Impact of Channel Prediction on Adaptive Coded Modulation Performance in Rayleigh Fading

Geir Egil Øien, *Member, IEEE,* Henrik Holm, *Member, IEEE,* and Kjell Jørgen Hole, *Member, IEEE*

Abstract

Adaptive coded modulation (ACM) is a promising tool for increasing the spectral efficiency on time-varying mobile channels while keeping a predictable bit error rate. An important restriction in systems with such a transmission scheme is that the transmitter needs to have accurate channel state information (CSI). Earlier analysis of ACM systems usually assume that the transmitter has perfect knowledge of the channel, or that the CSI is accurate but outdated. In this paper, we investigate the effects of predicting the CSI using a linear fading envelope predictor, in order to enhance the performance of an ACM system. For the case when multidimensional trellis codes are used on Rayleigh fading channels, we obtain approximative closed-form expressions for bit error rate and average spectral efficiency. Numerical examples are given for the case of Jakes correlation profile and maximum a posteriori-optimal predictor coefficients.

This work was supported in part by the Research Council of Norway (NFR), under contract numbers 140418/431 and 146946/431 (http://www.tele.ntnu.no/projects/beats), and in part by the NTNU project IKT-webtek (http://www.ntnu.no/satsingsomraader/ikt/focus1.htm).

G. E. Øien is with the Department of Electronics and Telecommunications, Norwegian University of Science and Technology, O. S. Bragstads plass 2B, N-7491 Trondheim, Norway. (e-mail: geir.oien@tele.ntnu.no).

H. Holm is with the Department of Electronics and Telecommunications, Norwegian University of Science and Technology. He is currently visiting the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA. (e-mail: henrik@ece.umn.edu).

K. J. Hole is with the Department of Telecommunications, Norwegian University of Science and Technology, and the Department of Informatics, University of Bergen, HiB, N-5020 Bergen, Norway (e-mail: kjell.hole@ii.uib.no).

Index Terms

(1) Fading Channels, (2) Adaptive Modulation, (3) Channel Estimation, (4) Antenna Diversity

I. Introduction

Wireless radio communication is frequently perturbed by multipath fading. A flexible model for such fading is *Nakagami multipath fading* (NMF) [1]. The much-used *Rayleigh* fading model, which will be assumed throughout this paper, is a special case of NMF. One way of coping with the varying channel quality resulting from NMF in general and Rayleigh fading in particular is by employing adaptive coded modulation (ACM) [2]–[5]. An important restriction in ACM systems is that the transmitter needs to have accurate channel state information (CSI). Earlier analysis of ACM systems usually assume that the transmitter has perfect knowledge of the channel, or that the CSI is accurate but outdated. In a practical ACM scheme, the CSI will have to be estimated at the receiver and then fed back to the transmitter. The estimation will not be perfect, and there will be some nonzero and possibly time-varying delay in the feedback channel. In this paper, we obtain closed-form expressions describing how estimation error and feedback delay affect some key performance measures of an ACM system.[1]

To improve the reliability of signal detection, pilot symbols can be inserted regularly into the data stream. This technique is called pilot symbol assisted modulation (PSAM) [6]–[8]. PSAM is also well suited for channel estimation in an ACM system. The requirements of the estimation technique are then slightly different from the signal detection case. Specifically, a channel *predictor* should be implemented by estimating the channel signal-to-noise ratio (CSNR) at the time in the future when it will be used. This way, the system takes into account the feedback delay.

A system as shown in Fig. 1 (baseband model) is assumed from now on. The figure also covers the case where the receiver has more than one receive antenna, enabling a potential *diversity combining gain* [9, Ch. 6]

When performing discrete rate adaptation by switching between the N available transmitter–receiver pairs, the transmitter must rely on the accuracy of CSI periodically fed back from

[1]Note that the “feedback delay” does not consist of the actual transmission delay only; it will also incorporate the time it takes to perform the estimation and the processing time needed by the transmitter to activate the transmission mode that should be employed. For simplicity, the sum of these delays is referred to as feedback delay.

the receiver end. The CSI in practice provides information about the CSNR *as predicted at the receiver* at the time t of signal reception. The *true* channel quality at the transmitter update time $t+\tau$, where τ is the return channel delay, may therefore deviate from the value available to the transmitter. Thus the transmitter may make the wrong assumption about the channel quality and transmit at a too high or too low rate. This may lead to changes in average bit error rate (BER) and *average spectral efficiency* (ASE) compared to the case of perfect CSI. Exactly what kind of changes that will occur depends on several factors, the most important being the fading correlation, the feedback delay τ, the average channel quality, the number of codes/signal constellations used (i.e., the degree of adaptivity), and how the CSNR is predicted. In this paper we will analyze how these factors influence the BER and the ASE.

The rest of the paper is organized as follows: In Section II, an in-depth description of the system model is given. Section III discusses linear pilot symbol assisted channel prediction. Section IV analyzes some key features of the system (ASE and BER) for an arbitrary linear prediction algorithm. In Section V, the predictor which is optimal in the *maximum a posteriori* (MAP) sense is introduced, and properties that are necessary to perform the analysis of the system are derived. Section VI concerns experiments on a certain example set of codes. Finally, the contributions of the paper are summarized in Section VII.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Denoting the transmitted complex baseband signal (after pilot symbol insertion) at time index k by $x(k)$, the received signal after transmission on a flat-fading channel can be written as $y(k) = z(k) \cdot x(k) + n(k)$. For a system with antenna diversity like the one suggested in Fig. 1, the received signal on the hth subchannel is denoted as $y_h(k) = z_h(k) \cdot x(k) + n_h(k)$. Here, the random variable (RV) $z_h(k)$ is the *complex fading amplitude*, and $n_h(k)$ is complex-valued additive white Gaussian noise (AWGN) with statistically independent real and imaginary components. $x(k)$ is the information signal, except for time instants $k = pL$ ($p \in \mathbb{Z}$, $L \in \mathbb{Z}^+$), when the *pilot symbols* are transmitted. The value of L has to be determined as a trade-off between the requirement for adequate sampling of the fading process, calling for a small L, and the requirement for high spectral efficiency, calling for a large L. We assume that all pilot symbols have the same absolute value, $|x(pL)| = a_p$. The pilot symbols will in practice often be modulated according to a pseudo-random sequence, in order to avoid spectral peaks (see for instance [10, Sec. 4.5]).

To be able to represent the results in closed-form, it is assumed that each subchannel is perturbed by flat Rayleigh fading, which is a special case of NMF with Nakagami parameter $m = 1$ [9, p. 53]. Then, $z_h(k)$ is a complex-valued Gaussian variable with zero mean. Furthermore, each subchannel is assumed to be wide-sense stationary (WSS). The fading power $\Omega = E[|z_h|^2]$ is therefore time-invariant and the autocorrelation of z_h measured at two different time indices k_1 and k_2 is only dependent on the time difference. Note that although the well-known *Jakes spectrum* will be used in our numerical results later in the paper, our analytical framework will not place any constraints on the actual shape of the autocorrelation sequence and corresponding Doppler spectrum.

Furthermore, a constant average transmit power P [W] and a one-sided power spectral density N_0 [W/Hz] of the complex AWGN in every subchannel are assumed. For a one-sided information bandwidth B [Hz] the received instantaneous CSNR on subchannel h at a given time k is then

$$\gamma_h(k) = \frac{|z_h(k)|^2 \cdot P}{N_0 B}, \tag{1}$$

with expected value $E[\gamma_h] = \bar{\gamma}_h = \Omega P/(N_0 B)$.

In addition to the time-invariance assumptions, it is also assumed that the subchannels are statistically independent and identically distributed. The fading power Ω is therefore independent of h, as is the expectation $\bar{\gamma}_h$. However, the latter is indexed by h anyway to indicate that this is not the *overall* expected CSNR on the channel.

Under the assumption of H statistically independent antenna branches, maximal ratio combining (MRC) can be implemented in the receiver. The overall received CSNR at time k will then be [9, Sec. 6.3]:

$$\gamma(k) = \sum_{h=1}^{H} \gamma_h(k). \tag{2}$$

On a Rayleigh fading channel, $\gamma_h(k)$ is exponentially distributed. It can be shown [11] that $\gamma(k)$ after MRC is gamma-distributed with a shape parameter H and expectation $H\bar{\gamma}_h$. In practice, this means that the overall channel with MRC behaves like an NMF channel with Nakagami parameter $m = H$ and fading power $H\Omega$.

In an ACM system like the one depicted in Fig. 1, the adaptive coder/modulator has available a set of N transmitter–receiver pairs, denoted as transmitter–receiver pair $n = 1, \ldots, N$. Transmitter n has a rate of R_n information bits per symbol, such that $R_1 < R_2 < \cdots < R_N$. The

CSNR range is split into $N+1$ *CSNR intervals*, as depicted in Fig. 2, and transmitter–receiver pair n is to be used when CSNR γ falls in the interval $[\gamma_n, \gamma_{n+1}\rangle$. In most ACM references, γ_n is—for each $n = 1, \ldots, N$—chosen as the lowest CSNR necessary for transmitter–receiver pair n to be able to operate at a BER below the designer-specified target bit error rate, BER_0, at transmit power P on an AWGN channel. However, we stress that the theoretical analysis in this paper does not depend on the CSNR intervals being chosen according to this criterion. In our numerical examples, however, this is assumed.

Letting $\gamma_0 = 0$ and $\gamma_{N+1} = \infty$ will result in $\gamma_{n+1} > \gamma_n$ for all $n \in \{0, \ldots, N\}$. No available transmitter–receiver pair satisfies the BER requirement when the CSNR is lower than interval γ_1; hence, no information is transmitted when γ falls in the interval $[0, \gamma_1\rangle$, and there will consequently be an outage during which information must be buffered at the transmitter end. For a more thorough introduction to how the ACM system operates, see for instance [5], [12].

The transmitter needs CSI in order to select the best code, and this is accomplished by letting the transmitter perform periodical prediction of the CSNR, at a rate equal to (at least) the highest possible transmitter adaptation rate. Here we will assume that the CSNR is predicted and transmitted back to the transmitter every time a new pilot symbol arrives. The return channel delay (here understood as the time delay between the time of CSNR prediction and the closest subsequent time of allowed transmitter update) is assumed to be an integer number of pilot symbol intervals, i.e., $\tau = kLT_\text{s}$, where $k \in \mathbb{Z}^+$ is the number of pilot symbol intervals and T_s [s] is the duration of one channel symbol. Finally, the return channel is assumed to be free of errors.

As a final comment we note that channel *prediction* is only necessary for the CSI which is fed back to the transmitter. For the actual signal detection in the receiver, the signal may be buffered before detection (assuming that a certain receiver delay is acceptable, such as in data transmission)—cf. Fig. 1. CSI estimation can then be accomplished using an optimal non-causal Wiener interpolator filter [6]–[8], [13], which will smooth the noise and improve CSI reliability beyond what can be achieved by a predictor, and allow for true coherent detection to be used. We therefore assume perfect coherent detection in the receiver.

