} \right) > c_0^{1/b}$ . Clearly,  $\phi$  is monotonic increasing with  $\theta_o$  and monotonic decreasing with  $\rho$ . In order for the high-rate expression (13) to be accurate, we need that  $\sigma_\epsilon^2/\sigma_o^2 \ll 1$ , and therefore,

$$\phi(\rho, \theta_o) \gg [h(1 + A^2)]^{1/b}. \quad (14)$$

In other words, sufficient channel bandwidth must be available to support a high enough quantization rate. With  $b$  and  $\theta_o$  fixed, increasing  $\rho$  reduces the allowable quantization rate. Hence, the high-rate model (13) is valid only for “small”  $\rho$ . However, as will be seen in Section 6, HDA-PQ provides a useful coding gain only in this regime (typically  $\rho < 30\%$ ) anyway, as higher  $\rho$  results in low quantization rates at which predictive coding does not yield a considerable gain over pure analog transmission.

Before proceeding, it is worth noting that when prediction is good, the prediction error resembles a white Gaussian process [8]. For the transmission of the latter, analog transmission will be nearly optimal (exactly optimal if the source and channel are bandwidth matched). However,

transmitting the open-loop prediction error itself in analog form is not possible in predictive quantization as it would result in channel error propagation in a closed-loop decoder.

### 3.2 HDA-TC

A detailed description of transform coding (TC) can be found in [8]. For stationary Gaussian sources, it is known that both PQ and TC can asymptotically (in rate) achieve the same MMSE, provided that PQ uses infinite memory linear prediction and TC uses a Karhunen-Loeve transform (KLT) of infinite dimension [8]. However, at low bit-rates, the performance of PQ for Gaussian sources drops below that of TC, due to the degradation of closed-loop predictions based on quantized samples.

Consider an HDA-TC system which transforms a Gaussian input vector  $\mathbf{X} \in \mathbb{R}^M$  using a  $M \times M$  orthonormal transform  $\mathbf{T}$ . Suppose that the transform coefficients  $\mathbf{S} = \mathbf{T}\mathbf{X}$  are quantized by ECQs with a bit allocation  $\mathbf{r} = (r_1, \dots, r_M)^T$ , where  $R = \frac{1}{M} \sum_{i=1}^M r_i$  is the bit-rate in bits/sample. If we assume asymptotically optimal ECQ, then the quantization error variance of the  $i$ -th transform coefficient  $s_i$  is  $\frac{\pi e}{6} \sigma_{s_i}^2 2^{-2r_i}$  where  $\sigma_{s_i}^2$  is the variance of the transform coefficient  $s_i$ . Note that  $\mathbf{S}$  is an uncorrelated Gaussian vector whose covariance matrix is given by  $\mathbf{T}^T \mathbf{C}_X \mathbf{T}$  where  $\mathbf{C}_X$  is the covariance matrix of  $\mathbf{X}$ . Let the reconstructed value of  $X_i$  from the quantized value of  $\mathbf{S}$  be  $\hat{X}_i$  and the quantization error be  $\epsilon_i = X_i - \hat{X}_i$ . Since  $\mathbf{T}$  is an orthonormal transform, the error variance  $E(X_i - \hat{X}_i)^2 = \sigma_\epsilon^2$  is the same for all  $i = 1, \dots, M$ . In the HDA-TC system, the quantized values of  $\mathbf{S}$  are transmitted digitally and the analog quantization error vector  $\boldsymbol{\epsilon} = \mathbf{X} - \hat{\mathbf{X}}$  is transmitted over the same bandwidth by using linear bandwidth expansion, as in the case of HDA-PQ. Now for given  $\rho$  and  $\theta_o$ , the maximum allowable rate can be found by (4) for which the optimal bit allocation  $r_i^*$ ,  $i = 1, \dots, M$  can be found by minimizing

$$\sigma_\epsilon^2(\rho, \theta_o) = \frac{\pi e}{6M} \sum_{i=1}^M \sigma_{s_i}^2 2^{-2r_i}, \quad (15)$$

subject to  $\sum r_i = R(\rho, \theta_o)$  and  $r_i > 0$ . The Lagrangian formulation of this problem leads to the well known reverse water-filling solution [21]. Without a loss of generality, assume that  $\sigma_{s_1}^2 \geq \sigma_{s_2}^2 \geq \dots \geq \sigma_{s_M}^2$ . Let  $G_m = (\prod_{i=1}^m \sigma_{s_i}^2)^{1/m}$  be the geometric mean of the  $m$  largest variances. Suppose we find  $m \leq M$  such that  $\sigma_{s_i}^2 \geq h G_m 2^{-2R/m}$  for  $i \leq m$  and  $\sigma_{s_i}^2 < h G_m 2^{-2R/m}$  otherwise, where  $h = \frac{\pi e}{6}$ . Then the optimal bit allocation is given by [8]

$$r_i = \begin{cases} \frac{1}{M} R(\rho, \theta_o) + \frac{1}{2} \log_2 \left( \frac{\sigma_{s_i}^2}{G_m} \right) & i = 1, \dots, m \\ 0 & i = m+1, \dots, M. \end{cases} \quad (16)$$

The total MSE of the optimal bit allocation is given by

$$\sigma_\epsilon^2(\rho, \theta_o) = h G_m 2^{-2R(\rho, \theta_o)}, \quad (17)$$

where the integer  $m \leq M$  and hence  $G_m$  is a function of  $\rho$  and  $\theta_o$ . While this solution is simple to determine for any given  $(\rho, \theta_o)$ , unlike (13), it does not seem to have a closed-form expression in terms of  $\rho$  and  $\theta_o$ .

### 4 Robust HDA systems for fading channels

Consider the MSE  $D(\rho, \theta_o, \theta)$  in (9), where  $\theta$  is a random variable (but assumed to remain constant at least for the duration of a single channel codeword), where  $\sigma_\epsilon^2(\rho, \theta_o)$  is given by either (13) or (17). This is the MMSE of an asymptotically optimal HDA-PQ or HDA-TC for a particular  $(\rho, \theta_o)$ . The choice of  $\rho$  and  $\theta_o$  determines how the MMSE varies with the CSNR  $\theta$ . If  $\theta$  is known to the transmitter,  $\rho = 0$  (purely digital) will achieve the lowest MMSE for any  $\theta$ , since in this case  $D(0, \theta_o, \theta) = \sigma_\epsilon^2(0, \theta_o)$  can be minimized by choosing  $\theta_o = \theta$ . In this case, both PQ and TC achieve the maximum possible coding gain. If however the receiver CSNR  $\theta$  is not available to the transmitter, a purely digital system must be designed for some  $\theta_o$  which will be different from  $\theta$ , resulting in a system that is not robust against CSNR variations. On the one hand, the receiver MSE of such a system remains constant even when  $\theta > \theta_o$  despite the increase in the available channel capacity. On the other hand, the channel code and hence the system fail when  $\theta < \theta_o$ , i.e., system goes into outage. We refer to  $\theta_o$  as the outage CSNR of the digital decoder. When the transmitter cannot be adapted to varying  $\theta$ , allocating power to the analog transmission ( $\rho > 0$ ) while keeping  $\theta_o$  fixed will increase the quantization MSE  $\sigma_\epsilon^2(\rho, \theta_o)$  but will make the overall MSE  $D(\rho, \theta_o, \theta)$  to decrease with  $\theta$ . For fixed  $\rho$ , increasing  $\theta_o$  reduces  $\sigma_\epsilon^2(\rho, \theta_o)$  but will increase the *outage probability* and hence the AMMSE. In order to obtain a robust system which is optimal in some sense over a range of  $\theta$ , we design the transmitter for  $\rho$  and  $\theta_o$  which minimizes the AMMSE  $E\{D(\rho, \theta_o, \theta)\}$  with respect to the distribution of  $\theta$ . Such a design is ideal for a system with a single receiver which experiences slow fading or a broadcast environment with a large number of receivers whose empirical CSNR converges to the fading distribution [26].