III. LINEAR PILOT SYMBOL ASSISTED CHANNEL PREDICTION

For any pilot symbol time instant $i - kL$ ($i = pL$, $p \in \mathbb{Z}$; $k \in \{0, 1, 2 \ldots\}$), divide the noisy signal received on each subchannel by the known pilot symbol value. The result can be interpreted as a memoryless maximum-likelihood (ML) estimate of $z_h(i - kL)$ based on one received observation [14]:

$$\tilde{z}_h(i - kL) = z_h(i - kL) + \frac{n_h(i - kL)}{a_p}. \tag{3}$$

The two terms are statistically independent and are also both zero-mean complex Gaussian, so their sum is a complex Gaussian with variance equal to the sum of their variances. Let us assume that a transmit update is desired j channel symbols ahead in time of the last received pilot symbol, i.e., at time $i+j$. This means that we should predict the channel fading at this time instant. Throughout the paper we will limit j to be an integer number of pilot symbol periods, i.e., $j = qL$ for $q \in \{0, 1, 2, \ldots\}$, i.e. we predict the channel only in pilot symbol instants. This is not a restrictive assumption, since a pilot symbol should be transmitted at least every 10th to 20th symbol (this will be demonstrated in Section V-B).

A prediction of $z(i + j)$ can be made from the K memoryless ML estimates given by (3). Here, $K \in \mathbb{Z}^+$ is a designer-chosen constant. Since we are dealing with complex Gaussian processes, the optimal fading envelope predictor in the MAP (and ML) sense is known to be a *linear* function of these observations [15]. We therefore restrict our study to linear prediction from now on.

Allowing for complex predictor filter coefficients, *any* linear predictor of order K can be written on the form

$$\hat{z}_h(i + j) = \sum_{k=0}^{K-1} f_{j,h}^*(k) \cdot \tilde{z}_h(i - kL) = \mathbf{f}_{j,h}^{\mathsf{H}} \tilde{\mathbf{z}}_{h,i}, \tag{4}$$

where $\mathbf{f}_{j,h}^{\mathsf{H}} = [f_{j,h}^*(0), \ldots, f_{j,h}^*(K-1)]$ is the predictor filter coefficient vector corresponding to subchannel h and delay j, and where

$$\tilde{\mathbf{z}}_{h,i} = [\tilde{z}_h(i), \tilde{z}_h(i - L), \ldots, \tilde{z}_h(i - (K-1)L)]^{\mathsf{T}} \tag{5}$$

is the vector of the K memoryless ML estimates of the complex fading amplitude, in the K last pilot symbol instants.

The optimal filter coefficient vector shall be stated in Section V; for the moment, we just assume that the coefficient vector is known. Due to the WSS assumption, the time index i will from now on be omitted when referring to $\tilde{\mathbf{z}}_{h,i}$ and other vectors where it is applicable.

It can be shown [8], [11] that the predicted fading envelope is described by the Rayleigh distribution with power $E[\hat{\alpha}_h^2] = \widehat{\Omega} = r\Omega$ for some constant r which depends on the predictor coefficients. Expressions for r will be provided for the general case in the Appendix, and for the MAP-optimal case in Section V.

A. Maximal Ratio Combining

As discussed in Section II, the receiver will implement MRC. The composite channel will then consequently behave like a Nakagami-H channel, and the true CSNR will be gamma-distributed with shape parameter H and expectation $\bar{\gamma} = E[\gamma] = H\bar{\gamma}_h$. Since the receiver implements MRC, a reasonable prediction of the squared overall channel gain is as follows:

$$\hat{\alpha}^2 = \sum_{h=1}^{H} \hat{\alpha}_h^2. \tag{6}$$

Again, as was the case for the actual effective channel gain, we have a sum of squared Gaussian RVs. The effective predicted channel gain hence obeys a Nakagami-H distribution with $E[\hat{\alpha}^2] = H\widehat{\Omega} = rH\Omega$. The predicted overall CSNR $\hat{\gamma} = \frac{\hat{\alpha}^2 P}{N_0 B}$ is then gamma-distributed with shape parameter H and expectation

$$E[\hat{\gamma}] = \bar{\hat{\gamma}} = H\bar{\hat{\gamma}}_h = rH\bar{\gamma}_h. \tag{7}$$

Hence, the true and the expected CSNR are both gamma-distributed. The parameter r can be viewed both as the ratio between the expectation of the predicted and the true subchannel CSNR and as the ratio between the squared mean (variance) of the predicted and the true complex fading amplitude. It follows from the definition that the ratio r is independent of the number of receive antennas.

B. Correlation Coefficient

An important factor affecting the error performance in an ACM system is the correlation between the predicted and the true CSNR. The BER expressions which will be developed in the next section will be functions of the normalized correlation coefficient ρ between the predicted and the true CSNR, i.e.,

$$\rho = \frac{\mathrm{Cov}(\hat{\gamma}, \gamma)}{\sqrt{\mathrm{Var}(\hat{\gamma})\,\mathrm{Var}(\gamma)}} = \frac{E[\hat{\alpha}^2\alpha^2] - H^2\Omega^2 r}{H\Omega^2 r}. \tag{8}$$

Additionally, it can be shown [11], [16] that the overall correlation coefficient is the same as the correlation coefficient for the channel between the transmitter and any of the H receive antennas. It follows that ρ may be expressed as

$$\rho = \frac{E[\hat{\alpha}_h^2\alpha_h^2] - \Omega^2 r}{\Omega^2 r} = \rho_h. \tag{9}$$

IV. System Analysis

A. BER Analysis

During our analysis, it will be assumed that γ_h stays in the same CSNR interval between two successive pilot symbols. Again, this is not a restrictive assumption, since a pilot symbol should be transmitted at least every 10th to 20th symbol (cf. Section V-B). With parameters suggested in the same Section and assuming that the correlation is described by the *Jakes Spectrum*, the correlation ρ of instantaneous SNR γ with a separation of 10 or 20 symbols will be 0.9995 or 0.998, respectively. The parameters suggested in Section V-B include among others a terminal speed of $v = 30\,$m/s ($108\,$km/h); as such quite conservative. A lower speed will yield higher correlation.

The BER (averaged over all codes and all CSNRs) is given as the average number of bits in error, divided by the average number of bits per symbol transmitted, i.e., the average spectral efficiency (ASE) [3], [16]:

$$\overline{\mathrm{BER}} = \frac{\sum_{n=1}^{N} R_n \cdot \overline{\mathrm{BER}}_n}{\sum_{n=1}^{N} R_n P_n}, \tag{10}$$

where R_n is the information rate of code n, P_n is the probability that code n will be used, and $\overline{\mathrm{BER}}_n$ is the average BER experienced when code n is used.

The probability P_n is simply the probability that the predicted CSNR falls in the interval $[\gamma_n, \gamma_{n+1}\rangle$, and for Rayleigh fading with MRC it can be shown [3] that

$$P_n = Q\Big(H, \frac{\gamma_n}{r\bar{\gamma}_h}\Big) - Q\Big(H, \frac{\gamma_{n+1}}{r\bar{\gamma}_h}\Big) \tag{11}$$

where $Q(x, y)$ is the *normalized incomplete gamma function* [17, Eq. 11.3].

In (10), only strictly positive values of n are used. Note that n can also be zero, corresponding to the CSNR interval $[0, \gamma_1\rangle$. When the CSNR is smaller than γ_1, no information will be sent. However, for applications without strict delay (real-time) constraints, this is of little importance as long as the ASE is maximized (thus minimizing the overall transmit time for large information sets).

The average BER for code n may be written [3]:

$$\overline{\mathrm{BER}}_n = \int_{\gamma_n}^{\gamma_{n+1}} \int_0^{\infty} \mathrm{BER}_n(\gamma \mid \hat{\gamma}) f_{\gamma,\hat{\gamma}}(\gamma, \hat{\gamma})\, d\gamma d\hat{\gamma}, \qquad n = 1, 2, \ldots, N, \quad (12)$$

where $\mathrm{BER}_n(\gamma \mid \hat{\gamma})$ is the BER experienced when applying code n. The choice of n is based on the belief that the CSNR is $\hat{\gamma}$, while it actually is γ. That is, n—and therefore all functions of n—should be viewed as dependent on $\hat{\gamma}$ in the expressions to follow. Furthermore, $f_{\gamma,\hat{\gamma}}(\gamma, \hat{\gamma})$ is the joint distribution of the actual and the predicted CSNR. In our case this will be a *bivariate gamma distribution* [18]. This is because γ and $\hat{\gamma}$ are both individually gamma distributed (cf. Section III-A) and mutually correlated (cf. Section III-B). The reader is referred to Tang *et al.* [8] for more details.

Definition 1 (Bivariate Gamma Distribution): The RVs X_1 and X_2 are described by a bivariate gamma distribution with common shape parameter $\alpha > 0$, scale parameters $\beta_1 > 0$, $\beta_2 > 0$, respectively, and correlation coefficient $\rho = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\,\mathrm{Var}(X_2)}}$ $(0 \leq \rho \leq 1)$ if their joint PDF is given by

$$\begin{aligned} f_{X_1 X_2}(x_1, x_2) = {} & \frac{(x_1 x_2)^{(\alpha-1)/2}}{\Gamma(\alpha)(\beta_1\beta_2)^{(\alpha+1)/2}(1-\rho)\rho^{(\alpha-1)/2}} \\ & \times \exp\Big(-\frac{x_1/\beta_1 + x_2/\beta_2}{1-\rho}\Big) \\ & \times I_{\alpha-1}\Big(\frac{2\sqrt{\rho}}{1-\rho}\sqrt{\frac{x_1 x_2}{\beta_1\beta_2}}\Big) U(x_1) U(x_2). \end{aligned} \quad (13)$$

Here, $\Gamma(\cdot)$ is the *gamma function* (see e.g. [19, Eq. 8.310-1]), $I_\nu(\cdot)$ is the *Modified Bessel function of the first kind and order* ν (see e.g. [19, Eq. 8.445]) and $U(\cdot)$ is the *Heaviside step function* (see e.g. [19, p. xliv]). We use the shorthand notation $X_1, X_2 \sim \mathcal{G}(\alpha, \beta_1, \beta_2, \rho)$ to denote that X_1 and X_2 follow a bivariate gamma distribution. When $\rho \to 0$, the joint PDF reduces to the product of two univariate gamma PDFs (with the aid of [20, Eq. 9.6.7].)

The key to further analysis of the general BER expression (10) is to approximate the BER–CSNR relationship for code n by an analytical expression which will make the integral (12) solvable. Such an expression will be dependent of the type of codes assumed. Hole *et al.* [5], [16] demonstrated that a very good fit to the actual BER–CSNR relationship for *multidimensional trellis codes* on AWGN channels could be found by applying the expression

$$\mathrm{BER}_n(\gamma \mid \hat{\gamma}) = \begin{cases} a_n \cdot \exp\left(-\frac{b_n \gamma}{M_n}\right) & \text{when } \gamma \geq \gamma_n^{\mathrm{l}} \\ \frac{1}{2} & \text{when } \gamma < \gamma_n^{\mathrm{l}} \end{cases} \tag{14}$$

where a_n and b_n are code-dependent constants which may be found by least-squares curve fitting to simulated BER–CSNR data on AWGN channels. M_n is the number of points in the symbol constellation used by the trellis code. The exponential part of the formula in (14) is a very good approximation when the CSNR is high (see e.g. [5] for a theoretical argument for this fact), but is not very good for low CSNRs. Indeed, this BER approximation could become larger than 0.5 for low CSNRs. The BER value 0.5 is therefore introduced along with the boundary $\gamma_n^{\mathrm{l}} = \ln(2a_n)M_n/b_n$, which is the smallest CSNR ensuring that the exponential BER approximation does not exceed 0.5. The resulting curve approximation is depicted in Fig. 3, for the example codes $n = 1, 2, 3$ used in [5]. For clarity of presentation, and since the same trends are seen for all the codes, we have excluded codes $5, \ldots, 8$ in this figure.

It is seen that the exponential approximation in (14) is very accurate for BER levels below 10^{-1}. For BERs between 10^{-1} and 0.5, the approximation tends to produce a somewhat larger value than the true BER, in effect acting as an upper bound. One would perhaps expect that the overall effect would be to make our results somewhat conservative. However, experiments have shown that the overall approximation of the average BER in our ACM context is in fact quite insensitive to this over-estimation of the individual BER curves at values above 10^{-1}. We attribute this to the fact that when the individual BER curves are averaged, the values in this region are weighted by very small probabilities—in other words, it is in practice very unlikely that a code resulting in instantaneous BER above 10^{-1} will be chosen for use in the ACM system. Therefore, it is of little or no consequence to our results that the BER approximation is too large in this region.

Also, one should keep in mind that the simulated BER performance points have been found assuming perfect coherent detection. At low CSNRs this is hard to maintain in practical systems;

thus the simulated BER points are probably too optimistic compared to a real system, and the true BER performance will lie closer to our upper bound for low CSNRs.

In the subsequent analysis, we assume that multidimensional trellis codes are used as component codes in the adaptive coder. This is justified by their good performance in ACM as demonstrated e.g. in [5]. (14) will therefore be employed from now on.

After applying (14) to (12), the result can be expressed as a sum of three integrals:

$$\begin{aligned}\overline{\mathrm{BER}}_n &= \int_{\gamma_n}^{\gamma_{n+1}} \left\{\mathcal{J}1(n,\hat{\gamma}) - (\mathcal{J}21(n,\hat{\gamma}) - \mathcal{J}22(n,\hat{\gamma}))\right\} d\hat{\gamma} \\ &= \mathcal{I}1(n) - (\mathcal{I}21(n) - \mathcal{I}22(n)),\end{aligned} \tag{15}$$

where

$$\mathcal{J}1(n,\hat{\gamma}) = \int_0^{\infty} a_n \exp\left(-\frac{b_n\gamma}{M_n}\right) f_{\gamma,\hat{\gamma}}(\gamma,\hat{\gamma})\, d\gamma \tag{16}$$

$$\mathcal{J}21(n,\hat{\gamma}) = \int_0^{\gamma_n^1} a_n \exp\left(-\frac{b_n\gamma}{M_n}\right) f_{\gamma,\hat{\gamma}}(\gamma,\hat{\gamma})\, d\gamma \tag{17}$$

$$\mathcal{J}22(n,\hat{\gamma}) = \frac{1}{2}\int_0^{\gamma_n^1} f_{\gamma,\hat{\gamma}}(\gamma,\hat{\gamma})\, d\gamma \tag{18}$$

and $\mathcal{I}1(n)$, $\mathcal{I}21(n)$, and $\mathcal{I}22(n)$ are integrals of $\mathcal{J}1(n,\hat{\gamma})$, $\mathcal{J}21(n,\hat{\gamma})$, and $\mathcal{J}22(n,\hat{\gamma})$, respectively. Manipulations as shown in [11] yield closed-form expressions for these three integrals. They can now be expressed as follows:

$$\begin{aligned}\mathcal{I}1(n) = a_n&\left(\frac{1}{\frac{b_n\bar{\gamma}_h}{M_n}+1}\right)^H \\ &\times\left[Q\left(H, \frac{\gamma_n}{\bar{\gamma}_h r}\cdot\frac{\frac{b_n\bar{\gamma}_h}{M_n}+1}{(1-\rho)\frac{b_n\bar{\gamma}_h}{M_n}+1}\right)\right. \\ &\quad\left. - Q\left(H, \frac{\gamma_{n+1}}{\bar{\gamma}_h r}\cdot\frac{\frac{b_n\bar{\gamma}_h}{M_n}+1}{(1-\rho)\frac{b_n\bar{\gamma}_h}{M_n}+1}\right)\right],\end{aligned} \tag{19}$$

$$
\begin{aligned}
\mathcal{I}21(n) = a_n \sum_{k=0}^{\infty} & \frac{\Gamma(k+H)}{\Gamma(k+1)\Gamma(H)} \left(\frac{\rho}{1-\rho}\right)^k \\
& \times \left(\frac{1}{\frac{b_n \bar{\gamma}_h}{M_n} + \frac{1}{1-\rho}}\right)^{k+H} \\
& \times \left[1 - Q\Big(k+H, \gamma_n^1 \Big(\frac{b_n}{M_n} + \frac{1}{(1-\rho)\bar{\gamma}_h}\Big)\Big)\right] \\
& \times \left[Q\left(k+H, \frac{\gamma_n}{(1-\rho)\bar{\gamma}_h r}\right)\right. \\
& \quad \left. - Q\left(k+H, \frac{\gamma_{n+1}}{(1-\rho)\bar{\gamma}_h r}\right)\right],
\end{aligned} \tag{20}
$$

and

$$
\begin{aligned}
\mathcal{I}22(n) = \frac{1}{2} \sum_{k=0}^{\infty} & \frac{\Gamma(k+H)}{\Gamma(k+1)\Gamma(H)} \rho^k \left(1-\rho\right)^H \\
& \times \left[1 - Q\left(k+H, \frac{\gamma_n^1}{(1-\rho)\bar{\gamma}_h}\right)\right] \\
& \times \left[Q\left(k+H, \frac{\gamma_n}{(1-\rho)\bar{\gamma}_h r}\right)\right. \\
& \quad \left. - Q\left(k+H, \frac{\gamma_{n+1}}{(1-\rho)\bar{\gamma}_h r}\right)\right].
\end{aligned} \tag{21}
$$

While these expressions are admittedly very complex, it is easy to see that $\mathcal{I}1(n)$, $\mathcal{I}21(n)$, and $\mathcal{I}22(n)$ are strongly dependent on the value of the ratio r of the expectations, and on the correlation coefficient ρ of the predicted and the true CSNR. In addition, they are all functions of the number of receive antennas H and of the expected value $\bar{\gamma}_h$ of the CSNR. It will also be seen later that the expressions yield numerical results that are easy to interpret and which correspond with intuition. Other parameters affecting the average BER are the curve fitting parameters a_n, b_n, M_n, γ_n^1 (constant for a given set of transmitter–receiver pairs), and the CSNR interval thresholds γ_n (constant for a given target BER).

The last component of (10) that is needed in order to calculate BER, is the information rate of each code, $R_n \in \{1, 2, \ldots, N\}$. For the case when $2G$-dimensional ($G \in \mathbb{Z}^+$) trellis codes are used, code n's information rate can be expressed as follows [5], [11], [16]:

$$
R_n = (\log_2(M_n) - 1/G) \cdot \frac{L-1}{L}. \tag{22}
$$

Every Lth channel symbol is a pilot symbol and consequently can not convey information; this is reflected in (22).

We are now able to combine (11), (15), (19)–(21), and (22), to obtain the average BER from (10).

B. Average Spectral Efficiency

The ASE is defined as the average transmission rate divided by the bandwidth, and it can be calculated as the average number of bits per channel symbol; i.e., the denominator of (10):

$$\begin{aligned}\mathrm{ASE} &= \sum_{n=1}^{N} R_n P_n \\ &= \sum_{n=1}^{N} \Big(\log_2(M_n) - 1/G\Big) \cdot \left(\frac{L-1}{L}\right) \\ &\qquad \cdot \left(Q\Big(H, \frac{\gamma_n}{r\bar{\gamma}_h}\Big) - Q\Big(H, \frac{\gamma_{n+1}}{r\bar{\gamma}_h}\Big)\right) \qquad (23)\end{aligned}$$

Note that the ASE as defined above is a meaningful measure of system performance *only as long as the target BER constraint is fulfilled*. If the BER becomes greater than BER_0, the system does not provide the desired transmission reliability, and the received data stream might be meaningless to the end user.

V. MAP-Optimal Rayleigh Fading Envelope Prediction

In this section we state the MAP-optimal solution for a fading envelope predictor in Rayleigh fading, and analyze its properties. Note that in an actual implementation of an ACM system, it is probably more feasible to utilize a less complex and thus sub-optimal predictor. The choice of a MAP-optimal solution is made out of a desire to investigate the influence on ACM performance of the *best possible* linear predictor (in an mean-square error sense), for any predictor order. The trade-off between complexity and optimality in a practical system is a subject for further research.

It is shown in [21] and [11] that the MAP-optimal prediction filter coefficient vector for the complex fading amplitude on a Rayleigh channel can be expressed as follows:

$$\mathbf{f}_{j,\mathrm{MAP}}^{\mathsf{T}} = \mathbf{r}_j^{\mathsf{T}} \left(\mathbf{R} + \frac{1}{\bar{\gamma}_h}\mathbf{I}\right)^{-1}. \qquad (24)$$

Remember that the j time index denotes the delay counted in number of channel symbols. As noted earlier, we shall restrict ourselves to predicting the channel only at pilot symbol instants (meaning that it is assumed that every transmitted codeword starts with a pilot symbol). The components of (24) consist of the vector $\mathbf{r}_j^{\mathsf{T}}$ which contains the correlation between the fading at the pilot symbol instants and the fading to be predicted at time instant $i+j$, and the matrix $\mathbf{R}$, which is the autocorrelation matrix of the fading at the pilot symbol instants. Introducing the correlation function $R(\tau)$, which describes the normalized correlation between two instances of fading as a function of the time delay τ between them, components of $\mathbf{r}_j^{\mathsf{T}}$ and $\mathbf{R}$ can be expressed as follows:

$$[\mathbf{r}_j]_k = \frac{1}{\Omega}E[z_h(i+j)z_h(i-kL)] = R((j+kL)T_{\mathrm{s}}) \qquad \text{and} \tag{25}$$

$$[\mathbf{R}]_{kl} = R(|k-l|LT_{\mathrm{s}}). \tag{26}$$

When the familiar *Jakes spectrum* is used (see for instance [3, Sec. 5], [9, Sec. 2.1], or [14, Sec. 12.2.3],) the correlation function $R(\tau)$ will have the following expression:

$$R(\tau) = J_0(2\pi f_{\mathrm{D}}\tau) \tag{27}$$

where $J_0(x)$ is the 0th order Bessel function of the first kind. $f_{\mathrm{D}} = \frac{v}{c}f_c$ [Hz] is the Doppler spread [9] due to terminal mobility at velocity v [m/s] and carrier frequency f_c [Hz]. c [m/s] is the speed of light. In the numerical results to follow, we have used fading correlation described by (27). However, we repeat that this is not a restriction in our theoretical framework.