The AMMSE of HDA-PQ or HDA-TC is given by

$$\bar{D}(\rho, \theta_o) = \begin{cases} E\{D(\rho, \theta_o, \theta) | \theta \geq \theta_o\} (1 - P_o) + \sigma_X^2 P_o & \rho < 1 \\ E\{D_a(\theta)\} & \rho = 1 \end{cases} \quad (18)$$

where  $D_a(\theta) = \frac{\sigma_X^2}{1+b\theta}$  is the MMSE of the optimal analog system and  $P_o = \Pr(\theta < \theta_o)$  is the outage probability, and we assume that in the event of an outage, the decoder output is set to  $\hat{x}'_n = E\{x_n\}$ . It is assumed that the distribution of  $\theta$  is a priori known to the system designer. Our main

focus is the Rayleigh fading channel in which the CSNR  $\theta$  is exponentially distributed [27]. The pdf of  $\theta$  is given by

$$p(\theta) = \frac{1}{\bar{\theta}} \exp\left(-\frac{\theta}{\bar{\theta}}\right), \quad (19)$$

where  $\bar{\theta} = E\{\theta\}$  is the mean CSNR. For the case of Rayleigh fading, from (9), (18), and (19), it follows that

$$\begin{aligned} \bar{D}(\rho, \theta_o) &= \sigma_\epsilon^2(\rho, \theta_o) \int_{\theta_o}^{\infty} \frac{\exp\left(-\frac{\theta}{\bar{\theta}}\right)}{\bar{\theta}(1+b\theta\rho)} d\theta + \sigma_X^2 P_o \\ &= \sigma_\epsilon^2(\rho, \theta_o) \frac{\exp\left(-\frac{1}{b\rho\bar{\theta}}\right)}{b\rho\bar{\theta}} E_1\left(\frac{1+b\rho\theta_o}{b\rho\bar{\theta}}\right) + \sigma_X^2 P_o, \end{aligned} \quad (20)$$

where  $P_o = \left(1 - \exp\left(-\frac{\theta_o}{\bar{\theta}}\right)\right)$  and  $E_1(x) = \int_x^{\infty} \frac{\exp(-t)}{t} dt$  is the exponential integral [28].  $E_1(x)$  is available as a standard function in most numerical software [e.g., `expint(x)` in Matlab]. The AMMSE depends on the choice of the power allocation  $\rho$  and the outage probability, or equivalently  $\theta_o$ . We define the optimal robust HDA system as the one which achieves the minimum AMMSE. The optimal values of  $\rho$  and  $\theta_o$  can be found by solving the problem

$$\begin{aligned} (\rho^*, \theta_o^*) &= \arg \min_{\rho, \theta_o} \bar{D}(\rho, \theta_o) \\ \text{subject to} \quad &0 \leq \rho < 1 \\ &\theta_o > 0. \end{aligned} \quad (22)$$

For fixed  $\theta_o$ ,  $\bar{D}(\rho, \theta_o)$  is convex in  $\rho \in (0, 1)$ . This can be deduced from (20):  $\sigma_\epsilon^2(\rho, \theta_o)$  monotonically increases with  $\rho$  while the term inside the integral monotonically decreases. This represents the trade-off between the coding gain of PQ or TC due to source memory and the robustness against CSNR variations. There must be a

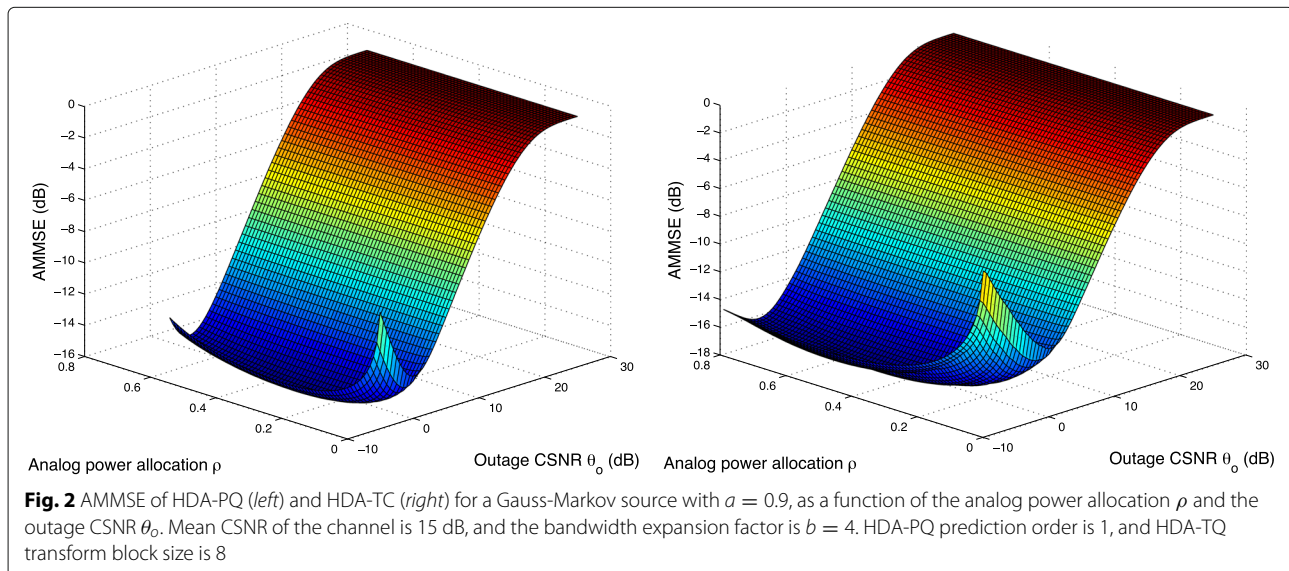
value for  $\rho \in (0, 1)$ , which minimizes the AMMSE. Now if  $\rho$  is fixed,  $\bar{D}(\rho, \theta_o)$  is quasi-convex in  $\theta_o > 0$ . This is because, as  $\theta_o$  is increased ( $P_o$  increases), the first term of the sum in (18)  $E\{D(\rho, \theta_o, \theta) | \theta \geq \theta_o\}$  decreases while the second term  $\sigma_X^2 P_o$  increases. A minimum for  $\bar{D}(\rho, \theta_o)$  occurs for some  $\theta_o < \infty$ . The quasi-convexity follows from the fact that, as  $\theta_o \rightarrow \infty$ , the system will be always in outage and hence  $\bar{D}(\rho, \theta_o) \rightarrow \sigma_X^2$ . Figure 2 shows the AMMSEs of HDA-PQ and HDA-TC as a function of  $\rho$  and  $\theta_o$  for the Gauss-Markov process, which we will refer to as the  $GM(a)$  source,

$$X_n = aX_{n-1} + W_n. \quad (23)$$

Figure 2 illustrates the convexity with respect to  $\rho$  and quasi-convexity with respect to  $\theta_o$ . Below we present an efficient method to determine the optimal solution for  $(\rho, \theta_o)$  in the case of HDA-PQ. Due to the lack of a closed-form expression for the AMMSE, such a simple procedure cannot be devised for HDA-TQ.

#### 4.1 Optimal HDA-PQ

In general, it is difficult to find a closed-form solution to the constrained non-linear minimization problem in (22). In the following, we present a simple *coordinate-descent* (CD) method [15] to solve this problem. In the CD method,  $\bar{D}(\rho, \theta_o)$  is minimized alternately with respect to  $\rho$  (for fixed  $\theta_o$ ) and  $\theta_o$  (for fixed  $\rho_o$ ), until the solution converges. Unlike the joint minimization problem in (22), these two sub-problems are much easier to solve. Since the solution to each problem is conditionally optimal, the CD algorithm is guaranteed to converge to the minimum of  $\bar{D}$ . In actual numerical examples, it was found that this method only required 2–3 iterations to converge. In the





following, we present the solutions to two sub-problems solved in each CD iteration.

Before proceeding, it should be noted that in the case of HDA-PQ, an additional constraint is required to ensure that (14) is not violated. This can be stated as

$$f_1(\rho, \theta_o) < 0, \quad (24)$$

where  $f_1(\rho, \theta_o) \triangleq c_1 - \phi(\rho, \theta_o)$  with  $c_1 = \nu[h(1 + A^2)]^{1/b}$  and  $\nu > 1$  is a sufficiently large constant chosen to ensure that (14) is not violated at low quantization rates. In our experiments, we have used  $\nu = 2$ . If the constraint (24) becomes active, the solution is not guaranteed to be optimal. However, note that, as the quantization rate decreases, the HDA-PQ performance approaches that of analog-only transmission. Therefore, when HDA-PQ outperforms purely analog transmission, (24) is unlikely to be active. For example, with both Gauss-Markov sources and speech signals, numerical results presented in Section 6 show that when  $\rho$  exceeds about 30%, the difference between HDA-PQ and analog systems becomes negligible. It is in this range of  $\rho$  that (24) becomes active.