A. Ratio and Correlation Coefficient

When the MAP-optimal predictor is used, the ratio r of the expected CSNRs (γ, $\hat{\gamma}$) and the correlation coefficient (ρ) between the estimated and the true squared channel gain are equal, and given by the following equation [11]:

$$\rho = \mathbf{r}_j^{\mathsf{T}}(\mathbf{R} + \tfrac{1}{\bar{\gamma}_h}\mathbf{I})^{-1}\mathbf{r}_j = r. \tag{28}$$

This can be shown as follows: For a MAP-optimal predictor, the filter coefficients $\mathbf{f}_j$ are described by (24). Inserting this expression into the formula for the ratio r in (32) from the Appendix, and noting that all the quantities in the expression are real-valued, we obtain

$$r = \mathbf{f}_j^{\mathsf{T}}(\mathbf{R} + \frac{1}{\bar{\gamma}_h}\mathbf{I})\mathbf{f}_j = \mathbf{r}_j^{\mathsf{T}}(\mathbf{R} + \frac{1}{\bar{\gamma}_h}\mathbf{I})^{-1}\mathbf{r}_j. \tag{29}$$

From (24) and (29) we see that the numerator of the expression for ρ in (41) becomes r^2 when the MAP-optimal $\mathbf{f}_j$ from (24) is inserted, which yields the desired result, (28).

B. Illustration of the Prediction Advantage

The discussion here will be concerned with the correlation coefficient ρ. Since ρ and r are shown to be equal in the MAP-optimal case, it will apply to both of them.

An illustration of the advantage of employing an optimized predictor is shown in Fig. 4. The correlation coefficient is there plotted as a function of the normalized time delay $f_{\mathrm{D}}T_{\mathrm{s}} \cdot j$ for the MAP-optimal predictor (solid) and for the case when the last received pilot symbol—divided by the pilot symbol value—is used as a prediction (dash-dotted). The latter is also shown for the case when there is no noise on the channel (dashed). Relevant parameters are the carrier frequency $f_c = 2\,\mathrm{GHz}$ and a terminal velocity $v = 30\,\mathrm{m/s}$ ($108\,\mathrm{km/h}$); the resulting Doppler spread is consequently $f_{\mathrm{D}} = 200\,\mathrm{Hz}$. Assuming a transmission bandwidth of $B = 400\,\mathrm{kHz}$ and Nyquist signalling, the symbol duration is $T_{\mathrm{s}} = 2.5\,\mu\mathrm{s}$, and the normalized Doppler spread will be $f_{\mathrm{D}}T_{\mathrm{s}} = 5 \cdot 10^{-4}$. Other parameters employed in Fig. 4 are expected CSNR $\bar{\gamma}_h = 20\,\mathrm{dB}$ and prediction filter length $K = 1000$. The pilot symbols are inserted every 10th channel symbol, i.e., $L = 10$. Thus, the past horizon used by the predictor in this case spans the last 10 000 symbol periods, sampled at 10-symbol intervals.

The expression for ρ in (28) is easily seen to be dependent on $\bar{\gamma}_h$—an increase in the expected CSNR will naturally cause an increase of ρ, as is illustrated in Fig. 5 by letting $\bar{\gamma}_h$ run from $0\,\mathrm{dB}$ to $40\,\mathrm{dB}$.

Some comments on the choice of pilot symbol period are in order. The pilot symbol transmission and detection can be viewed as sampling of a band-limited process. The Doppler spectrum is typically band-limited to f_{D} [9, Ch. 2]; thus the pilot symbols should be transmitted at a rate of at least $2 \cdot f_{\mathrm{D}}$, according to the sampling theorem [22]. Hence, $L \leq 1/(2f_{\mathrm{D}}T_{\mathrm{s}})$, and any $L \leq 1000$ would yield adequate sampling when the Doppler spread is $f_{\mathrm{D}}T_{\mathrm{s}} = 5 \cdot 10^{-4}$. However, note that this is only valid when the pilot symbols are not corrupted by noise, which is always present in a wireless communication system. Meyr *et al.* [14, Sec. 14.2.2] suggest that if the system throughput is to be maximized, pilot symbols need to be transmitted at much smaller intervals than the Nyquist interval suggested by the Doppler spectrum on a noisy channel. As Fig. 6 illustrates, relatively small variations of L yields substantial improvement of ρ—even when L is several orders of magnitude smaller than 1000. The variations in correlation are most notable for the smaller CSNRs. For $\bar{\gamma}_h = 0$ or $10\,\mathrm{dB}$, the relative gain of decreasing the pilot symbol spacing from $L = 15$ to 5 is substantial compared to when the CSNR is more advantageous. As will be shown in the next section, the BER performance also benefits from the oversampling following from a smaller L.

The number of pilot symbols upon which the prediction is done must of course be limited, yet remain sufficiently high. Meyr *et al.* [14, Sec. 14.2.2] claim that quasi-optimal performance will be achieved if $(K/2)L \gg 1/(f_{\mathrm{D}}T_{\mathrm{s}})$, for non-causal detection. A practical problem is then that the filter order K can easily become very large if L is to be kept small. For instance, with parameters from the discussion above ($f_{\mathrm{D}}T_{\mathrm{s}} = 5 \cdot 10^{-4}$ and $L = 10$), $K \gg 400$. For $L = 5$, $K \gg 800$. Fortunately, the results indicate that L may be decreased without necessarily increasing K correspondingly; note for instance that K is kept constant at $K = 1000$ in Fig. 6. The effect of increasing the filter order while keeping L constant is shown in Fig. 7. Several authors have viewed the fading as an auto-regressive process [23], [24], which may explain the minute advantage of increasing K compared to the improvement resulting from decreasing L shown in Fig. 6.

To sum up, our choice of predictor order $K = 1000$ in subsequent experiments is not seen as a practical predictor length, but rather intended as an approximation to the use of an infinite-order predictor, taking the entire past into account. This results in quasi-optimal performance, i.e., it approximates the best possible predictor performance for our channel model. Practical predictors will always result in a performance deterioration compared to these results. Some more experimental results illustrating the effects of reducing the predictor order are provided in [25].

VI. EXAMPLE SYSTEM

In order to assess bounds for the performance of a system incorporating antenna diversity and channel prediction, a specific system similar to the one in [5] is investigated. The a_n and b_n parameters for the BER expression of the individual codes are summarized in Table I with the calculated thresholds γ_n between the CSNR intervals.

Parameters not directly tied to the codes, albeit dependent on the implementation, are the carrier frequency $f_c = 2\,\mathrm{GHz}$, a bandwidth of $B = 400\,\mathrm{kHz}$, and a terminal velocity of $v = 30\,\mathrm{m/s}$. A prediction filter length of $K = 1000$ is utilized.

A. BER Performance

In order to calculate the average BER, the expressions from Section IV-A are employed. In Fig. 8, the BER has been plotted as a function of subchannel CSNR and of feedback delay, for a pilot symbol spacing of $L = 10$. It is assumed that the system uses two receive antennas and combines the signals with MRC. When a target BER_0 is decided, the operation of the system will be acceptable whenever $\mathrm{BER} < \mathrm{BER}_0$. When $\mathrm{BER} > \mathrm{BER}_0$, the system does not operate properly. The shape of the BER surface is therefore not significant, except for the contour at $\mathrm{BER}_0 = 10^{-4}$. In Fig. 9, the contour lines at $\mathrm{BER} = 10^{-4}$ have been plotted for pilot symbol spacing $L = 5$, 10, 15, and for $H = 1$, 2, and 4 receive antennas. The middle (dashed) line for $H = 2$ corresponds to the contour curve at $\mathrm{BER} = 10^{-4}$ in Fig. 8. The figure shows that a large improvement (in the form of lowering the CSNR requirements, or equivalently, allowing for a longer delay) can be achieved by using more than one receive antenna. Naturally, decreasing the pilot symbol spacing will also lead to some performance gain since the quality of the predicted instantaneous CSNR is increased.

When the CSNR is $10\,\mathrm{dB}$ and when every $L = 10$th channel symbol is a pilot symbol, $H = 1$ receive antenna and the vehicle speed, carrier frequency and transmission bandwidth assumed earlier lead to an acceptable normalized delay of $j \cdot f_{\mathrm{D}} T_{\mathrm{s}} = 0.038$. This corresponds to an actual delay of $j \cdot T_{\mathrm{s}} = 190\,\mu\mathrm{s}$ or $j = 76$ symbols. Increasing the number of receive antennas to $H = 2$ leads to an acceptable delay of $350\,\mu\mathrm{s}$. Note that the CSNR here is the CSNR *per antenna branch*, and the CSNR after MRC will be approximately $13\,\mathrm{dB}$.

B. ASE Performance

As explained in the previous section, the BER can largely be divided into two regions—acceptable (smaller than BER_0) and unacceptable (larger than BER_0). The BER analysis is therefore significant mainly for determining acceptable operation regions. As long as the system can be relied upon to operate at acceptable BER levels, the ASE is the key performance feature.

Using the expression in (23), the ASE is in Fig. 10 plotted as a function of CSNR and of delay (similarly to in Fig. 8). The contour shown on the resulting 3-dimensional curve corresponds to the BER contours in Fig. 9, and divides the ASE curve into a "relevant" part (where the BER conatraints are fulfilled, to the left of the contour) and a "irrelevant" part (where the BER constraints are violated, to the right of the contour). As Fig. 10 indicates, as long as the system is in the acceptable operating region, ASE seems to be almost independent of the normalized time delay. This is reasonable, since the delay only affects the expected value $r\bar{\gamma}_h$ of the predicted CSNR, which appears in the second argument to the normalized incomplete gamma function in (23). Another way to state this is that the probability of using each individual code is not changed very much by varying the delay, even though increased delay of course means increased probability of using each code incorrectly. Thus, at some given normalized time delay value, which is very dependent on the expected subchannel CSNR, the system slides into the "unacceptable" region to the right of the contour.

In Fig. 11, the ASE is plotted for zero delay, as a function of the expected subchannel CSNR. Plots have been made for $L = 5$, 10, 15, and for $H = 1$, 2, and 4 receive antennas. It is apparent that the ASE reaches a ceiling when the CSNR grows large, the ceiling being dependent on L. The reason is that the $2G$-dimensional codes have a spectral efficiency of $\log_2 M_n - 1/G$. For the system considered here, $G = 2$. The spectral efficiency of the largest code (and hence the largest possible spectral efficiency for the set of codes) is consequently 8.5. Remember, however, that the spectral efficiency of the system under consideration is also affected by the pilot symbol spacing L. A smaller L naturally leads to a smaller relative part of the transmission time being available for transmission of information. Thus the maximal possible ASE is $R_8 = 8.5 \cdot (L-1)/L$. For $L = 15$ the maximal possible ASE is 7.933, decreasing to 6.8 for $L = 5$.

It is here emphasized that the CSNR under consideration is again the subchannel CSNR, thus most of the difference between the $H = 1$, 2, and 4 curve bundles in Fig. 11 stems from the $\bar{\gamma} = H\bar{\gamma}_h$ gain. Even if the results for the BER in Fig. 9 were also affected by the beneficial total CSNR of an MRC system when increasing the number of antennas, the BER takes advantage of the lower variance of the underlying gamma distribution when H is increased. This is not the case for the ASE.

VII. Discussion and Conclusions

The results obtained reveal that delay in the feedback channel and erroneous CSI due to noise may have a significant impact on the BER in an ACM system. The increased BER limits the region where the system can operate reliably with respect to permitted delay and average CSNR. However, using MAP-optimal fading envelope prediction, the results show that ACM is still feasible also for practical feedback delays and channel model parameters. In an actual implementation of an ACM system, it might however be more feasible to utilize a less complex and thus sub-optimal predictor. The trade-off between complexity and optimality is a subject for further research.

The system under consideration utilizes pilot symbols with fixed power, and equal pilot symbol and information symbol power. It has not been claimed here that this is optimal; indeed, for uncoded adaptive modulation it has recently been shown that it is not [26]. More power to the pilot symbols yields a better estimate, whereas less pilot symbol power would leave more to the actual transmission of data. A system where the pilot symbol power is adapted to the channel quality might also be considered. Similar arguments apply to the pilot symbol spacing. Decreasing the pilot symbol spacing under adverse channel conditions might yield a better channel estimate which in turn could cause a lower overall BER—or equivalently a larger operating region.

Finally, we re-emphasize the following: Our numerical examples here are obtained for an ACM system with CSNR thresholds chosen to fulfill the target BER constraint under the assumption of perfect CSI. However, our analytical framework allows for any choice of CSNR thresholds. By varying the thresholds we will perform a trade-off between ASE and system robustness with respect to faulty CSI. Increasing the thresholds will mean a smaller ASE, but a higher tolerance against feedback delay and noise. Analytically based approaches for performing such an increase are suggested in [26] for uncoded adaptive modulation, and in [27] for a similar ACM framework as that considered in this paper. One avenue for future research is to combine the analytical framework in this paper with these approaches, to explore the possible trade-offs in the ASE-robustness design space.

APPENDIX
STATISTICAL PROPERTIES OF THE PREDICTED FADING AMPLITUDE

Tang *et al.* [8] derive general formulas for the expectation of the square of a linear, non-causal estimate of the fading amplitude, and for the power correlation coefficient between the estimate and the true fading amplitude. It is here demonstrated that the formulas also apply to the case of linear prediction.

A. *Expectation of* $\hat{\alpha}^2$

The expectation of $\hat{\alpha}^2$ (or, equivalently, the variance of $\hat{z}_h$) can be expressed as follows:

$$\begin{aligned}\widehat{\Omega} = E[\hat{\alpha}^2] = E[|\hat{z}_h|^2] = E[|\mathbf{f}_j^{\mathsf{H}}\tilde{\mathbf{z}}_h|^2] = \mathbf{f}_j^{\mathsf{H}} E[\tilde{\mathbf{z}}_h\tilde{\mathbf{z}}_h^{\mathsf{H}}]\mathbf{f}_j \\ = \mathbf{f}_j^{\mathsf{H}}\left(E[\mathbf{z}_h\mathbf{z}_h^{\mathsf{H}}] + \frac{1}{a_p^2}E[\mathbf{n}_h\mathbf{n}_h^{\mathsf{H}}]\right)\mathbf{f}_j \end{aligned} \quad (30)$$

where $\mathbf{n}_h$ is the noise accompanying the fading $\mathbf{z}_h$. Since the noise is assumed to be white, its covariance matrix is simply the $K \times K$ identity matrix multiplied by the noise variance N_0B. Thus,

$$\widehat{\Omega} = \Omega\mathbf{f}_j^{\mathsf{H}}\mathbf{R}\mathbf{f}_j + \frac{N_0B}{a_p^2}\|\mathbf{f}_j\|^2 \quad (31)$$

Defining r as the ratio of the $\widehat{\Omega}$ to the Ω, and assuming that $a_p = \sqrt{P}$ where P is the average transmit power, we obtain

$$r = \frac{\widehat{\Omega}}{\Omega} = \mathbf{f}_j^{\mathsf{H}}\mathbf{R}\mathbf{f}_j + \frac{\|\mathbf{f}_j\|^2}{\bar{\gamma}_h}. \quad (32)$$

B. *Correlation Coefficient*

In calculating an expression for the correlation coefficient $\rho = \frac{E[\hat{\alpha}_h^2\alpha_h^2] - \Omega^2 r}{\Omega^2 r}$ between the predicted and the true value of the CSNR at time index $(n+j)$, we first concentrate on the correlation between $\hat{\alpha}_h^2$ and α_h^2:

$$\begin{aligned} E[\hat{\alpha}_h^2\alpha_h^2] &= E[|\mathbf{f}_j^{\mathsf{H}}\tilde{\mathbf{z}}_h|^2 \cdot |z_h(n+j)|^2] \\ &= \mathbf{f}_j^{\mathsf{H}} E[(\mathbf{z}_h + \frac{1}{a_p}\mathbf{n})(\mathbf{z}_h^{\mathsf{H}} + \frac{1}{a_p}\mathbf{n}^{\mathsf{H}}) \cdot |z_h(n+j)|^2]\mathbf{f}_j. \end{aligned} \quad (33)$$

Since noise and fading are statistically independent and both zero-mean,

$$\begin{aligned} E[\hat{\alpha}_h^2\alpha_h^2] &= \mathbf{f}_j^{\mathsf{H}} E[\mathbf{z}_h\mathbf{z}_h^{\mathsf{H}} \cdot |z_h(n+j)|^2 \\ &\quad + \frac{1}{a_p^2}\mathbf{n}\mathbf{n}^{\mathsf{H}} \cdot |z_h(n+j)|^2]\mathbf{f}_j \\ &= \mathbf{f}_j^{\mathsf{H}} E[\mathbf{z}_h\mathbf{z}_h^{\mathsf{H}} \cdot |z_h(n+j)|^2]\mathbf{f}_j + \Omega\frac{N_0B}{a_p^2}\|\mathbf{f}_j\|^2. \end{aligned} \quad (34)$$

Again invoking the assumption that $a_p = \sqrt{P}$, using $\bar{\gamma}_h = \frac{\Omega \cdot P}{N_0 B}$, and introducing $\mathbf{z}_{r,h} = \mathrm{Re}(\mathbf{z}_h)$ (and similarly for the scalar $z_h(n+j)$), where $\mathrm{Re}(\cdot)$ indicates the real part (in-phase or I-component) of a complex baseband symbol,

$$E[\hat{\alpha}_h^2 \alpha_h^2] = 2\mathbf{f}_j^{\mathsf{H}} E[\mathbf{z}_{r,h}\mathbf{z}_{r,h}^{\mathsf{T}} \cdot z_{r,h}^2(n+j)]\mathbf{f}_j \\ + 2\mathbf{f}_j^{\mathsf{H}} E[\mathbf{z}_{r,h}\mathbf{z}_{r,h}^{\mathsf{T}}] E[z_{r,h}^2(n+j)]\mathbf{f}_j + \frac{\Omega^2}{\bar{\gamma}_h}\|\mathbf{f}_j\|^2. \tag{35}$$

Since $E[\mathbf{z}_h\mathbf{z}_h^{\mathsf{H}}] = E[\mathbf{z}_{r,h}\mathbf{z}_{r,h}^{\mathsf{T}}] + E[\mathbf{z}_{i,h}\mathbf{z}_{i,h}^{\mathsf{T}}] = \Omega\mathbf{R}$, and $E[\mathbf{z}_{r,h}\mathbf{z}_{r,h}^{\mathsf{T}}] = E[\mathbf{z}_{i,h}\mathbf{z}_{i,h}^{\mathsf{T}}]$,

$$E[\mathbf{z}_{r,h}\mathbf{z}_{r,h}^{\mathsf{T}}] = \frac{\Omega}{2}\mathbf{R}, \tag{36a}$$

and it turns out that similar expressions exist for the scalar variance/cross-correlation expressions, namely,

$$E[\mathbf{z}_{r,h} z_{r,h}(n+j)] = \frac{\Omega}{2}\mathbf{r}_j \quad \text{and} \quad E[z_{r,h}^2(n+j)] = \frac{\Omega}{2}. \tag{36b}$$

Thus it can be concluded that,

$$E[\hat{\alpha}_h^2 \alpha_h^2] = 2\mathbf{f}_j^{\mathsf{H}} E[\mathbf{z}_{r,h}\mathbf{z}_{r,h}^{\mathsf{T}} \cdot z_{r,h}^2(n+j)]\mathbf{f}_j \\ + \frac{1}{2}\Omega^2\mathbf{f}_j^{\mathsf{H}}\mathbf{R}\mathbf{f}_j + \frac{\Omega^2}{\bar{\gamma}_h}\|\mathbf{f}_j\|^2 \tag{37}$$

It is known (see for instance [28, Sec. 8-2]) that, for zero-mean, real, Gaussian RVs a, b, c, and d, the following formula can be used to find the fourth-order moment:

$$E[abcd] = E[ab]E[cd] + E[ac]E[bd] + E[ad]E[bc]. \tag{38}$$

It is demonstrated in [11] that the above formula may also be used when two of the RVs are column vectors,

$$E[ab\mathbf{c}\mathbf{d}^{\mathsf{T}}] = E[ab]E[\mathbf{c}\mathbf{d}^{\mathsf{T}}] + E[a\mathbf{c}]E[b\mathbf{d}^{\mathsf{T}}] + E[b\mathbf{c}]E[a\mathbf{d}^{\mathsf{T}}]. \tag{39}$$

With (39) in mind, and utilizing the expression for r from (32),

$$\begin{aligned} E[\hat{\alpha}_h^2 \alpha_h^2] &= 2\mathbf{f}_j^{\mathsf{H}}\Big(E[\mathbf{z}_{r,h}\mathbf{z}_{r,h}^{\mathsf{T}}]E[z_{r,h}^2(n+j)] \\ &\quad + 2E[\mathbf{z}_{r,h} z_{r,h}(n+j)]E[\mathbf{z}_{r,h}^{\mathsf{T}} z_{r,h}(n+j)]\Big)\mathbf{f}_j \\ &\quad + \frac{1}{2}\Omega^2\mathbf{f}_j^{\mathsf{H}}\mathbf{R}\mathbf{f}_j + \frac{\Omega^2}{\bar{\gamma}_h}\|\mathbf{f}_j\|^2 \\ &= 4\mathbf{f}_j^{\mathsf{H}}\Big(\frac{\Omega}{2}\mathbf{r}_j\frac{\Omega}{2}\mathbf{r}_j^{\mathsf{T}}\Big)\mathbf{f}_j + \Omega^2\mathbf{f}_j^{\mathsf{H}}\mathbf{R}\mathbf{f}_j + \frac{\Omega^2}{\bar{\gamma}_h}\|\mathbf{f}_j\|^2 \\ &= \Omega^2|\mathbf{f}_j^{\mathsf{H}}\mathbf{r}_j|^2 + \Omega^2 r \end{aligned} \tag{40}$$

By inserting (40) in (9), the final expression for the correlation coefficient is obtained as follows:

$$\rho = \frac{|\mathbf{f}_j^{\mathsf{H}}\mathbf{r}_j|^2}{r} = \frac{|\mathbf{f}_j^{\mathsf{H}}\mathbf{r}_j|^2}{\mathbf{f}_j^{\mathsf{H}}\mathbf{R}\mathbf{f}_j + \frac{1}{\bar{\gamma}_h}\|\mathbf{f}_j\|^2} \tag{41}$$

ACKNOWLEDGMENTS

The authors would like to express their thanks to the anonymous reviewers, whose comments significantly improved the quality of this paper.

REFERENCES

[1] M. Nakagami, "The m-distribution—a general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*, W. C. Hoffman, Ed. Pergamon Press (London, Oxford, New York, Paris), June 1960, pp. 3–36, proceedings of a Symposium held at the University of California, Los Angeles, June 18–20, 1958.

[2] 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.