#### 4.1.1 Optimal power allocation for fixed outage CSNR

For a fixed  $\theta_o$ , optimal power allocation can be found by solving

$$\rho^* = \arg \min_{\rho} \bar{D}(\rho, \theta_o) \quad (25)$$

$$\text{subject to} \quad 0 \leq \rho < \rho_{\max},$$

where  $\rho_{\max} \in (0, 1]$ . In this case, (24) simplifies to

$$\rho_{\max} < \rho_1 \triangleq \frac{(1 + \theta_o)c_1^{-1} - 1}{\theta_o}. \quad (26)$$

and therefore

$$\rho^* = \min\{\rho', \rho_1, 1\}, \quad (27)$$

where  $\rho'$  is the solution to  $f_2(\rho) \triangleq \partial \bar{D} / \partial \rho = 0$ . Using (9) and (13), it can be readily shown that  $f_2(\rho) = 0$  is equivalent to

$$\left[ \frac{b\theta_o}{h\sigma_o^2} \frac{\sigma_\epsilon^2(\rho, \theta_o)\phi^b(\rho, \theta_o)}{1 + \rho\theta_o} - \frac{(1 + b\rho\bar{\theta})}{b\rho^2\bar{\theta}} \right] E_1\left(\frac{1 + b\rho\theta_o}{b\rho\bar{\theta}}\right) + \frac{\exp\left(-\frac{1+b\rho\theta_o}{b\rho\bar{\theta}}\right)}{\rho(1 + b\rho\theta_o)} = 0,$$

which can be solved in the interval  $0 \leq \rho < \rho_{\max}$  using a single-variable root-finding method.

#### 4.1.2 Optimum outage CSNR for fixed power allocation

For fixed  $\rho$ , the optimal outage CSNR can be found by solving

$$\theta_o^* = \arg \min_{\theta_o} \bar{D}(\rho, \theta_o) \quad (28)$$

$$\text{subject to} \quad \theta \geq \theta_{o,\min},$$

where, from (24)

$$\theta_{o,\min} \geq \theta_{o1} \triangleq \max\left\{0, \frac{c_1 - 1}{1 - \rho c_1}\right\}. \quad (29)$$

Using (9) and (13), it can be verified that  $\partial \bar{D} / \partial \theta_o = 0$  is equivalent to

$$\frac{\sigma_X^2}{\bar{\theta}} - \sigma_\epsilon^2(\rho, \theta_o) \left[ \sigma_\epsilon^2(\rho, \theta_o) \frac{(1 - \rho)}{h\sigma_o^2} \frac{(1 + \theta_o)^{b-1}}{(1 + \rho\theta_o)^{b+1}} \frac{\exp\left(\frac{(1+b\rho\theta_o)}{b\rho\bar{\theta}}\right)}{\rho\bar{\theta}} E_1\left(\frac{1 + b\rho\theta_o}{b\rho\bar{\theta}}\right) + \frac{1}{\bar{\theta}(1 + b\rho\theta_o)} \right] = 0 \quad (30)$$

for  $0 < \rho < 1$  and

$$\sigma_X^2 - \sigma_\epsilon^2(0, \theta_o) \frac{(1 + b\bar{\theta})}{1 + \theta_o} = 0 \quad (31)$$

for  $\rho = 0$ . Given  $\rho$ , optimal  $\theta_o$  can be found by locating the root of (30) or (31) in the interval  $\theta_o \in [\theta_{o,\min}, \theta_{o,\max}]$ , where  $\theta_{o,\max}$  is a suitable value chosen to truncate the pdf  $p(\theta)$ .

## 5 Comparisons and performance limits

### 5.1 Analog transmission with block decoding

Consider using only the analog part of the HDA system to transmit the Gaussian AR source. To be useful, any HDA system must perform better than this analog system. Since the sequence of analog channel samples being transmitted is now correlated, the optimal (MMSE) decoder is given by  $\hat{x}_n = E\{x_n | Y_o, \theta\}$  where  $Y_o$  is the observed sequence of channel outputs and  $\theta$  is the receiver CSNR. As  $\{x_n\}$  is a Gaussian sequence, the optimal decoder is linear. In a system with bandwidth expansion  $b = \frac{L}{M}$ , a vector of  $M$  samples  $\mathbf{X}$  from the source is mapped to  $L$  analog channel symbols  $\mathbf{Y} = \alpha \mathbf{F} \mathbf{X}$  where  $\mathbf{F}$  is a UTF and  $\alpha$  is chosen such that the variance of the channel symbols is  $\sigma_Y^2 = \mathcal{P}_T$ . Let  $\mathbf{C}_X$  be the covariance matrix of  $\mathbf{X}$ . The covariance matrix of  $\mathbf{Y}$  is  $\mathbf{C}_Y = \alpha^2 \mathbf{F} \mathbf{C}_X \mathbf{F}^T$ . Therefore, it follows that  $\sigma_Y^2 = \frac{1}{L} \text{trace}\{\mathbf{F}^T \mathbf{F} \mathbf{C}_X\} = \alpha^2 \sigma_X^2$ . Let the corresponding  $L$ -dimensional channel output vector be  $\mathbf{Y}'$ . With a linear decoder, the decoded source vector is given by  $\mathbf{X}' = \mathbf{G}_a \mathbf{Y}'$ , where  $\mathbf{G}_a$  is a  $M \times L$  matrix. Following along the same lines as for (6), we find that the optimal linear decoder is

$$\mathbf{G}_a^* = \frac{1}{g\alpha} \mathbf{C}_X \mathbf{F}^T \left( \mathbf{F} \mathbf{C}_X \mathbf{F}^T + \frac{\sigma_X^2}{\theta} \mathbf{I}_L \right)^{-1}, \quad (32)$$

whose MSE is

$$D_{\text{analog}}(\theta) = \frac{1}{M} \text{trace} \left\{ \left( \mathbf{C}_X^{-1} + \frac{\theta}{\sigma_X^2} \mathbf{F}^T \mathbf{F} \right)^{-1} \right\}. \quad (33)$$

We can write  $\mathbf{C}_X^{-1} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$ , where  $\mathbf{U}$  is the  $M \times M$  matrix whose columns are unit-norm eigenvectors of  $\mathbf{C}_X^{-1}$



and  $\mathbf{\Lambda}$  the  $M \times M$  diagonal matrix whose diagonal elements are  $\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_M}$  where  $\lambda_i, i = 1, \dots, M$  are the eigenvalues of  $\mathbf{C}_X$ . It can be verified that

$$D_{\text{analog}}(\theta) = \frac{\sigma_X^2}{M} \sum_{i=1}^M \frac{\tilde{\lambda}_i}{1 + \tilde{\lambda}_i b \theta}, \quad (34)$$

where  $\tilde{\lambda}_i = \lambda_i / \sigma_X^2$ . For a Rayleigh fading channel with the average CSNR  $\bar{\theta}$ , the AMMSE of the analog system is given by

$$\bar{D}_{\text{analog}} = \frac{\sigma_X^2}{M} \sum_{i=1}^M \tilde{\lambda}_i \frac{\exp\left(\frac{1}{\tilde{\lambda}_i b \bar{\theta}}\right)}{\tilde{\lambda}_i b \bar{\theta}} E_1\left(\frac{1}{\tilde{\lambda}_i b \bar{\theta}}\right). \quad (35)$$

This analog system achieves no coding gain from source correlation, but it does achieve a gain at the receiver due to linear block decoding. Therefore, (35) is not necessarily worse than (21), though it will be so when source correlation is high. However, since sample-by-sample analog encoding and decoding is a special case of HDA coding, (35) is an upper bound to (21) when the source correlation is ignored in (32), that is when  $\tilde{\lambda}_i = 1$  in (35).