[3] M.-S. Alouini and A. Goldsmith, "Adaptive modulation over Nakagami fading channels," *Wireless Personal Communications*, vol. 13, pp. 119–143, 2000.

[4] L. Hanzo, C. H. Wong, and M.-S. Yee, *Adaptive Wireless Transceivers: Turbo-Coded, Turbo-Equalized and Space-Time Coded TDMA, CDMA and OFDM Systems*. John Wiley & Sons, LTD/IEEE Press, 2002.

[5] K. J. Hole, H. Holm, and G. E. Øien, "Adaptive multidimensional coded modulation over flat fading channels," *IEEE J. Select. Areas Commun.*, vol. 18, no. 7, pp. 1153–1158, July 2000.

[6] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[7] J. M. Torrance and L. Hanzo, "Comparative study of pilot symbol assisted modem schemes," in *Sixth International Conference on Radio Receivers and Associated Systems*. Bath, UK: IEE, Sept. 1995, pp. 36–41.

[8] X. Tang, M.-S. Alouini, and A. Goldsmith, "Effect of channel estimation error on M-QAM BER performance in Rayleigh fading," *IEEE Trans. Commun.*, vol. 47, no. 12, pp. 1856–1864, Dec. 1999.

[9] G. L. Stüber, *Principles of Mobile Communication*, 2nd ed. Norwell, MA: Kluwer Academic Publishers, 2001.

[10] ETSI, "Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television," European Brodcasting Union; European Telecommunications Standards Institute (ETSI), F-06921 Sophia Antipolis Cedex, France, Standard ETSI EN 300 744 v1.2.1 (1999-07), 1999.

[11] H. Holm, "Adaptive coded modulation performance and channel estimation tools for flat fading channels," Ph.D. dissertation, Norwegian University of Science and Technology, Trondheim, Norway, Mar. 2002, http://www.tele.ntnu.no/projects/beats/Documents/HenrikHolmThesis.pdf.

[12] K. J. Hole and G. E. Øien, "Spectral efficiency of adaptive coded modulation in urban microcellular networks," *IEEE Trans. Veh. Technol.*, vol. 50, no. 1, pp. 205–222, Jan. 2001.

[13] G. E. Øien and K. J. Hole, "Maximum average spectral efficiency for slowly varying Rayleigh fading channels with pilot-symbol-assisted channel estimation," in *Proc. European Personal Mobile Communications Conference*, Vienna, Austria, Feb. 2001.

[14] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing*. New York, NY: Wiley Series in Telecommunications and Signal Processing, 1998.

[15] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*, ser. Signal Processing Series. Englewood Cliffs, NJ: Prentice Hall, 1992.

[16] K. J. Hole, H. Holm, and G. E. Øien, "Performance analysis of adaptive coded modulation with antenna diversity and feedback delay," *Telektronikk – Information Theory and its Applications*, vol. 98, no. 1, pp. 106–113, 2002, published by Telenor, a Norwegian telecommunications group. Complete paper available from the authors, upon e-mail request to oien@tele.ntnu.no.

[17] N. M. Temme, *Special Functions: An Introduction to the Classical Functions of Mathematical Physics*. New York: John Wiley & Sons, Inc, 1996.

[18] M. Nagao and M. Kadoya, "Two-variate exponential distribution and its numerical table for engineering applications," *Bull. Disas. Prev. Res. Inst., Kyoto Univ.*, vol. 20, no. 178, pp. 183–210, March 1971.

[19] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6th ed. San Diego: Academic Press, 2000.

[20] M. Abramowitz and I. E. Stegun, Eds., *Handbook of Mathematical Functions*, tenth printing ed. United States Department of Commerce, 1972.

[21] G. E. Øien, H. Holm, and K. J. Hole, "Channel prediction for adaptive coded modulation on Rayleigh fading channels," in *Proc European Signal Processing Conference (EUSIPCO) 2002*, Toulouse, France, Sept. 2002.

[22] A. Duel-Hallen, S. Hu, and H. Hallen, "Long-range prediction of fading signals," *IEEE Signal Processing Mag.*, vol. 17, no. 3, pp. 62–75, May 2000.

[23] A. Aghamohammadi, H. Meyr, and G. Ascheid, "Adaptive synchronization and channel parameter estimation using an extended Kalman filter," *IEEE Trans. Commun.*, vol. 37, no. 11, pp. 1212–1219, Nov. 1989.

[24] Y. Zhang, M. P. Fritz, and S. B. Gelfand, "Soft output demodulation on frequency-selective Rayleigh fading channels using AR channel models," in *IEEE Global Telecommunications Conference (GLOBECOM '97)*, vol. 1, 1997, pp. 327–331.

[25] G. Øien, R. K. Hansen, D. V. Duong, H. Holm, and K. J. Hole, "Bit error rate analysis of adaptive coded modulation with mismatched and complexity-limited channel prediction," in *Proc. IEEE Nordic Signal Processing Symposium (NORSIG-2002)*, Norway, Oct. 2002.

[26] X. Cai and G. Giannakis, "Adaptive PSAM accounting for channel estimation and prediction errors," *submitted to IEEE Trans. on Wireless Communications*, 2003.

[27] G. E. Øien, H. Holm, and K. J. Hole, "Adaptive coded modulation with imperfect channel state information: System design and performance analysis aspects," in *ISWC-2002: Book of Extended Abstracts*, Victoria, BC, Canada, Sept. 2002, pp. 19–20.

[28] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill International Editions, 1991.

Geir E. Øien (S'90–M'2001) was born in Trondheim, Norway in 1965. He received the M.Sc. degree and the Ph.D. degree, both from the Norwegian Institute of Technology (NTH), Trondheim, Norway, in 1989 and 1993, respectively. From 1994 until 1996 he was an associate professor with Stavanger University College in Stavanger, Norway. In 1996 he joined The Norwegian University of Science and Technology in Trondheim as associate professor, and in 2001 he was promoted to full professor. His current research interests include analysis and design of bandwidth-efficient channel coding and modulation schemes for fading channels; wireless channel analysis, estimation, and prediction; analysis and characterization of spatial diversity and MIMO systems; and OFDM-based wireless systems.

Henrik Holm (S'96–M'2003) was born in Sarpsborg, Norway, on January 25, 1972. He received the M.Sc. and Ph.D. degrees in electrical engineering from the Department of Telecommunications at the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 1997 and 2002, respectively.

Since May 2002, he has been a Post Doctoral Associate in Prof. Mohamed-Slim Alouini's group at the Department of Electrical and Computer Engineering at the University of Minnesota, with funding from the NTNU. His current research interests are in the areas of channel estimation and modulation techniques for wireless communications.

PLACE PHOTO HERE

Kjell J. Hole (S'89–M'90) was born in Molde, Norway, on June 1, 1961. He received the B.Sc., M.Sc., and Ph.D. degrees in computer science from the University of Bergen, Norway, in 1984, 1987, and 1991, respectively.

From August 1988 to May 1990, he was a Visiting Scholar in Prof. Jack K. Wolf's group at the Center for Magnetic Recording Research at the University of California San Diego. During 1993, he worked in Dr. Paul Siegel's group at IBM Almaden Research Center, San José, California. Since 1995, he has been a Research Scientist at the University of Bergen and the Norwegian University of Science and Technology with funding from the Norwegian Research Council. His current research interests are in the areas of coding theory and wireless communications.

LIST OF FIGURES

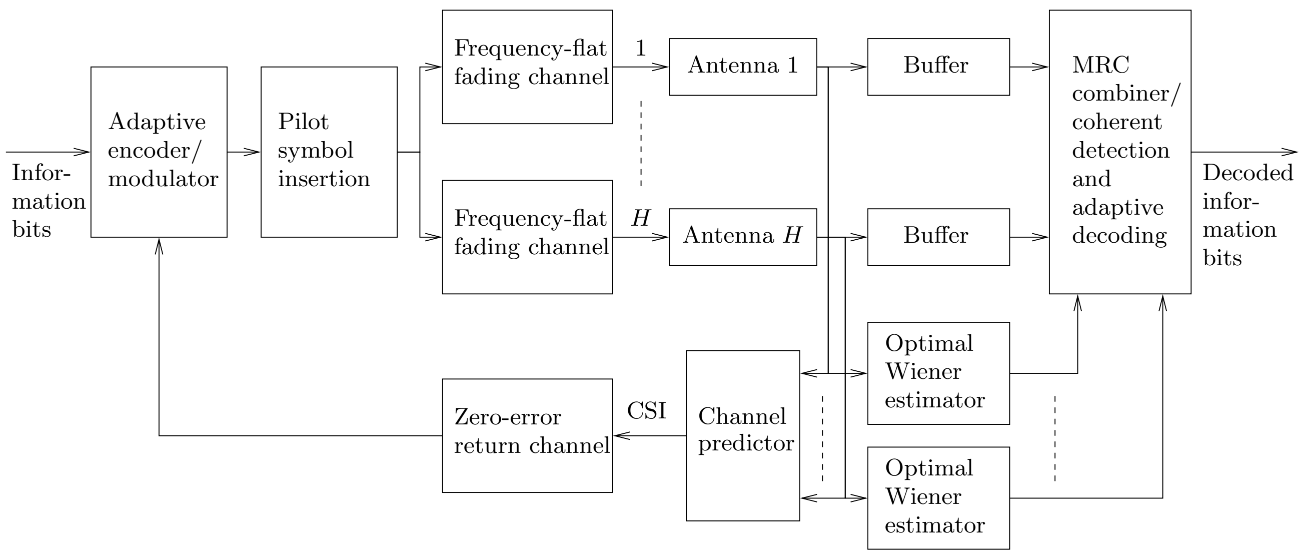

Fig. 1. ACM system with pilot symbol assisted channel estimation (for coherent detection purposes) and prediction (for transmitter adaptation purposes).

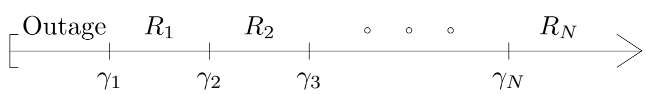

Fig. 2. The CSNR range is split into $N+1$ CSNR intervals. When the instantaneous CSNR falls in the lowest interval, an outage occurs; whereas in the upper N intervals, a transmission mode with rate R_n is employed.

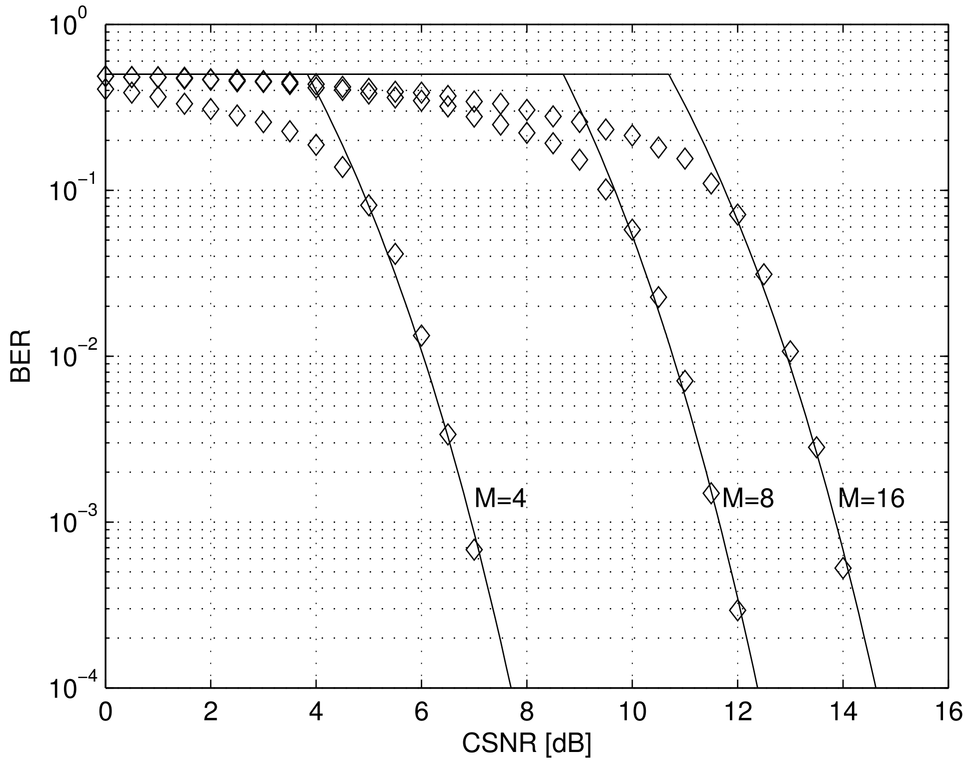

Fig. 3. Accuracy of the BER approximation in (14), for example codes $n = 1, 2, 3$ from [5].

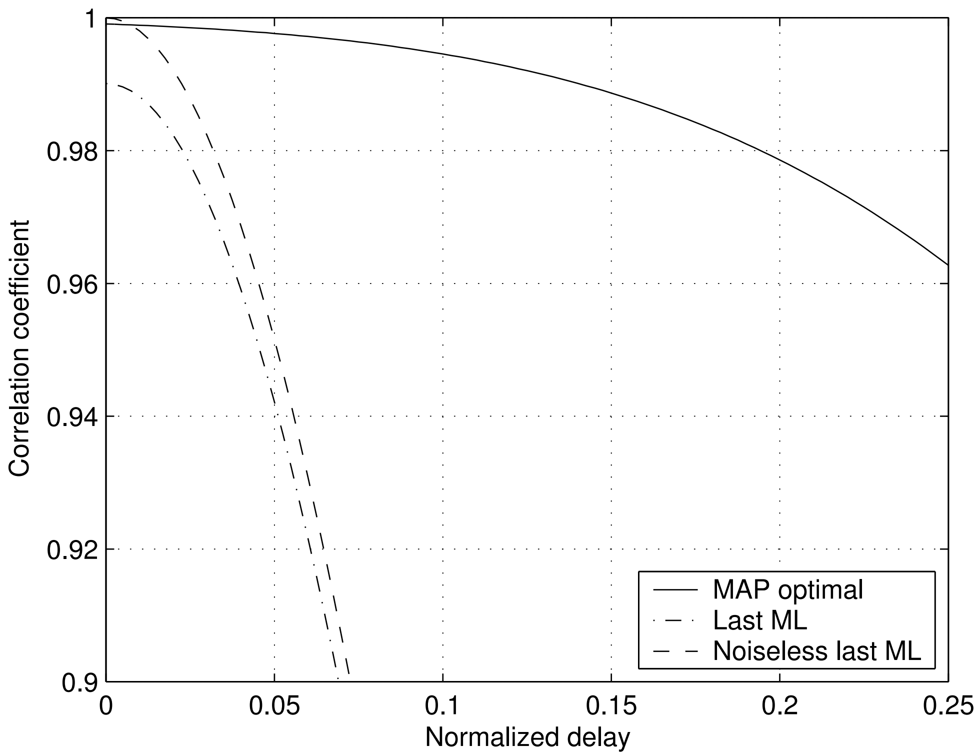

Fig. 4. The correlation coefficient as a function of normalized time delay $f_{\mathrm{D}}T_{\mathrm{s}} \cdot j$, with CSNR $20\,\mathrm{dB}$, $L = 10$, and $K = 1000$.

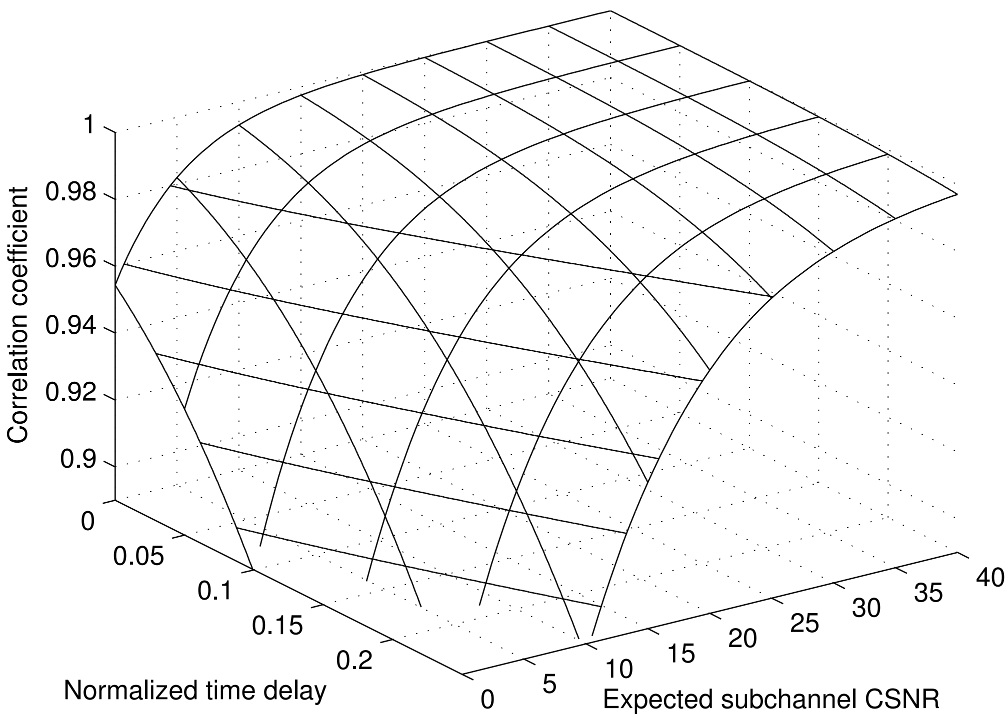

Fig. 5. The correlation coefficient as a function of normalized delay $f_{\mathrm{D}}T_{\mathrm{s}} \cdot j$ and expected subchannel CSNR $\bar{\gamma}_h$ (in dB).

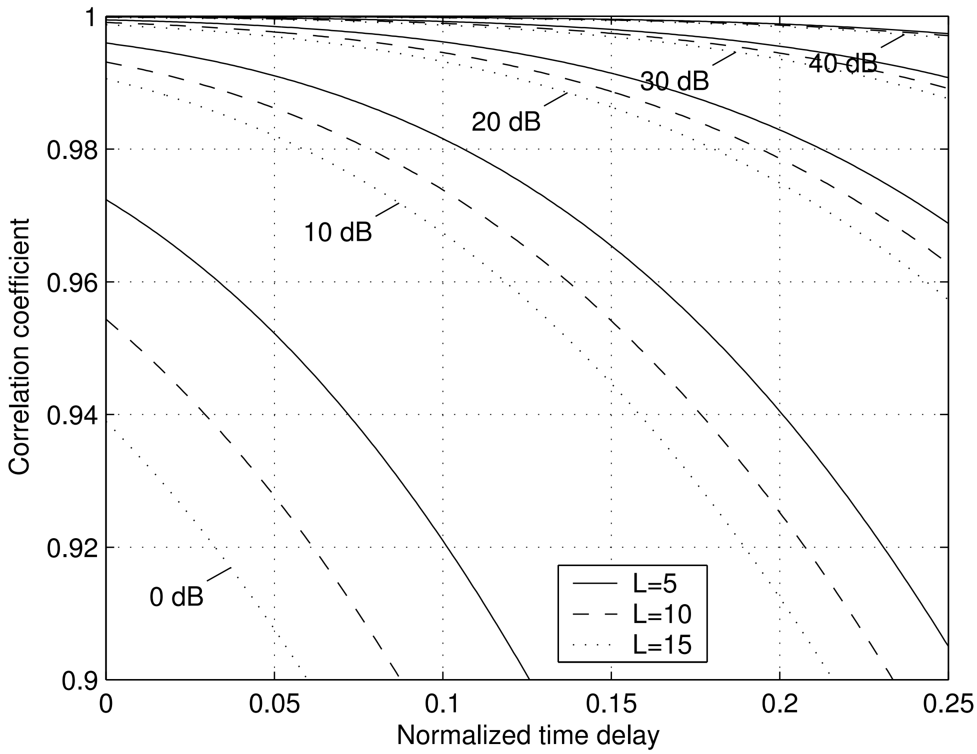

Fig. 6. Correlation coefficient ρ as a function of delay, plotted for expected subchannel CSNR $\bar{\gamma}_h = 0$, 10,..., 40 (measured in dB), and for pilot symbol spacing $L = 5$, 10, and 15. The prediction filter length is $K = 1000$.

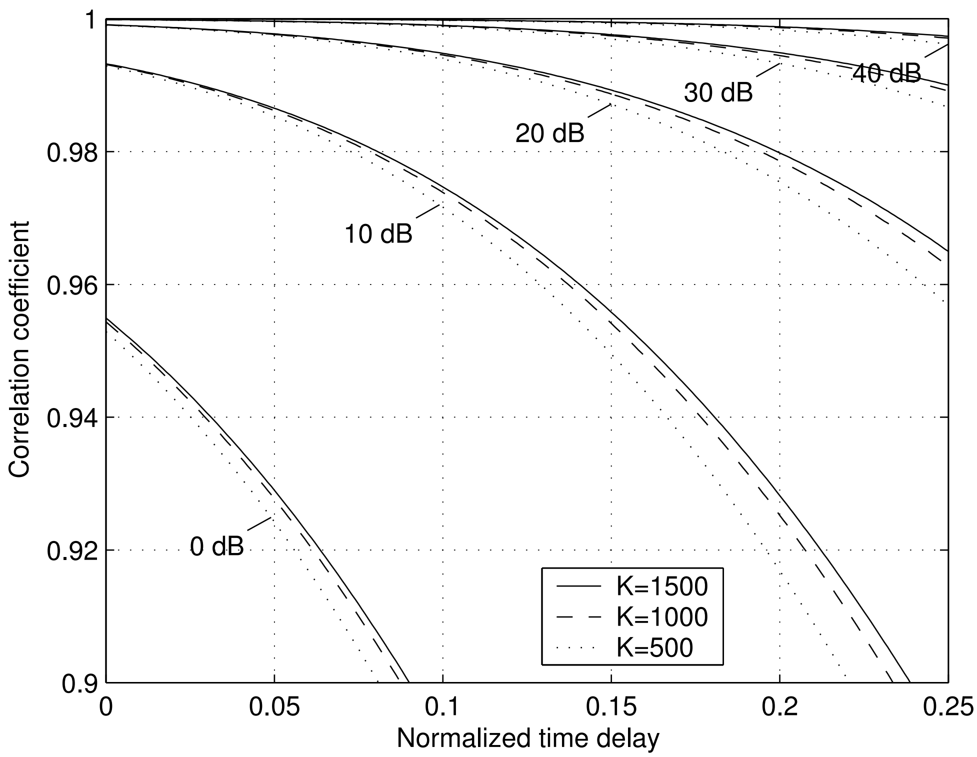

Fig. 7. Correlation coefficient plotted analogously to Fig. 6, however, the pilot symbol spacing is kept still at $L = 10$ and the filter order is varied from $K = 500$ to $K = 1500$.