## 5.2 HDA vector quantization (HDA-VQ) lower bound

HDA systems considered in this paper can asymptotically achieve performance (AMMSE) that cannot be achieved with either purely analog transmission or purely digital transmission. On an absolute scale, the upper bound to HDA system performance is the optimum performance theoretically attainable (OPTA) when CSI is only available at the receiver. Unfortunately, this bound cannot be determined in any reasonable way, even for a Gaussian source. One obvious upper bound that is easily computed for a Gaussian source is the OPTA when the CSI is available at both transmitter and receiver. This can be found by evaluating the distortion-rate function of the Gaussian process [29] at the rate equal to the capacity of an AWGN channel with the given channel power gain. A more meaningful upper bound for the case when CSI is only available to the receiver can be obtained by replacing the PQ or TC in the HDA coding setup by an optimal (rate-distortion achieving) VQ for the source. The HDA-VQ of a memoryless Gaussian source over a non-fading AWGN channel has previously been studied in [11, 12]. Below, we derive an expression for the AMMSE of HDA-VQ for the  $GM(a)$  source and Rayleigh-fading AWGN channel.

Let the distortion-rate function of  $GM(a_1)$  be  $D_G(R)$ . The latter function is known in closed-form for rates  $R \geq \frac{1}{2} \log_2(1 + a_1)^2 \sigma_X^2$  and in parametric form otherwise [29]. Suppose we use an optimal VQ as  $\Gamma_q$  in the HDA system in Fig. 1. The maximum possible rate achievable at an

outage CSNR of  $\theta_o$  is given by (4). From (9), it follows that the MMSE of the HDA-VQ system at a CSNR of  $\theta$  is

$$D_{\text{HDA-VQ}}(\rho, \theta_o, \theta) = \frac{\delta_G(\rho, \theta_o)}{1 + b\rho\theta}, \quad (36)$$

where  $\delta_G(\rho, \theta_o)$  is  $D_G(R)$  expressed as function of  $\rho$  and  $\theta_o$ . The AMMSE of the HDA-VQ over a Rayleigh fading channel with a mean CSNR of  $\bar{\theta}$  is given by

$$\begin{aligned} \bar{D}_{\text{HDA-VQ}}(\rho, \theta_o) &= \delta_G(\rho, \theta_o) \int_{\theta_o}^{\infty} \frac{1}{\bar{\theta}} \frac{\exp(-\theta/\bar{\theta})}{1 + b\rho\theta} d\theta + \sigma_X^2 P_o \\ &= \frac{\exp(1/(b\rho\bar{\theta}))}{b\rho\bar{\theta}} E_1\left(\frac{1 + b\rho\theta_o}{b\rho\bar{\theta}}\right) \delta_G(\rho, \theta_o) \\ &\quad + \sigma_X^2 P_o. \end{aligned} \quad (37)$$

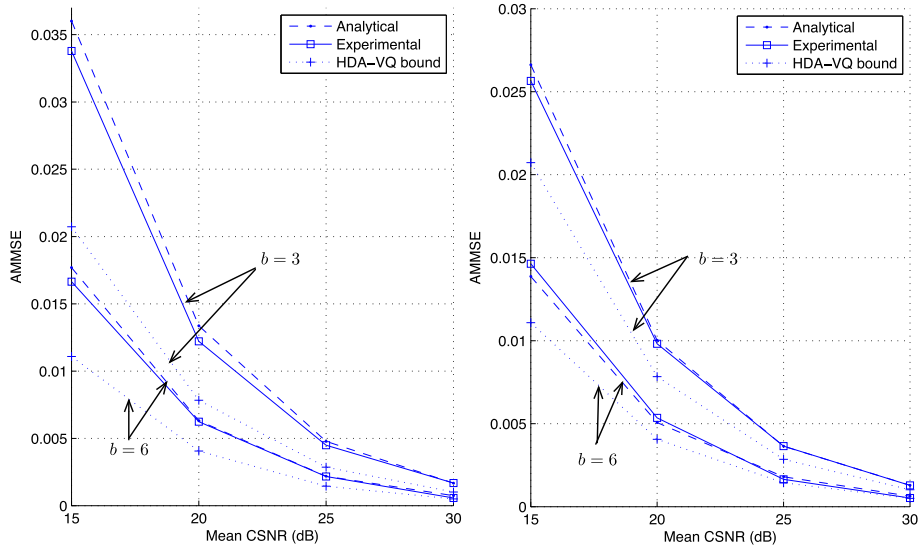
Since neither PQ nor TC can outperform optimal VQ, the AMMSE in (18) is bounded below by the minimum value of  $\bar{D}_{\text{HDA-VQ}}(\rho, \theta_o)$ . There is no apparent simple way to determine this minimum value since a closed-form expression for  $\delta_G(\rho, \theta_o)$  is not available for all  $\rho$  and  $\theta_o$ . Numerical values of this bound shown in Section 6 have been obtained by performing a grid-search over the  $(\rho, \theta_o)$  space where  $0 \leq \rho \leq 1$  and  $0 \leq \theta_o \leq \theta_{o,\max}$  (a suitable upper limit) to determine the minimum of (37).

## 6 Numerical results and discussion

In this section, we use numerical examples to demonstrate the theoretical performance achievable with asymptotically optimal HDA systems as well as the actual performance of finite-rate HDA systems designed using power allocations and quantizer rates obtained through asymptotic analysis. It is useful to compare the minimum AMMSE of actual HDA-PQ and HDA-TC designs with the HDA-VQ bound for the same source-channel pair. While the latter bound can be difficult to evaluate for a general Gaussian source, it can be numerically evaluated for a Gauss-Markov source (Section 5.2). We also compare the HDA systems with the purely analog system in Section 5.1 and purely digital systems (PQ and TC). We do so for both  $GM(a)$  source and speech signals modeled by a Gaussian AR source.

### 6.1 Performance for Gauss-Markov sources

Figure 3 shows the AMMSE as a function of mean CSNR for HDA-PQ and HDA-TC together with the corresponding HDA-VQ upper bound for the  $GM(0.9)$  source. These figures show both the AMMSEs of the asymptotically optimal HDA systems (labeled analytical) obtained by minimizing the expression (18) with respect to  $\rho$  and  $\theta_o$ , as well as the experimental AMMSE of actual HDA-PQ and HDA-TC systems which use these  $(\rho, \theta_o)$  values. For HDA-PQ, a prediction order of 1 has been used while in HDA-TA, a transform block size of 8 has been used. For



**Fig. 3** A comparison of analytical and experimental AMMSEs of HDA systems designed for the unit variance GM(0.9) source, HDA-PQ (left), and HDA-TC (right) ( $b$  is the bandwidth expansion).  $\rho$  and  $\theta_o$  values for HDA-PQ and HDA-TC systems shown here are listed in Table 1. HDA-PQ prediction order is 1, and HDA-TQ transform block size is 8

solving (22) for HDA-PQ, the CD algorithm presented in Section 4.1 was used. On the other hand, for HDA-TC, an exhaustive grid search over the solution space of  $(\rho, \theta_o)$  was used to locate the minimum. Table 1 shows the power allocations ( $\rho^*$ ) and outage CSNRs ( $\theta_o^*$ ) in this manner. Practical HDA PQ/TC systems used ECQs designed (using training set  $10^5$  of source samples) for the rates corresponding to optimal  $(\rho, \theta_o)$  values, by combining the algorithms in [30] and ([8] Table 13.1). These quantizers were used to simulate the HDA encoders and decoders. In order to simplify the simulations, the equivalent digital channel (with channel coding and digital modulation) was assumed error-free for source-coding rates below the capacity of the AWGN channel. The channel output was assumed undecodable at CSNRs below the outage CSNR.

**Table 1** The power allocations and outage CSNRs of HDA-PQ and HDA-TC systems shown in Fig. 3

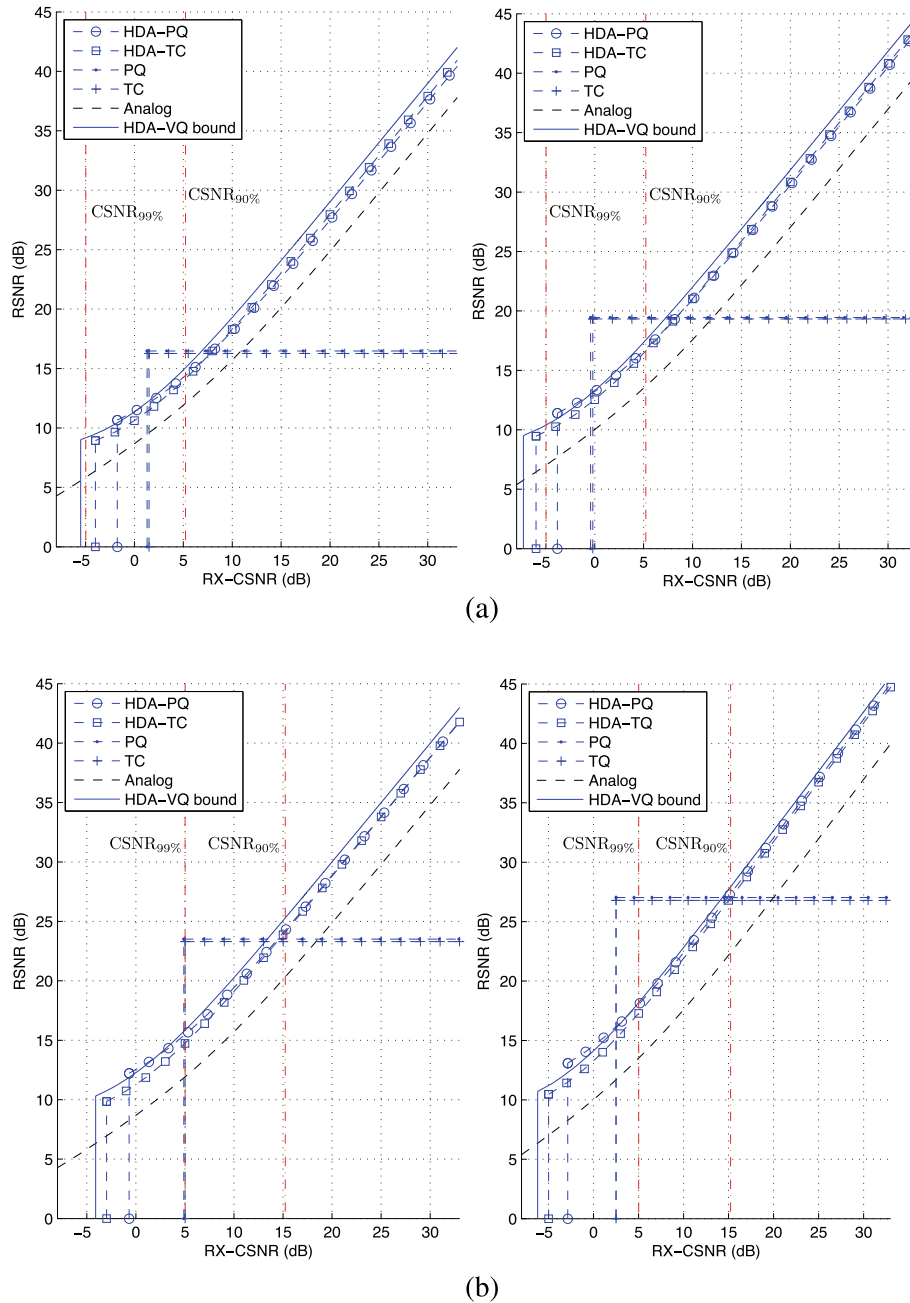
$\bar{\theta}$ (dB)	$b$	HDA-PQ		HDA-TC	
		$\rho^*$ (%)	$\theta_o^*$ (dB)	$\rho^*$ (%)	$\theta_o^*$ (dB)
15	3	27	-1.9	40	-3.8
20	3	27.5	-1.3	42	-3.0
25	3	27.5	-0.8	43	-2.8
30	3	27.5	-0.4	44	-2.5
15	6	28.5	-4.6	43	-6.5
20	6	28.5	-4.1	45	-6.0
25	6	28.5	-3.8	47	-5.3
30	6	28	-3.4	48	-5.5

The AMMSEs of practical HDA-PQ/TC systems were estimated by numerical integration of the receiver MSE  $\hat{D}(\theta)$  over the pdf of CSNR  $\theta$ , where  $\hat{D}(\theta)$  for each  $\theta$  value was determined by Monte-Carlo simulation of the HDA system. As usual, Karhunen-Loeve transform (KLT) [8] has been used as the transform in HDA-TC. Since transform dimensions larger than 8 provided no significant improvement in AMMSE of HDA-TC, the performance shown in Fig. 3 is for  $K = 8$ . Notice that the experimental AMMSE values observed for finite-rate HDA-PQ and HDA-TC systems closely agree with those predicted by high-rate analysis when the available channel capacity is high. At lower mean CSNRs and small bandwidth expansion factors, the AMMSEs of the practical designs are in fact lower than that predicted by high-rate analysis. This is because, at low rates (below about 2 bits/sample), high-rate expressions overestimate the MSE of quantizers. The performance of HDA systems degrades (relative to HDA-VQ bound) as the quantization rates become low, i.e., when the bandwidth expansion and mean CSNR are low. However, at low bit-rates, the gain achieved by coding of a source with memory diminishes as well, and since the quantization error is no longer small compared to the source variance, a large fraction of the transmitter power gets allocated to the analog transmission. In this regime, neither HDA-PQ nor HDA-TC is worth the effort since similar performance can be achieved by the simple purely analog system described in Section 5.1. This observation also shows that the use of asymptotic quantizer expressions to determine the optimal power allocation is indeed reasonable.

Table 1 lists the  $\rho$  and  $\theta_o$  values of HDA-PQ and HDA-TC systems whose AMMSEs are shown in Fig. 3. In general, the power allocated to the analog component of both HDA-PQ and HDA-TC increases with average CSNR  $\bar{\theta}$ , but decreases with the increasing bandwidth. The former effect is due to the fact that, when  $\bar{\theta}$  of a Rayleigh fading channel increases, so does the variance of the CSNR. The latter effect can be explained as follows.

When more channel bandwidth is made available, the AMMSE can be reduced by increasing the quantization rate and hence the prediction gain.

We have used the AMMSE as a design criterion to achieve a good (asymptotically optimal) trade-off between the digital coding gain and the analog robustness over a wide range of CSNRs. This design procedure determines the best power allocation factor  $\rho$  and the outage CSNR

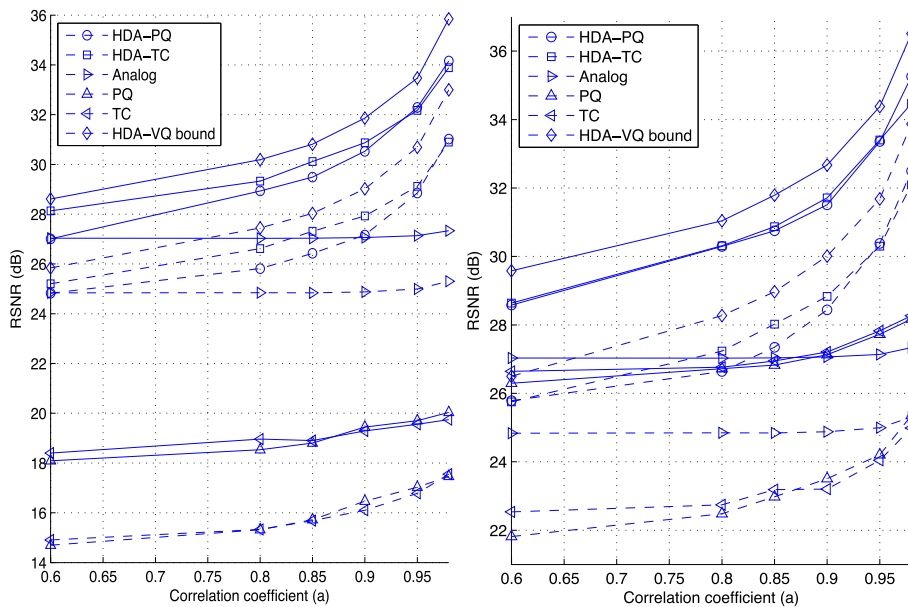


**Fig. 4** Experimental RSNR of asymptotically optimal designs as a function of RX-CSNR for GM(0.9) source. HDA-PQ prediction order is 1, and HDA-TQ transform block size is 8. **a**  $\bar{\theta} = 15$  dB, left:  $b = 3$ , right:  $b = 5$ . **b**  $\bar{\theta} = 25$  dB, left:  $b = 3$ , right:  $b = 5$