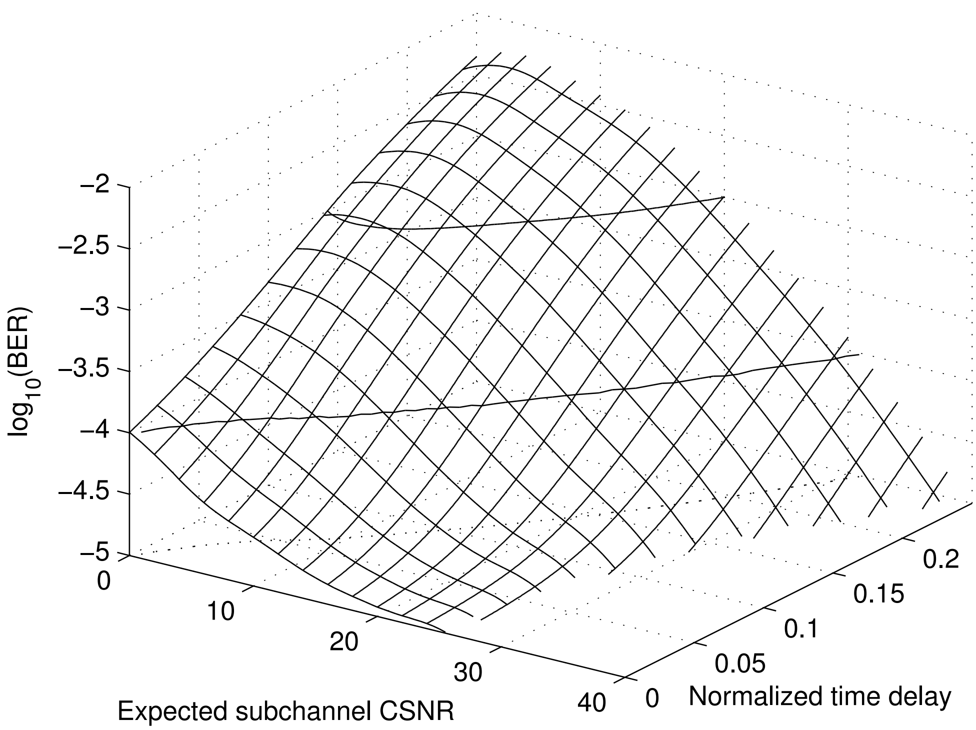

Fig. 8. BER as a function of feedback delay and of expected subchannel CSNR (in dB). $H = 2$ receive antennas are utilized and the pilot symbol spacing is $L = 10$. In this plot—as in all subsequent plots—the prediction filter length is kept constant at $K = 1000$.

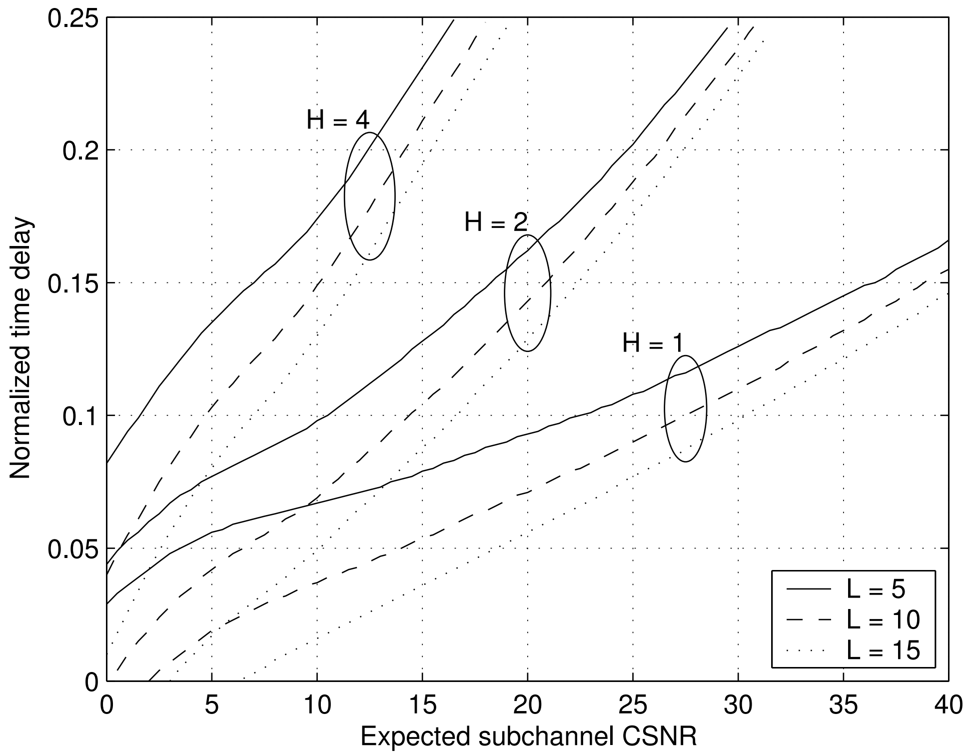

Fig. 9. Regions for which system performance is acceptable, plotted for pilot symbol spacing $L = 5$, 10, 15, and for $H = 1$, 2, and 4 receive antennas. The curves indicate largest delay that is allowed in order to achieve the BER requirements, for a given expected subchannel CSNR (in dB). Thus, acceptable performance results when the point specified by a CSNR/delay combination is below and to the right of the curve for system parameters L and H.

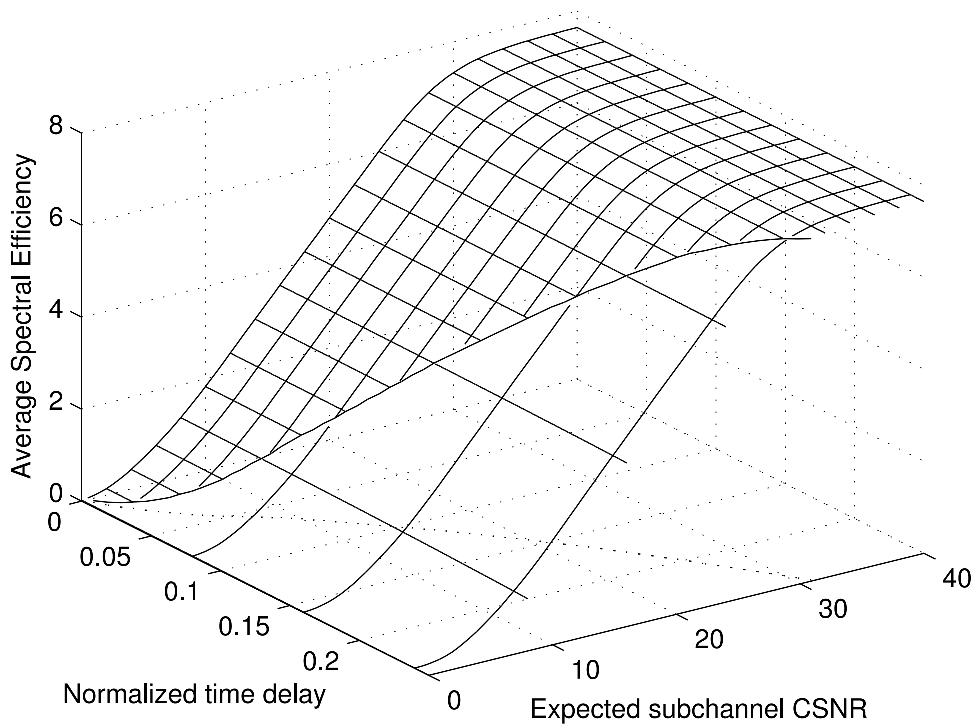

Fig. 10. ASE as a function of feedback delay and expected subchannel CSNR (in dB). $L = 10$, $H = 2$. Note that the ASE is almost independent of the feedback delay when normalized time delay is in the region 0–0.25 (the biggest delay corresponding to an actual delay of $1.25\,\mathrm{ms}$ for the simulation parameters otherwise employed.)

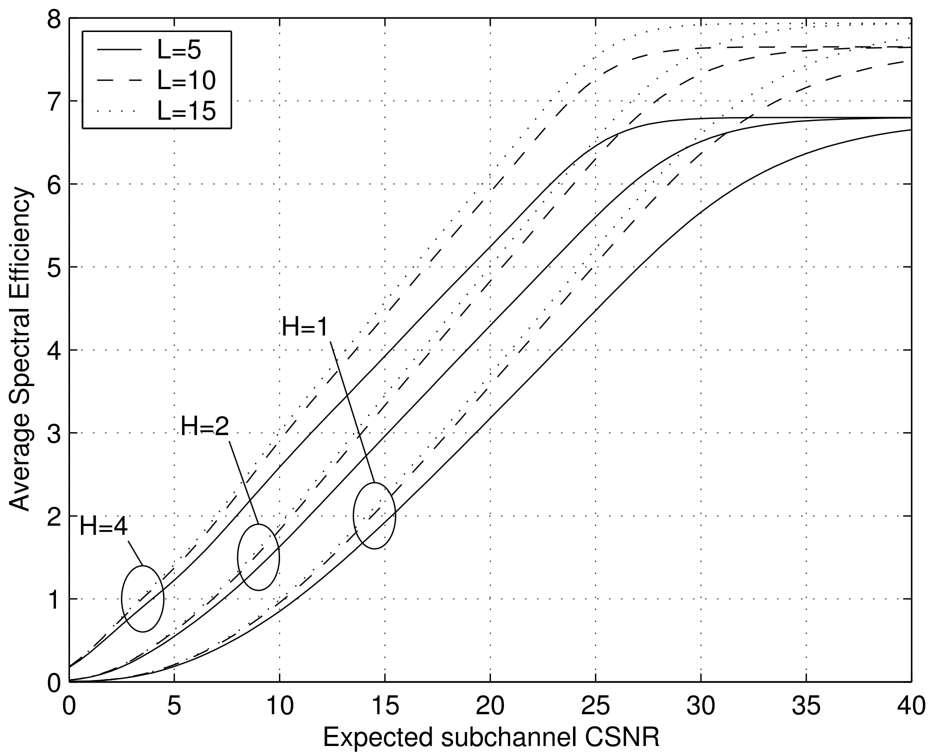

Fig. 11. ASE as a function of expected subchannel CSNR (in dB), plotted for various L and H, for zero delay.

LIST OF TABLES

TABLE I

PARAMETERS a_n AND b_n FOR THE EXAMPLE CODEC, ALONG WITH THRESHOLDS γ_n FOR TARGET $\mathrm{BER}_0 = 10^{-4}$.

n	M_n	a_n	b_n	γ_n (dB)
1	4	188.7471	9.8118	7.7
2	8	288.8051	6.8792	12.4
3	16	161.6898	7.8862	14.6
4	32	142.6920	7.8264	17.6
5	64	126.2118	7.4931	20.8
6	128	121.5189	7.7013	23.7
7	256	79.8360	7.1450	26.9
8	512	34.6128	6.9190	29.7