$\theta_o$  (or equivalently the quantization rate) to be used over a given fading channel ( $\bar{\theta}$ ). However, given that the channel has slow fading, an important performance measure from the point of view of the individual users in the system is the MSE  $D(\theta)$  of a receiver with a given instantaneous CSNR  $\theta$  (which we refer to as RX-CSNR in figures). Consider a receiver operating in an AMMSE-optimized system, whose CSNR is  $\theta$ . Figure 4 shows several examples for receiver *reconstruction signal-to-ratio* (RSNR)  $10 \log_{10} \frac{\sigma_x^2}{D(\theta)}$  as a function of the RX-CSNR, where the top two figures are for a channel with  $\bar{\theta} = 15$  dB and the bottom ones are for a channel with  $\bar{\theta} = 25$  dB. The figures also show the performance of purely digital systems and the analog system in Section 5.1 with a decoding block size of 8. Other than the HDA-VQ bound, these results are experimentally measured performance of actual systems designed with asymptotically optimal power allocations and outage CSNRs. The purely digital systems have been designed with the same procedure as the HDA systems, but by setting  $\rho = 0$  and optimizing only with respect to  $\theta_o$ . To have a perspective of CSNR variations, these figures show the values of CSNR above which each channel remains 90 and 99% of the time, respectively (CSNR<sub>90%</sub> and CSNR<sub>99%</sub>). The effect of designing HDA systems to minimize the AMMSE can be clearly seen. Unlike the digital systems, the performance of HDA systems increases limitlessly while having lower outage probabilities. HDA systems outperform the digital-only counterparts

85 – 90% of the time in all cases. Increasing the bandwidth expansion on a given channel (hence increasing the capacity) not only boosts the instantaneous RSNR at all RX-CSNRs above the outage value but also reduces the outage CSNR. The gap between the HDA systems and the analog system is due to the source-coding gain of HDA systems (of course, for memoryless Gaussian sources and unit bandwidth expansion, purely analog transmission is optimal [7]).

In order to highlight the fact that HDA-PQ and HDA-TC proposed in this paper are useful only with correlated sources, Fig. 5 presents the dependence of RSNR on the correlation coefficient  $a$  at 20 dB RX-CSNR. Note that as  $a$ , and hence the source-coding gain drops, the performance of both HDA-PQ and HDA-TC approach that of the analog system. On the other hand, for high  $a$ , the HDA systems substantially improve over the analog system. Tables 2 and 3 present the  $\rho$  and  $\theta_o$  values of HDA-PQ and HDA-TC systems shown in Fig. 5. As the source correlation increases, the optimal solution allocates more power to the digital transmission to benefit from the resulting source-coding gain. In interpreting these results, note also that that higher  $\bar{\theta}$  and  $b$  means higher overall channel capacity. Therefore, the higher the channel capacity, the higher the gap between HDA systems and the analog system. Note that the analog power allocation of HDA-TC does not monotonically decrease with increasing  $a$  (see Tables 2 and 3). This is because, as the source correlation increases, the



**Fig. 5** Experimental RSNR of HDA systems for GM( $a$ ) source, as a function of the correlation coefficient  $a$  at RX-CSNR = 20 dB. Left:  $\bar{\theta} = 15$  dB, right:  $\bar{\theta} = 25$  dB. Dash-lines:  $b = 3$ , solid-lines:  $b = 5$ .  $\rho$  and  $\theta_o$  values for HDA-PQ and HDA-TC systems shown here are listed in Tables 2 and 3. HDA-PQ prediction order is 1, and HDA-TQ transform block size is 8

**Table 2** The power allocations and outage CSNRs of HDA-PQ and HDA-TC systems shown in Fig. 5 (left)

$a$	$b$	HDA-PQ		HDA-TC	
		$\rho^*$ (%)	$\theta_o^*$ (dB)	$\rho^*$ (%)	$\theta_o^*$ (dB)
0.60	3	100	–	61	–3.3
0.80	3	30	–1.2	48.5	–3.4
0.85	3	28	–1.4	44.5	–3.5
0.90	3	27	–1.9	39.5	–3.8
0.95	3	25.5	–2.7	45	–5.1
0.98	3	23.5	–3.7	33	–5.5
0.60	5	100	–	62.5	–5.2
0.80	5	28.5	–3.1	51	–5.4
0.85	5	28.5	–3.4	47	–5.6
0.90	5	28	–3.9	42	–5.8
0.95	5	27.5	–4.8	35.5	–6.3
0.98	5	25.5	–5.9	36	–7.6

The average CSNR is 15 dB

number coefficients with non-zero bit allocations shrink. Therefore optimal  $\rho$  is not a continuous function of  $a$ . Note also that there is a sharp decrease in analog power allocation in HDA-PQ when the source correlation coefficient  $a$  changes from 0.6 to 0.8. For  $a = 0.6$ , the prediction gain achievable is too small for digital coding to be useful. For  $a = 0.8$ , the prediction gain is significant. It was observed that when  $\rho$  exceeds about 30%, the low quantization rates result in poor predictions through the feedback loop, making predictive coding ineffective. Hence, the sharp increase of analog power allocation from 28 – 30% to 100% is seen in Table 2.

**Table 3** The power allocations and outage CSNRs of HDA-PQ and HDA-TC systems shown in Fig. 5 (right)

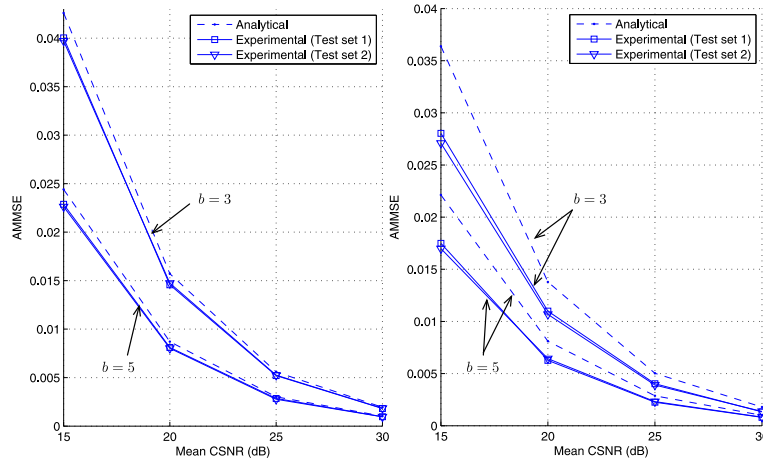
$a$	$b$	HDA-PQ		HDA-TC	
		$\rho^*$ (%)	$\theta_o^*$ (dB)	$\rho^*$ (%)	$\theta_o^*$ (dB)
0.6	3	29.5	0.7	63	–2.2
0.8	3	28	0.1	53	–2.3
0.85	3	27.5	–0.3	48	–2.5
0.9	3	27.5	–0.8	43	–2.8
0.95	3	27.5	–1.7	36.5	–3.3
0.98	3	26.5	–2.9	38.5	–4.6
0.6	5	25.5	–1.2	55	–3.4
0.8	5	27	–2.0	45.5	–3.8
0.85	5	27.5	–2.4	50.5	–4.7
0.9	5	28	–2.9	45.5	–5.0
0.95	5	28.5	–3.9	39	–5.5
0.98	5	28	–5.1	41	–6.9

The average CSNR is 25 dB

## 6.2 HDA speech transmission

One of the key applications of predictive coding is in moderate-to-high bit-rate speech coding [13]. We designed and simulated HDA-PQ, and for comparison HDA-TC systems, for 4 kHz speech signals sampled at 8 kHz. It is known that speech can be well modeled by a 10th-order auto-regressive process [31]. Therefore, a 10th-order linear predictor was used in predictive coding, while a transform block size of 10 was used for transform coding. In the latter case, the discrete cosine transform (DCT) [8], which is a more practical choice than the KLT for non-Gaussian vectors, was used. The designs were then carried out using a source covariance matrix estimated from an actual training set of  $4 \times 10^5$  speech samples. This training set consisted of short sentences spoken by a number of male and female English speakers. As in the case of GM( $a$ ) source, the quantization rate (entropy) found by the asymptotic analysis for Gaussian sources were used to design the actual ECQs for HDA-PQ and HDA-TC. For experimentally evaluating the performance of the practical designs, two different test sets (*test set 1* and *test set 2*), each of  $4 \times 10^5$  samples, were used. The test set 1 included male and female English speakers, while the test set 2 included male and female French speakers.

Figure 6 compares the experimental AMMSEs for both test sets and the analytical values which are based on the source covariance matrix estimated from the training set (training and test sets have been normalized to unit-variance). Table 4 lists the power allocations and outage CSNRs of HDA-PQ and HDA-TQ designs shown in Fig. 6. HDA-TCs shown here use a transform block size of 10 (the same as predictor order in HDA-PQ). In all cases shown here, the AMMSEs for both test sets are nearly identical. However, while for HDA-PQ, there is a close agreement between analytical and experimental values, this is not so with HDA-TC. The actual HDA-TC systems perform noticeably better at low bit-rates (low mean CSNR) than the high-rate analysis predicts. This is in contrast to the performance of HDA-TC for Gauss-Markov processes where there is no model mismatch. In the case of speech signals, the asymptotic analysis assumes a stationary Gaussian process while the speech signals are in reality neither Gaussian nor stationary. In this case, Gaussian high-rate analysis of HDA-TC considerably exaggerates the average quantization distortion at the lower rates. Nonetheless, both HDA-PQ and HDA-TC designs perform well, with HDA-TC being slightly better. The main advantage of HDA-PQ is the simplicity of both the design and the implementation. Furthermore, it may be possible to improve HDA-PQ performance at low rates, since in that case, the quantization errors in an HDA-PQ system contain some residual correlation. One idea is to use a linear block decoder in the analog part (it is of course difficult to assume such a decoder during system



**Fig. 6** A comparison of analytical and experimental AMMSEs of HDA systems designed for unit-variance speech signals, HDA-PQ (left) and HDA-TC (right).  $\rho$  and  $\theta_o$  values for HDA-PQ and HDA-TC systems shown here are listed in Table 4. Both HDA-PQ prediction order and HDA-TQ transform block size are 10

optimization). Another possibility is to use a decoder of the form [32].

Figure 7 presents the decoder performance as a function of RX-CSNR in systems with  $b = 3$  (12 kHz channel bandwidth) and  $b = 5$  (20 kHz channel bandwidth). Both HDA-TC systems and the decoder in analog-only systems had their block size set to 10. In this case, the performance of decoded speech has been evaluated by using the short-term or *segmental SNR* (seg-SNR) given by ([33], Eq. 9.7) which is known to better reflect the perceptual quality of speech at moderate to high-bit-rates than the RSNR. In our experiments, we used a segment size of 240 samples, which corresponds to 30 ms. Despite the fact that speech is non-Gaussian, the results in Fig. 7 are qualitatively consistent with those in Fig. 4. For example, HDA system outperform analog system at all RX-CSNRs and digital systems 85–90% of the time. We also performed listening tests which supported the trends in Fig. 7. The

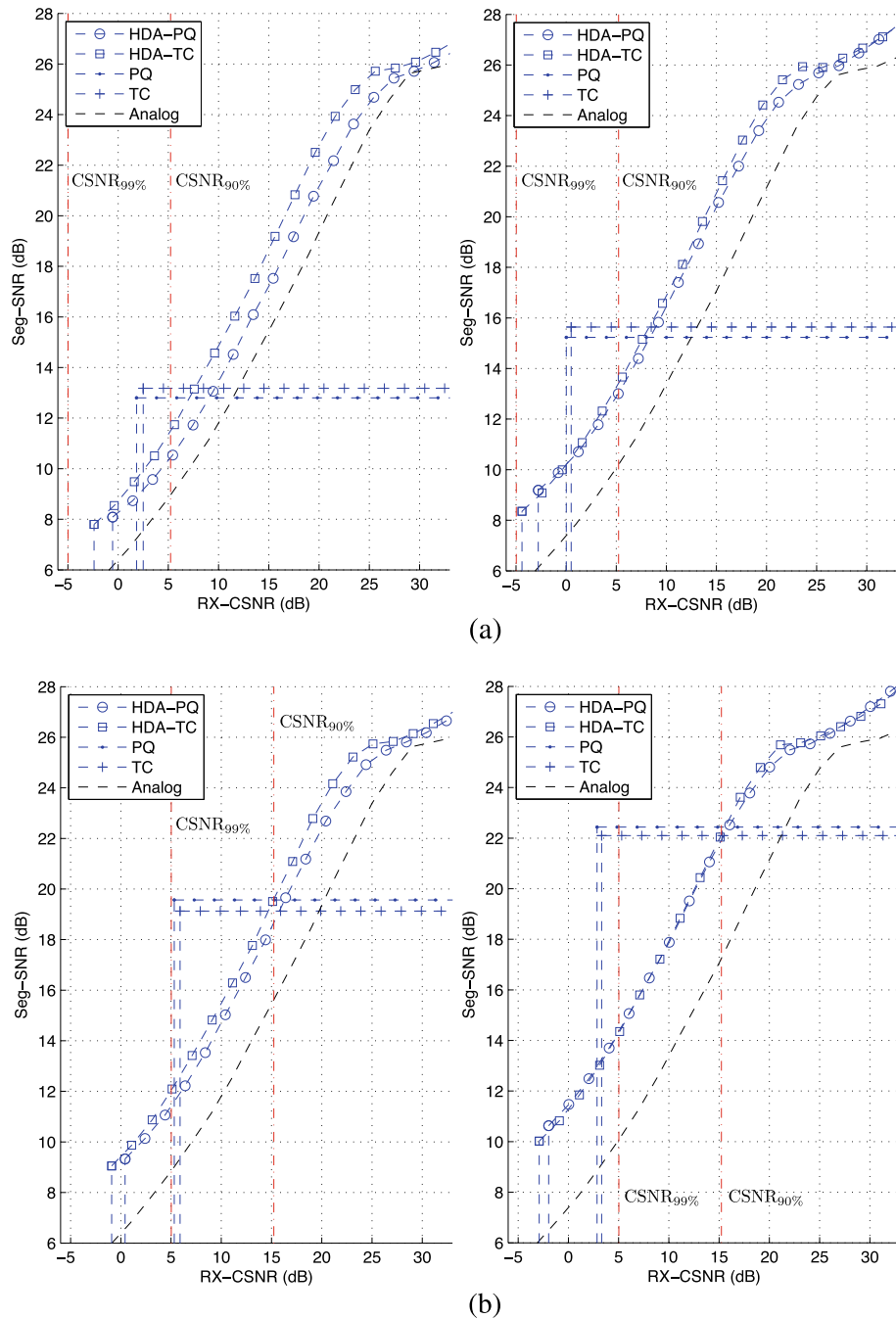
decoded speech from the HDA systems sounded the best, though when the available channel capacity is relatively low (e.g., when  $\bar{\theta} = 15$  dB and  $b = 3$ , see Fig. 7a), HDA-TC sounded less noisier than HDA-PQ. Both HDA-PQ and HDA-TC systems produced speech with white background noise (but free from any quantization noise) that dropped rapidly as the RX-CSNR is increased. Indeed, this “graceful” variation of the output quality is the goal we hoped to achieve with the proposed HDA designs. In comparison, the analog-only systems produced noticeably more background noise. The digital-only systems produced comparatively poor quality speech, with objectionably harsh quantization noise being clearly audible, except when the available channel capacity is relatively high (e.g., when  $\bar{\theta} = 25$  dB and  $b = 5$ , see Fig. 7b).

## 7 Conclusions

This paper presented an approach to designing HDA-PQ and HDA-TC systems for transmitting correlated Gaussian sources over frequency-flat, block Rayleigh fading channels, when CSI is not available to the transmitter. In this case, the encoder is designed to minimize the AMMSE over the receiver-CSNR distribution, so that the system operates well over a range of CSNRs. The main issue addressed in this paper is the joint optimization of the analog-digital power allocation and the outage CSNR (or equivalently the quantization rate) to minimize the AMMSE of HDA-PQ and HDA-TC systems. In particular, a simple algorithm for solving the optimization problem in the case of HDA-PQ was presented. While the power allocations and quantization rates obtained as suggested in this paper can only be asymptotically (in rate) optimal, they were found to be effective in actual HDA systems with finite rates.

**Table 4** The power allocations and outage CSNRs of HDA-PQ and HDA-TC systems shown in Fig. 6

$\bar{\theta}$ (dB)	$b$	HDA-PQ		HDA-TC	
		$\rho^*$ (%)	$\theta_o^*$ (dB)	$\rho^*$ (%)	$\theta_o^*$ (dB)
15	3	20	−0.7	38	−2.2
20	3	21.5	−0.1	39	−1.7
25	3	22	0.3	40	−1.0
30	3	22	0.7	33	−0.2
15	5	22	−2.9	39	−4.2
20	5	23	−2.4	34	−3.2
25	5	23	−2.0	35	−3.0
30	5	23	−1.7	35	−2.5



**Fig. 7** Experimental seg-SNR of speech test set 2 on channels with different mean CSNRs. Both HDA-PQ prediction order and HDA-TQ transform block size are 10. **a**  $\bar{\theta} = 15$ , left:  $b = 3$ , right:  $b = 5$ . **b**  $\bar{\theta} = 25$ , left:  $b = 3$ , right:  $b = 5$

Our experimental results showed that, despite the Gaussian assumption, the proposed HDA design approach also worked well with the speech signals. HDA-PQ in particular can be a good approach to adaptive speech coding (e.g., similar to ADPCM [13]) over fading channels and in broadcasting. HDA-PQ is amenable to adaptive quantization in real time due to the simplicity of the system

optimization algorithm presented in Section 4.1. A simple approach to adaptive speech coding with HDA-PQ is to use a finite-state model for the source signal, where the state is determined by a segment of consecutive speech samples and each state has a particular set of HDA-PQ parameters (predictor coefficients, quantizer-rate, and power allocation). The method described in this paper can



be used to determine optimal parameters for each state. Since unvoiced speech segments resemble white noise, experimental results in this paper suggest that purely analog transmission can likely be as nearly as good as HDA-PQ for such segments. On the other hand, for highly correlated voiced speech segments, a significant amount of total power will get allocated to the digital component.

## Endnote

<sup>1</sup> Since the elements of  $\mathbf{v}$  are linear combinations of  $M$  quantization errors, they will be approximately Gaussian if  $M$  is sufficiently large.

## Competing interests

The author declares that he has no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 22 November 2016 Accepted: 9 May 2017

Published online: 25 May 2017

## References

- WF Schreiber, Advanced television systems for terrestrial broadcasting: some problems and some proposed solutions. *Proc. IEEE*. **83**(6), 958–981 (1993)
- N Phamdo, U Mittal, A joint source-channel speech coder using hybrid digital-analog (HDA) modulation. *IEEE Trans. Speech Audio Process.* **10**(4), 222–231 (2002)
- TM Cover, Broadcast channels. *IEEE Trans. Inform. Theory*. **18**(1), 2–14 (1972)
- K Ramachandran, A Ortega, KM Uz, M Vetterli, Multiresolution broadcast for digital HDTV using joint source/channel coding. *IEEE J. Select. Areas Commun.* **11**(1), 6–23 (1993)
- I Kozintsev, K Ramachandran, in *IEEE ICIP*. Hybrid compressed-uncompressed framework for wireless image transmission (IEEE, Santa Barbara, 1997), pp. 77–80
- U Mittal, N Phamdo, Hybrid digital-analog (HDA) joint source-channel codes for broadcasting and robust communication. *IEEE Trans. Inform. Theory*. **48**(5), 1082–1102 (2002)
- TJ Goblick, Theoretical limitations on the transmission of data from analog sources. *IEEE Trans. Inform. Theory*. **11**(4), 558–567 (1965)
- A Gersho, RM Gray, *Vector Quantization and Signal Compression*. (Kluwer Academic Publishers, Norwell, 1992)
- L Yu, H Li, W Li, Wireless scalable video coding using hybrid digital-analog scheme. *IEEE Trans. Circ. Syst. Video Technol.* **24**(2), 331–345 (2014)
- Y Wang, F Alajaji, T Linder, in *IEEE Data Compression Conf.* Design of VQ-based hybrid digital-analog joint source-channel codes for image communications (IEEE, Snow Bird, 2005), pp. 193–202
- M Skoglund, N Phamdo, F Alajaji, Design and performance of VQ-based hybrid digital-analog joint source-channel codes. *IEEE Trans. Inform. Theory*. **48**(3), 708–720 (2002)
- M Skoglund, N Phamdo, F Alajaji, Hybrid digital-analog source-channel coding for bandwidth compression/expansion. *IEEE Trans. Inform. Theory*. **52**(8), 3757–3763 (2006)
- M Hasegawa-Johnson, A Alwan, in *Wiley Encyclopedia of Telecomm.* Speech coding: Fundamentals and applications (John Wiley, USA, 2003)
- H Gish, JP Peirce, Asymptotically efficient quantizing. *IEEE Trans. Inform. Theory*. **IT-14**, 676–683 (1968)
- DP Bertsekas, *Nonlinear Programming*, 2nd edn. (Athena Scientific, Belmont, 1999)
- M Rungeler, P Vary, in *IEEE ICASSP*. Hybrid digital analog transform coding (IEEE, Vancouver, 2013), pp. 1–5
- Z Song, R Xiong, S Ma, X Fan, W Gao, in *IEEE ICME*. Layered image/video softcast with hybrid digital-analog transmission for robust wireless visual communication (IEEE, Chengdu, 2014), pp. 1–6
- H Cui, C Luo, CW Chen, F Wu, in *IEEE INFOCOM 2014*. Robust uncoded video transmission over wireless fast fading channels (IEEE, Adelaide, 2014), pp. 73–81
- VK Goyal, J Kovacevic, JA Kelner, Quantized frame expansions with erasures. *Appl. Comput. Harmon. Anal.* **10**, 203–233 (2001)
- RG Gallager, *Information Theory and Reliable Communication*. (John Wiley, USA, 1968)
- TM Cover, JA Thomas, *Elements of Information Theory*, 2nd edn. (John Wiley, Upper Saddle River, 2006)
- SM Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. (Prentice Hall, Hoboken, 1993)
- DS Arnstein, Quantization error in predictive coders. *IEEE Trans. Commun.* **23**(4), 423–429 (1975)
- N Farvardin, JW Modestino, Rate-distortion performance of DPCM schemes for autoregressive sources. *IEEE Trans. Inform. Theory*. **IT-31**(3), 402–418 (1985)
- VR Algazi, JT DeWitte, Theoretical performance of entropy-encoded DPCM. *IEEE Trans. Commun.* **30**(5), 1088–1095 (1982)
- S Sesia, G Caire, G Vivier, in *Proc. IEEE Int. Symp. Inform. Theory*. Lossy transmission over slow-fading AWGN channels: a comparison of progressive, superposition and hybrid approaches, (2005), pp. 224–248
- GL Stuber, *Principles of Mobile Communications*, 2nd edn. (Kluwer Academic Publishers, New York, 2002)
- M Geller, EW Ng, A table of integrals of the exponential integral. *J. Res. Natl. Bureau of Standards-B Math. Math. Sci.* **73B**(3), 191–210 (1969)
- T Berger, *Rate Distortion Theory: a Mathematical Basis for Data Compression*. (Prentice-Hall, Englewood Cliffs, 1971)
- P Chou, T Lookabaugh, RM Gray, Entropy-constrained vector quantization. *IEEE Trans. Acoust. Speech Signal Process.* **37**(1), 31–42 (1989)
- H Abut, RM Gray, G Rebolledo, Vector quantization of speech and speech-like waveforms. *IEEE Trans. Acoust. Speech Signal Process.* **30**(3), 423–435 (1982)
- K Sayood, JC Borkenhagen, Use of residual redundancy in the design of joint source/channel coders. *IEEE Trans. Commun.* **39**, 838–846 (1991)
- JR Deller, JHL Hansen, JG Proakis, *Discrete Time Processing of Speech Signals*. (Wiley-InterScience, New Jersey, 1993)

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)